American Psychological Association

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

2021, Vol. 27, No. 2, 292–306
https://doi.org/10.1037/xap0000310

Who Knows What? Knowledge Misattribution in the Division of Cognitive Labor

Matthew Fisher[1] and Daniel M. Oppenheimer[2]
[1] Marketing Department, Southern Methodist University
[2] Department of Social and Decision Sciences, Carnegie Mellon University

As technology advances, people increasingly outsource cognitive tasks and can more easily access others' knowledge. While externalized aids often support human abilities, they may also make it more difficult for people to assess their own competence. Indeed, using online search engines leads people to treat searchable information as if they already know it (Fisher et al., *Journal of Experimental Psychology: General*, 2015, *144*, 674). Six primary and two supplemental studies (N = 3,262) extend previous research by exploring how illusions of knowledge result from reliance on other agents. After teaming with knowledgeable partners (artificially intelligent agents or human teammates) on a trivia quiz, people overestimated how well they would perform on future quizzes for which help was not available; this bias was not evident for participants who never received help. Moreover, overconfidence was insensitive to whether assistance was provided on hard versus easy problems, or even whether the assistance was accurate. Receiving outside assistance creates an ambiguity regarding who deserves credit for success or blame for failure. When this ambiguity is removed, people become better calibrated. These results indicate that reliance on technology and outside knowledge may change our view of ourselves—convincing us we are more capable than we really are.

Public Significance Statement

After receiving assistance on a task, either from other people or from technological aids, people overestimate how well they would be able to perform the task on their own (without assistance). Using assistive technology and working in teams can lead us to believe that we are more capable than we really are.

Keywords: transactive memory, cognitive offloading, metacognition, illusions of knowledge

Supplemental materials: https://doi.org/10.1037/xap0000310.supp

Just as physical labor is divided within a society (Smith, 1776), so too is *cognitive* labor (Kitcher, 1990). No single person can know everything. Instead, people must rely on the expertise of others to function in everyday life. At any given time, one may need to access the knowledge of a doctor, a car mechanic, or an IT specialist. Without the distribution of knowledge within a community we would be overwhelmed by everything we would need to know (Hardwig, 1985). People's impressive ability to navigate the networks of knowledge in which they are embedded is supported by competency, even in early childhood, in attributing knowledge to appropriate experts (Lutz & Keil, 2002) and an understanding of how knowledge clusters in the world (Keil et al., 2008). Thus, children and adults alike are able to seek out answers they need from relevant sources.

The sharing of knowledge across a social network has been conceptualized as a transactive memory system (Wegner, 1987). In its simplest form, these systems can be seen in the way romantic couples coordinate the retrieval of information. By using both explicit and implicit strategies, couples recall more information together than they would on their own or paired with a stranger (Harris et al., 2011, 2014; Wegner et al., 1991). This boost in performance is supported by expertise within a particular domain of knowledge combined with well-practiced communication skills—familiar groups more frequently acknowledge and elaborate on others' contributions (Gagnon & Dixon, 2008; Meade et al., 2009). At the dyadic level and beyond, individuals must recognize the relative expertise of others, tracking what is already known and who knows the information they might need in the future.

Just as social networks can support individual cognition, so too can the tools and technology in our environment. For example, the cognitive capacity of memory routinely relies on external tools like memos, reminder notes, and shopping lists (Block & Morwitz, 1999; Harris, 1980; Intons-Peterson & Fournier, 1986). Indeed, the mind can be understood as extending beyond its biological substrate to include resources in the environment (Barrett, 2011; Clark, 2008; Clark & Chalmers, 1998; Hutchins, 1995). In this view, cognitive processes are not confined to a single individual's nervous

This article was published Online First April 8, 2021.
Matthew Fisher https://orcid.org/0000-0002-4338-1638
All preregistration forms, de-identified raw data files, and analytic syntax are available at Open Science Framework (https://osf.io/mg3zn/).
Correspondence concerning this article should be addressed to Matthew Fisher, Marketing Department, Southern Methodist University, Dallas, TX 75275, United States. Email: matthewfisher@smu.edu

system but can be offloaded onto the community of knowledge and available technologies.

Could these networks of knowledge create miscalibrations where people fail to recognize their own limitations when no longer receiving assistance? The study of metacognitive calibration, particularly overconfidence, has been central to the field of decision making (Block & Harper, 1991; Fischhoff et al., 1977; Hoffrage, 2004; Moore & Healy, 2008; Park & Budescu, 2015; Soll & Klayman, 2004) and related fields such as marketing (Alba & Hutchinson, 2000), behavioral finance (Camerer & Lovallo, 1999), and forecasting (Pfeifer, 1994). According to one prominent model (Koriat, 1997), people do not have direct insight into their own cognitive processing, and thus metacognition is best thought of as a form of judgment, in which people integrate available cues to estimate their general knowledge. Notably, in addition to numerous demonstrations that people are often unaware of what they do and do not know, there is also evidence that people struggle to understand what they would or would not know when under conditions different from the status quo. For example, with the hindsight bias (Fischhoff, 1975), people who have been provided information are grossly inaccurate when attempting to guess how they would have behaved had they not known that information.

Typically, such investigations have examined individuals in isolation, rather than individuals as embedded in transactive memory networks and in the context of cognitive offloading. However, offloading denies people the internal, mnemonic cues that are so essential to metacognitive judgments (Koriat, 1997). Thus, one might expect offloading to yield metacognitive challenges above and beyond normal metacognitive miscalibration (cf., Hargis & Oppenheimer, 2020).

Indeed, offloading is accompanied by a host of cognitive consequences. For example, our reliance on new technologies like computers and the Internet changes our strategies for storing and retrieving information. People are better at remembering where newly learned information is stored on a computer than at remembering the information itself (Sparrow et al., 2011). People's estimates of how much they have learned are better calibrated when learning on paper versus a computer (Ackerman & Goldsmith, 2011) and after searching online, people are more likely to overestimate how much they have learned compared to those who are provided the study materials directly (Fisher et al., 2021). Relatedly, when museumgoers photographed objects during their tour, subsequent recall was impaired, suggesting that reliance on externalized memory inhibited internal memory capacities (Henkel, 2014). However, in other ways, these technologies can boost memory—after saving a file of previously learned material, people show improved memory for new content (Storm & Stone, 2015). By externally storing old information, people reduce cognitive interference from the old information and thus increase retention of novel information. Similarly, air traffic controllers showed improvement in prospective memory when tasks were automated, thus reducing extraneous cognitive load (Vortac et al., 1993).

However, while the offloading functions of new technologies have important consequences for memory, there has been limited research exploring how reliance on outside sources affects people's understanding of their own abilities. Some evidence suggests people are largely unaware of their dependence on external sources that supplement their memories. After searching online for the answers to a set of questions, people rated themselves as more knowledgeable (without outside assistance) compared to those who did not search online (Fisher et al., 2015). The topics participants found online (e.g., "why are there phases of the moon?") were separate from the topics used to assess participants' estimates of their own knowledge (e.g., topics from American history). Thus, an increase in knowledge ratings showed the conflation of external knowledge accessible through the Internet and knowledge "in the head." Converging evidence shows that Internet search increases one's cognitive self-esteem: after finding information through a Google search people report increased self-perceptions in domains supported by the Internet (e.g., thinking and memory; Ward, 2013). In other words, the cognitive boost provided by an external aid was misattributed to the self. Moreover, people think they have a better personal understanding of novel topics (e.g., glowing rocks, liquid helium weather systems) after being told that scientists understand the phenomenon (Sloman & Rabb, 2016).

While previous studies have demonstrated that outsourcing knowledge yields overconfidence effects, to date there is little understanding of why, mechanistically, those effects occur, nor where the boundary conditions lie. In the current research, we examine plausible boundary conditions and narrow in on a specific mechanism—performance ambiguity—that has not previously been considered a candidate mechanism in the aforementioned literature. Moreover, we directly test and find evidence for the ambiguity mechanism. This mechanistic understanding better explicates the domains where the effect will and will not be expected to occur, allowing for informed generalization to applied domains.

As reliance on technology, algorithms, artificial intelligence, and access to others' knowledge increases, it is important to understand the impact of outsourced knowledge. The Internet has accelerated the availability of information, creating an especially salient example of knowledge misattribution. However, this metacognitive dilemma could exist in many other real-world contexts as well. Recent technological developments like navigation apps (e.g., Google Maps) or fitness tracking devices (e.g., FitBit) make it unclear how well people can accomplish their goals without the outside assistance. People may mistakenly overestimate their own competence after relying on these external aids. Similarly, in social contexts people may fail to see their own limitations because of the aid they are receiving from others. For example, in a scientific collaboration, scientists may fail to recognize their dependency on others' expertise, believing they could have done more on their own than would have actually been possible.

The current work advances the previous literature in several key ways. First, we show metacognitive miscalibration occurs for passive receipt of outside assistance. Previous research has largely focused on actively seeking help (e.g., Google search), but in many applied contexts, the assistance occurs without explicit request. Second, we show miscalibration occurs beyond cases of clearly established sources of expertise like Google or scientists. In particular, people overestimate their own abilities not just after receiving technological assistance, but also after receiving social assistance from someone with whom they have no previous experience. Third, we show that plausible boundary conditions, like inaccurate assistance, still produce overconfidence. Finally, we explain these findings by presenting initial evidence for a novel mechanism underlying miscalibration: performance ambiguity.

In the current studies, we show that when paired with technology that helps complete trivia quizzes, people overestimate how accurate they would be on future, novel quizzes where help is not available (Study 1). This overconfidence persists when paired with human teammates (Study 2) and cannot be explained by simply viewing others' success, active participation is required (Study 3). Further, the same pattern occurs regardless of the accuracy of the outside help (Study 4). Lastly, we find this illusion of competence is explained by performance ambiguity: people can take credit for helpful teammates and attribute failure to unhelpful teammates. Once this ambiguity is removed, participants provide better calibrated predictions for their own future performance without outside help (Study 5).

Study 1

To explore how well people understand the division of cognitive labor, we developed a paradigm where participants completed a task either with or without the help of outside technology. Participants answered general knowledge questions (stating the capitals of 15 countries) by either working alone or partnering with a team of two artificially intelligent agents.

Method

Participants

Three hundred and one (122 males, 179 females; $M_{\mathrm{AGE}} = 37.45$, $SD = 11.47$) participants from the United States completed the study online through Amazon Mechanical Turk (Buhrmester et al., 2011).[1] Sample size was selected with a power analysis using an estimated effect size based on pilot testing (Power > .90). Large sample size was the primary strategy for maximizing power—later studies also employed attention checks, IP address checks, and use of stimuli as a random factor (details provided below). Participants were not allowed to participate in more than one of the current studies.

Materials

In the main task of Study 1, participants completed a 15-question country capital quiz. Fifteen countries were listed with a corresponding blank space for the capital. Using data from the popular quiz website sporcle.com, the quiz was constructed so that five of the capitals were difficult, five were intermediate, and five were easy (see Appendix A).

Design and Procedure

Participants in Study 1 were randomly assigned to either the Team or Alone condition. In the Team condition, participants were instructed, "You will be paired with the Wible algorithm and the Eyer Algorithm to complete the capital city quiz. You must work as a team to get the highest score you can!" When the quiz began, these participants chose an item to answer, filled in their own answer, submitted their response, and then were given feedback on whether the capital they submitted was correct or incorrect. Next, the Wible Algorithm would submit a response to one of the remaining questions. Then, the Eyer Algorithm would submit a response to another of the remaining questions. This process continued until the team had provided answers for all 15 countries—5 were completed by the participant, 5 were completed by the Wible algorithm, and 5 were completed by the Eyer Algorithm. The two algorithms selected items to answer at random. To control for the number of correct answers viewed by participants across conditions, the answers provided by the algorithms were not visible to participants: as participants proceeded through the five rounds of the quiz, the algorithms' answers were masked by a series of six asterisks. While participants could not see the actual text, after each response by their teammates, participants were told that the algorithm's answer was correct. Participants' previous answers and their teammates' masked answers from previous rounds remained visible throughout the study, so that participants could see the answers accrue one by one (see Open Science Framework Stimuli and Materials at https://osf.io/mg3zn/ for a visual walkthrough of the procedure).

In the Alone condition, participants were told that they would answer quiz questions about country capitals with no mention of other team members. Participants were presented with the same 15-question country capital quiz and were instructed to answer any 5 of the 15 questions, one at a time, receiving feedback after each submission about whether or not their answer was correct. The difficulty of the available questions for participants to answer could differ between the Alone and Team condition, since the algorithms could answer items that participants would have selected. Since participants are more likely to select countries whose capital they know, this means that, on average, the items in the Team condition would be more difficult, which could lower predicted performance in that condition. However, since we predict those in the Team condition will provide higher estimates of future performance, the direction of this difference in difficulty works against our hypothesis, making our test a conservative one. All participants, regardless of condition, were asked not to use any outside help when answering quiz questions. Participants in both conditions viewed their total number of correct responses once the quiz was complete.

After the initial country capital quiz, participants predicted the percentage of questions they could answer correctly without outside help for 10 different topics (see Appendix B). The country capital quiz prediction question always appeared first, while the other nine items appeared subsequently in a randomized order. All 10 items were displayed together on a single page.

To test the accuracy of their predictions, participants next completed a new country capital quiz (see Appendix C). For this quiz, they were shown 15 new countries and asked to fill in as many of the capitals as they could. Participants completed a corporate logo quiz (Appendix D), for which they had predicted their performance ahead of time, and a periodic table quiz (Appendix E), for which they had not previously predicted their performance. The three quizzes appeared in a randomized order. If participants in the Team Condition are influenced by the specific content on their teammates' answers, then they should only be inaccurate when predicting performance on the country capital quiz. After each of the three follow-up quizzes, participants viewed the total number of questions they had answered correctly. Lastly, participants reported demographic information.

[1] Note, the fact that participants completed these studies online does not drive the effect—the same pattern of results hold in an in-lab replication study designed to ensure participants did not cheat on the trivia questions (see Supplementary Study 1 in the online supplemental materials).

Preregistration and Open Data

All preregistration forms, de-identified raw data files, and analytic syntax are available at Open Science Framework (https://osf.io/mg3zn/).

Results

First, we analyzed participants' predictions for their performance on the quiz matching the topic from the main task (country capitals). Although all participants attempted the same number of questions and saw the same number of correct answers, those who had been paired with successful teammates thought they would perform better on a future quiz, even without the aid of those teammates. Participants in the Team condition predicted they would perform better on future country capital quizzes, $M = 38.68\%$, $SD = 25.64\%$, than those in the Alone condition, $M = 29.67\%$, $SD = 22.22\%$, $t(299) = -3.25$, $p = .001$, Cohen's $d = -.37$ (95% CI = [−.60, −.15]). This effect occurred for predictions of the matched task (another country capital quiz), but not for the other nine topics. To account for the within-subject design, a linear mixed-effect model using the lme4 and lmerTest packages in R (Bates et al., 2015; Kuznetsova et al., 2015; R Development Core Team, 2013), predicted participants' ratings for performance on other quizzes based on condition. Random intercepts and slopes for subjects and items were also included in the model. This analysis revealed no significant differences in self-assessments based on condition, $b = .88$, $SE = 1.86$, $p = .64$.

We next analyzed participants' accuracy in their predictions. To so do, we first created a difference score by subtracting participants' actual performance on the follow-up quiz from their predicted performance on that topic (see Table 1). For the follow-up country capital quiz, participants in the Team condition overestimated their performance to a greater degree ($M = 11.75\%$, $SD = 25.90\%$) than those in the Alone condition did ($M = 5.23\%$, $SD = 19.48\%$), $t(299) = -2.64$, $p = .01$, Cohen's $d = -.28$ (95% CI = [−.51, −.06]), see Figure 1. However, for the corporate logo quiz, for which participants had not previously received outside help, participants were equally accurate in their predictions of their performance ($M_{\text{Team}} = 1.78\%$, $SD = 25.29\%$; $M_{\text{Alone}} = 1.55\%$, $SD = 25.58\%$), $t(299) = -.08$, $p = .94$. There was also no difference in participants' performance in the periodic table quiz, a topic for which they had not predicted performance ($M_{\text{Team}} = 26.10\%$, $SD = 16.87\%$; $M_{\text{Alone}} = 28.03\%$, $SD = 17.15\%$), $t(299) = .98$, $p = .33$. There was also no difference on the initial country capital quiz score ($M_{\text{Team}} = 3.96$, $SD = 1.24$; $M_{\text{Alone}} = 4.14$, $SD = 1.20$), $t(299) = 1.29$, $p = .20$.

Lastly, a linear regression predicted participants' estimates of their follow-up country capital quiz performance using condition, their score in the main task (out of five), and the demographic variables of age and gender. The effect of experimental condition remained significant when controlling for these other factors. Independently, those who answered more capitals correctly in the main task predicted higher performance in the follow-up quiz. Moreover, consistent with previous research (Jonsson & Allwood, 2003), we found an individual difference based on gender, where males predicted higher scores on the future quiz than females (see Table 2).

To ensure the findings were robust, and demonstrate that arbitrary features of the paradigm were not driving the effect, we replicated the study with some minor variations (see Supplementary Study 2). The replication differed from Study 1 in three ways. First, participants in the alone condition answered all 15 questions instead of only 5 (ensuring that the observed effects were not because some questions were left unanswered). Second, the answers given by the algorithms were visible (ensuring that the observed effects were not because answers were hidden in the team condition). Finally, the algorithms chose which questions to answer based on difficulty: in the difficult-question condition the algorithms answered the hardest remaining question, in the easy-question condition the algorithms answered the easiest remaining question, and in the random-question condition, they algorithm chose questions randomly, as in Study 1. Despite these differences, the results replicated. Moreover, the difficulty of questions that the algorithms answered played

Figure 1
Quiz Score Overestimations (Predicted Score–Predicted Score) by Condition in Study 1. Error Bars, Mean ± Standard Error

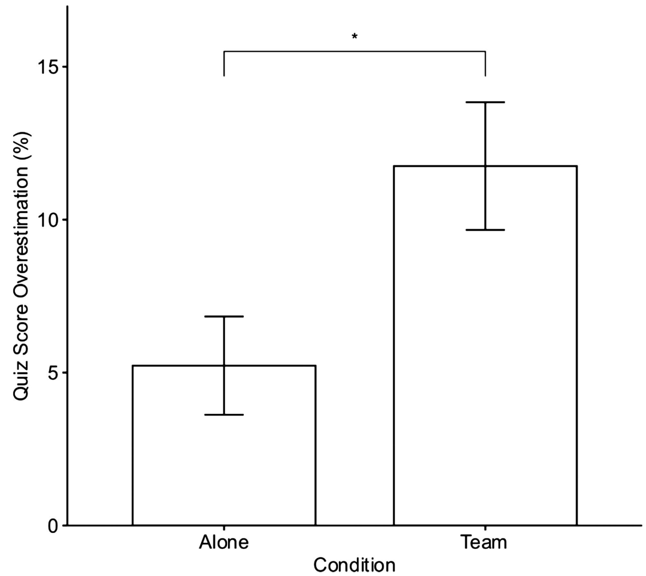


Table 1
Mean (SD) Predicted and Actual Quiz Performance in Study 1

	Predicted performance		Actual performance	
	Capital quiz	Logo quiz	Capital quiz	Logo quiz
Alone condition	29.67% (22.22%)	50.80% (26.61%)	24.44% (15.18%)	49.25% (13.50%)
Team condition	38.68% (25.64%)	52.00% (26.90%)	26.93% (19.19%)	50.22% (14.12%)

Table 2
Study 1 Linear Regression Results

Predictor	*b* (β)	*SE*
Intercept	5.08 (.02)	6.32 (.10)
Condition (team)	10.59 (.43)***	2.60 (.11)
Score	5.69 (.28)***	1.07 (.05)
Age	.18 (.08)	.11 (.05)
Gender (female)	−9.84 (−.40)***	2.68 (.11)

Note. $R^2 = .16$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

no role whatsoever on the strength of the effect (see Supplementary Study 2 for details of the methods and analysis). This demonstrated the reliability of the findings, and robustness to minor changes in testing procedure.

Study 2

Technological developments make receiving help from AI-agents, like those in the previous study, more common. At the same time, powerful communication technologies grant easier access to human knowledge—at any moment we can send a text message to a friend or video chat with a colleague. In Study 2, we explored the potential breadth of the previous results by testing a social transactive memory context. Will participants similarly inflate ratings of their own ability after being paired with human partners? If so, it would suggest metacognitive misattributions occur in a variety of domains: both technological and social.

Method

Participants

Four hundred fifty-one participants (180 males, 271 females; $M_{\text{AGE}} = 36.91$, $SD = 11.74$) from the United States completed the study online through Amazon Mechanical Turk.

Design and Procedure

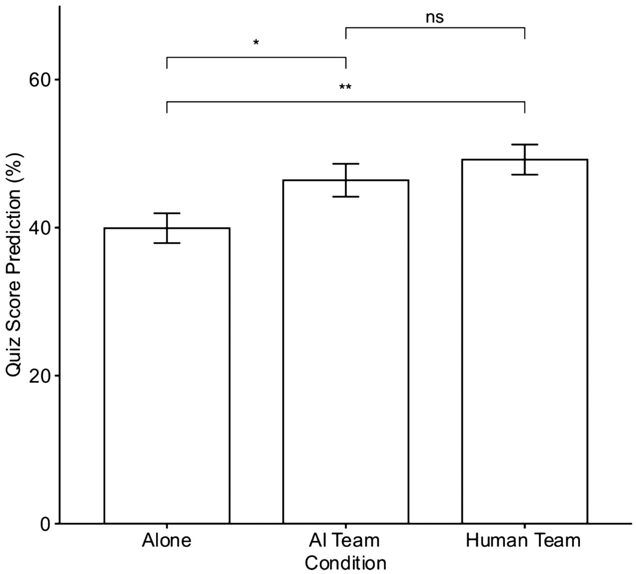
Study 2 used the same paradigm as the Study 1, but made several key changes. First, the Team condition was separated into two conditions: an AI Team condition and a Human Team condition. In the Human Team condition, participants were first instructed "You will be paired with the two other participants to complete the capital city quiz. You must work as a team to get the highest score you can!" If fact, participants were teamed with the same AI teammates as before but were under the impression that these were other participants replying in real time. To make the matching of teammates plausible, they were asked to "Please enter your first, middle (if applicable), and last initials for your username." Participants were then shown a loading icon with the following message, "You are being matched with your two teammates. This should take less than 15 s." After 5 s, participants were automatically advanced to the next page, where they were told, "You have been matched with the following teammates: Player 2 = MCF, Player 3 = DMO." Lastly, participants received these instructions: "You have been randomly selected to play as Player 1. All 3 players will input their answers simultaneously. When you choose one item to answer, MCF (Player 2) and DMO (Player 3) will be required to choose other items to answer. Once all 3 answers are submitted, you will see the results. You will not be able to see the answers of your teammates but will see whether they are correct or incorrect. This will continue for five total rounds until all the questions have been answered. Work as a team and get as many questions right as you can. Follow the directions at the top of each screen and do not use any outside help. Good luck!" The instructions about simultaneous input were included to make it plausible that the others' answers would appear so quickly. In order to make the conceit of the game dynamics plausible, before each round of the country capital quiz, participants were shown a list of unanswered countries and asked to select the country they would answer that round. After making their selection, participants continued to the input page, where they were instructed to "input the capital of the country you just selected." Like the Team conditions in Study 1, participants would then submit their answer, see if it was correct or incorrect, then click to see if their teammates' answers were correct. As in Study 1, all of the teammates' answers were masked by a series of six asterisks but were always correct.

Since the Human Team condition required additional elements in order for the cover story to be plausible, we altered the other two conditions to match it as closely as possible. At the beginning of the study, participants in the Alone and AI Team condition were shown the same loading screen as the Human Team condition but were shown the message "The capital quiz is loading, this should take less than 15 s." The instructions regarding simultaneous input were also included in the AI Team condition except the initials of the other players were replaced with the names of the algorithms. After completing the quiz, participants in all conditions were asked "If you were to take a different capital quiz on your own, what percent of the questions would you answer correctly without using any outside AIs?" [0–100 sliding scale]. Unlike Study 1, no other dependent measures were collected. Other than these changes, the Alone and AI Team conditions were identical to the Alone and Team conditions from Study 1.

Results

A linear regression using condition to predict future performance ratings found that participants in the Human Team ($M = 49.19\%$, $SD = 25.25\%$, $\beta = .29$, $SE = .11$, $p = .006$) and the AI Team condition ($M = 46.39\%$, $SD = 26.77\%$, $\beta = .24$, $SE = .11$, $p = .03$) gave higher ratings compared to those in the Alone condition ($M = 39.92\%$, $SD = 24.82\%$). Higher scores on the initial quiz corresponded with higher predictions of future performance ($\beta = .32$, $SE = .04$, $p < .001$) and males predicted higher performance than females ($\beta = .20$, $SE = .09$, $p = .03$; overall model fit: $R^2 = .13$). Independent sample *t*-tests showed the Alone condition predicted lower scores than the Human Team, $t(304) = -3.24$, $p = .001$, Cohen's $d = -.37$ (95% CI = [−.60, −.14]), and AI Team condition, $t(295) = -2.16$, $p = .03$, Cohen's $d = -.25$ (95% CI = [−.48, −.02]), while the predictions of the two Team conditions did not differ, $t(297) = -0.93$, $p = .35$ (see Figure 2). Furthermore, the scores on the initial quiz did not differ between the Alone condition ($M = 4.19$, $SD = 1.18$) and those in the Human Team ($M = 4.40$, $SD = .97$, $t(304) = -1.66$, $p = .10$) or AI Team condition ($M = 4.24$, $SD = .95$, $t(295) = .40$, $p = .69$); there was also no difference between the two Team conditions, $t(297) = -1.39$, $p = .17$. Note, that while follow-quiz scores were not measured in Study 2, participants in Study 1 scored

Figure 2
Quiz Score Prediction Ratings by Condition in Study 2. Error Bars, Mean ± Standard Error



an average of 25.71% on the follow-up country capital quiz, suggesting that participants in Study 2 are also overestimating their abilities. These results suggest that having help, whether human or AI, leads to a similar effect—a boost in people's assessments of their own abilities.

Study 3

One explanation for this effect is that merely seeing the task being successfully completed leads to a belief that the task is easy. If so, people would not be over-estimating their own knowledge, so much as under-estimating task difficulty because the team is doing so well. If this were the case, then passively observing AI agents complete the task (even without active human participation) would nonetheless invoke metacognitive miscalibration. In Study 3, we tested whether participants' active participation was necessary for the increase in self-assessed knowledge found in the previous studies.

Method

Participants

Two hundred ninety-nine (143 males, 156 females; M_{AGE} = 36.81, SD = 12.14) participants from the United States completed the study online through Amazon Mechanical Turk.

Design and Procedure

Like the previous studies, the main task in Study 3 consisted of a country capital quiz. Participants were randomly assigned to one of two conditions: the Participate condition or the Observe condition. In the Participate condition, participants were paired with the two algorithms to complete the 15-question quiz. This condition was identical to the Team condition from Study 1.

In the Observe condition, participants were instructed that they would be watching three algorithms complete a country capital quiz. This condition was identical to the Participate condition except that instead of participants answering one question in each of the five rounds, another algorithm (the Evans Algorithm) answered those questions. Like the other algorithms, the Evans Algorithm always provided a correct response. Thus, all participants in the Observe condition viewed the full set of 15 correct answers, as opposed to those in the Participate condition who would only see 10 correct answers, plus their own responses. If merely seeing more correct answers increases estimates of future performance, then this feature of the design provides a conservative test for the effect.

Results

Participants in the Participate condition predicted they would score higher on another country capital quiz (M = 37.13%, SD = 25.63%) than those in the Observe condition did (M = 29.80%, SD = 23.47%), $t(297) = 2.58$, $p = .01$, Cohen's $d = .30$ (95% CI = [.07, .53]), see Figure 3. Like the previous studies, this result indicates that simply seeing successful responses does not explain the results in the previous studies. When participants are not actively engaged in producing answers for the team, they are no longer likely to increase their ratings of their own abilites.[2]

Study 4a

To this point, we have found that when individuals receive help they overestimate their own abilities, regardless of whether the help comes from other humans or machines. Why might this be? One initial possibility is that they misattribute others' knowledge to themselves. Although the participants in our studies only answered (at most) five trivia questions correctly, their teammates successfully answered an additional ten out of ten questions. If participants conflate team performance with their personal performance, then their increased confidence could be attributable to the team success. If this account is accurate, we would expect participants' overconfidence to diminish or disappear when paired with less accurate teammates. That is, if participants confuse their own abilities with that of the team, then when paired with unreliable teammates it should reduce rather than increase their subsequent confidence in their abilities. Study 4a aimed to test this prediction.

Method

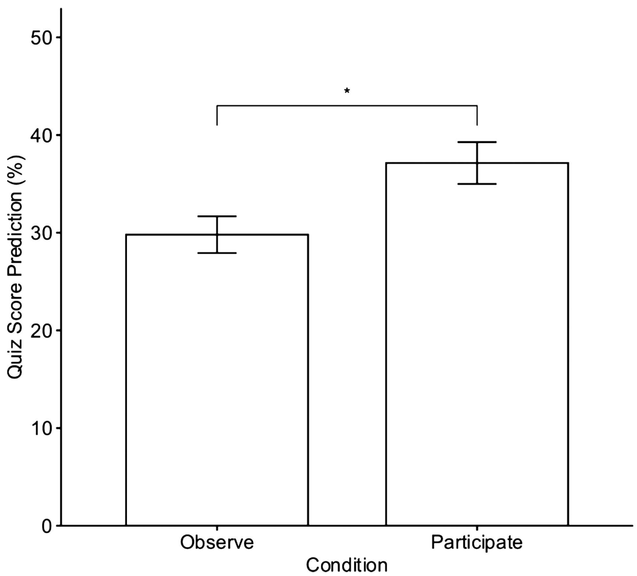
Participants

Anticipating participants from outside the United Sates would bypass Amazon's restrictions, we recruited more than the target participant total. A total of 545 participants (267 males, 278 females; M_{AGE} = 37.12, SD = 10.97) completed the study online through Amazon Mechanical Turk. Twenty-two participants could not be

[2] The psychology of why and how team membership and composition influences people is complex, nuanced, and hotly debated, answering the question of why membership is necessary remains outside the scope of the current set of studies, but remains a promising area for future research (see Charness et al., 2007; Mathieu et al., 2014).

Figure 3

Quiz Score Predictions by Condition in Study 3. Error Bars, Mean ± Standard Error



confirmed to be from the United States and per our preregistration, were not included in the analysis.

Design and Procedure

Participants in Study 4a were randomly assigned to one of three conditions. In the Reliable condition, participants were paired with 2 always correct algorithms to complete a 15-question country capital quiz (same as the AI Team condition in Study 2). In the Unreliable condition, participants completed the same country capital quiz, but one of the algorithms provided incorrect answers in rounds 1, 3, and 5 while the other algorithm provided incorrect answers in rounds 2, 3, and 4. Like Study 1, the text of the algorithms' answers was masked with asterisks, so participants could only see feedback stating if the algorithms' answers were correct or incorrect, but could not view the actual answer. Finally, in the Alone condition, participants answered 5 of the 15 questions by themselves on the same country capital quiz. After the quiz, all participants predicted how well they would perform on another capital quiz without using any outside sources. As preregistered, at the end of the study we asked all participants the following check question: "Did you use outside help to look up any of your own answers to the quiz questions? Please answer honestly." Twenty-six participants reported accessing outside materials and were thus not included in the analysis.[3]

Results

A linear regression using condition to predict future performance ratings found that those in the Reliable ($M = 43.57\%$, $SD = 26.22\%$, $\beta = .22$, $SE = .10$, $p = .03$) and the Unreliable condition ($M = 42.56\%$, $SD = 24.00\%$, $\beta = .18$, $SE = .10$, $p = .07$) provided higher ratings than those in the Alone condition ($M = 37.54\%$, $SD = 28.86\%$). Like previous studies, higher initial quiz scores corresponded with higher future performance ratings ($\beta = .40$, $SE = .04$, $p < .001$) and males predicted higher performance than females ($\beta = .30$, $SE = .08$, $p < .001$; overall model fit: $R^2 = .18$). Independent sample t-tests showed the Reliable condition provided higher ratings than the Alone condition, $t(331) = -2.00$, $p = .047$, Cohen's $d = -.22$ (95% CI = [−.44, −.003]), while the difference between the Unreliable and Alone condition approached significance, $t(326) = -1.71$, $p = .09$, Cohen's $d = -.19$ (95% CI = [−.41, .03]). The ratings of the Reliable and Unreliable conditions did not differ, $t(331) = 0.37$, $p = .71$ (see Figure 4). Like the previous studies, those in the Alone condition performed as well on the initial country capital quiz ($M = 3.92$, $SD = 1.44$) as those in the Reliable ($M = 3.94$, $SD = 1.16$, $t(331) = -.14$, $p = .89$) and Unreliable condition ($M = 3.97$, $SD = 1.11$, $t(326) = -.34$, $p = .73$); there was also no difference between the Reliable and Unreliable conditions, $t(331) = -.23$, $p = .82$. Overall, these results suggest that having help, regardless of the accuracy leads to overestimations of one's own abilities.

Study 4b

In contrast to our predictions, being on a team boosted confidence in future performance even when the teammates were not particularly accurate. If participants were conflating team performance with personal performance, then better performing teammates should lead to greater overestimation, however this was not the case in Study 4a.

Because the pattern of results was unexpected, and because some of the findings only approached significant, it is plausible this result was spurious. As such, we replicated this study with modifications to increase statistical power. At the time of conducting this study, there was a non-trivial uptick in poor quality responses on Amazon Mechanical Turk (Aruguete et al., 2019; Chmielewski & Kucker, 2019), so we increased the sample size and added several reading comprehension questions at the beginning of the study to screen out low quality participants (e.g., random clicking). Furthermore, while in Study 4a the teammates missed three of the five questions, in Study 4b the teammates missed all five questions, thus reducing team performance even further.

Finally, we added a new exploratory measure in which we asked participants to anticipate how well the algorithms would perform on future trivia questions. This assessed whether people's overestimation of their own performance would be reflected in their teammates' performance: do people's evaluations of all the team members increase as a function of team membership, or is this unique to the self? As we did not have a priori predictions about what would happen, we did not pre-register this question; it was exploratory in nature, and inferences should be tempered accordingly.

Method

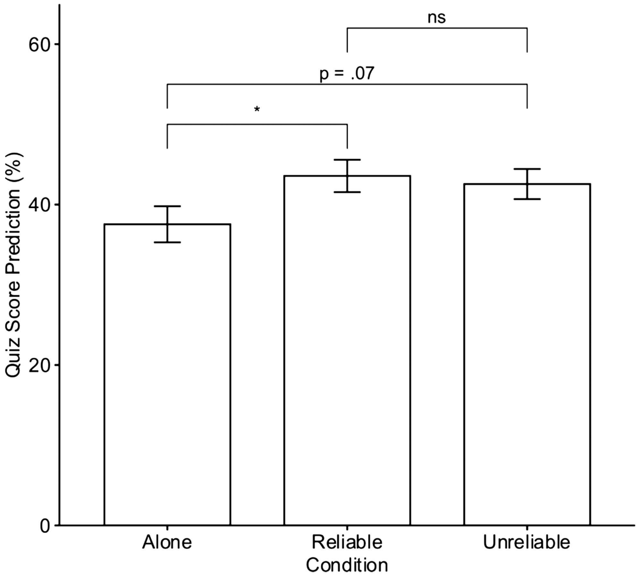
Participants

Six hundred participants (282 males, 318 females; $M_{AGE} = 38.09$, $SD = 12.56$) completed the study online through Amazon

[3] Note that even if the self-report screening did not detect all participants who used outside sources, as long as this behavior did not occur more often in one condition than the others, it would not affect the direction of the results.

Figure 4

Quiz Score Prediction Ratings by Condition in Study 4a. Error Bars, Mean ± Standard Error



Mechanical Turk. Eleven participants could not be confirmed to be from the United States and per our preregistration, were not included in the analysis. The reading comprehension questions at the beginning of the study screened out a total of 164 participants, and as per our preregistration were not included in the analysis.

Design and Procedure

Study 4b was identical to 4a except that Unreliable condition was replaced with the Incorrect condition. Instead of the algorithms incorrectly answering 6 out of the 10 items (as in the Unreliable condition of Study 4a), the algorithms answered all 10 incorrectly. The Reliable condition from Study 4a is identical to the Correct condition from Study 4b. Thirty-four participants reported accessing outside materials and as per the pre-registration were thus excluded from the analysis. Finally, we asked participants to rate the algorithms' future performance, on the same scale as predicted self-performance.

Results

A linear regression used experimental condition as a predictor of future quiz performance and found that those in the Incorrect (M = 46.60%, SD = 25.57%, β = .30, SE = .10, p = .002) and Correct conditions (M = 43.80%, SD = 24.97%, β = .20, SE = .10, p = .046) provided higher ratings than those in the Alone condition (M = 38.78%, SD = 25.25%). Number of correct answers during the quiz (β = .33, SE = .04, p < .001) and gender (β = .23, SE = .08, p = .005) were again independently significant predictors of future performance ratings (overall fit: R^2 = .14). Independent t-tests showed a significant difference between the Incorrect and Alone conditions, t(377) = −2.99, p = .003, Cohen's d = −.31 (95% CI = [−.51, −.10]), and the difference between the Correct and Alone conditions approached significance, t(358) = −1.90, p = .06, Cohen's d = −.20 (95% CI = [−.41, .01]). There was no difference between the Incorrect and Correct conditions, t(373) = −1.07, p = .29 (see Figure 5). Initial quiz scores did not differ between the Alone condition (M = 4.01, SD = 1.34) and those in the Correct (M = 3.97, SD = 1.16, t(358) = .29, p = .77) or Incorrect condition (M = 4.06, SD = .96, t(377) = −.42, p = .67); there was also no difference between the Correct and Incorrect conditions, t(373) = −.82, p = .41.

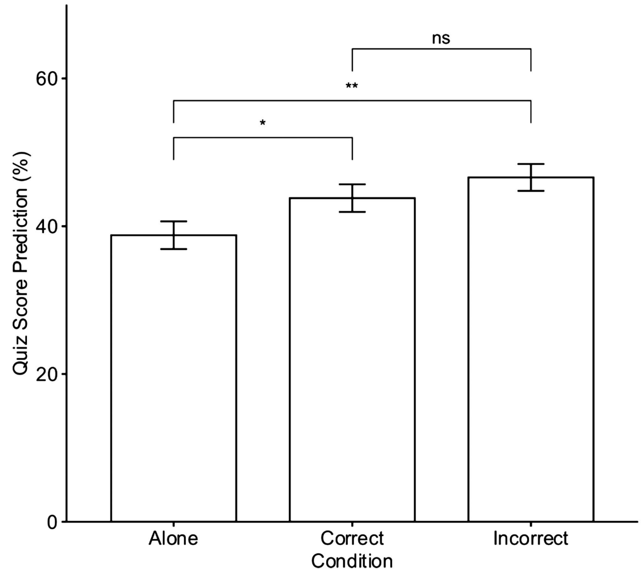
We also found that participants clearly distinguish between the correct and incorrect algorithms. When rating the algorithms' future performance, participants provide much higher rating for the correct algorithms (M = 89.12%, SD = 17.94%) than the incorrect algorithms (M = 12.49%, SD = 19.52%), t(373) = 43.75, p < .001, Cohen's d = 4.52 (95% CI = [4.14, 4.91]). First, this result serves as an attention check and demonstrates that participants were tracking their teammates performance. Second, ratings of future performance for one's teammates do not exhibit the same degree of overconfidence that occurred for self-ratings.

Discussion

Study 4b replicated and extended the findings from Study 4a—participants were more confident in their future performance after working with teammates, even when those teammates performed poorly. In fact, in Study 4b the teammates missed every single question, and participants still showed inflated perceptions of future self-performance. Studies 4a and 4b rule out the simple mechanistic explanation that people conflate the team's performance with their own.

Figure 5

Quiz Score Prediction Ratings by Condition in Study 4b. Error Bars, Mean ± Standard Error



Study 5

The previous results, especially Study 4b, rule out the most straightforward accounts of overconfidence. Participants do not simply confuse the team's performance for their own. One possible explanation is that being on a team opens up ambiguity about one's contribution to success. When working alone, there is no uncertainty about one's own knowledge. However, when a person's teammate answers (or fails to answer) a question, there is uncertainty about whether that person would have answered correctly in the absence of the teammate. This allows people to engage in biased and self-enhancing interpretations to resolve the uncertainty (Kunda, 1987, 1990). For example, when the algorithms increase the team's score, people can attribute the team's success to themselves, engaging in a credit-taking bias ("the reason our score was so high was because of me, I would do just as well by myself," Kruger & Gilovich, 1999; Ross & Sicoly, 1979). However, when the algorithms perform poorly, people can attribute the team's failure to the algorithms, and thus overestimate performance ("The reason our score was so low was because of my teammates, I would do better by myself").

To test this explanation, we created a version of the country capital quiz which made the participants' contribution unambiguous. After completing the country capital quiz, participants were asked to provide their own answers for the countries selected by their teammates (whose original answers were masked by asterisks). If contribution ambiguity is driving the results, participants who answer their teammates' items will not overestimate future performance while participants who never attempt to answer their teammates' items will show the previously observed overconfidence.

Method

Participants

Four hundred four participants (205 males, 199 females; $M_{\text{AGE}} = 38.28$, $SD = 21.74$) completed the study online through Amazon Mechanical Turk. Two participants could not be confirmed to be from the United States, 21 participants self-reported using outside sources to answer quiz questions, and 28 participants self-reported low effort on the quizzes. Per our preregistration, these participants were not included in the analysis.

Design and Procedure

In Study 5, participants were randomly assigned to either the Ambiguity or the No Ambiguity condition. The Ambiguity condition was identical to the Team condition from Study 1, where participants never attempt to answer the items completed by the AI agents. In the No Ambiguity condition, after the fifth and final round of the quiz and seeing the overall team score, participants were shown a list of the 10 countries for which their teammates had submitted masked answers. Participants were asked to name the capital for those 10 countries. After submitting their answers, they were told how many they answered correctly. Like the previous studies, participants next rated what percent of questions they would answer correctly on another country capital quiz using different countries than previously seen.

Following the procedure from Study 1, participants predicted their performance for quizzes on nine other topics. Finally, participants completed three follow-up quizzes, one additional country capital quiz, one corporate logo quiz (for which they had predicted performance), and one periodic table of the elements quiz (for which they had not predicted performance). These additional measures allowed us to determine the accuracy of participates' self-assessments.

Results

Participants in the No Ambiguity condition had significantly lower predictions of their performance on an additional country capital quiz ($M = 37.27\%$, $SD = 23.86\%$) than those in the Ambiguity condition ($M = 42.91\%$, $SD = 24.67\%$) $t(351) = 2.18$, $p = .03$, Cohen's $d = .23$ (95% CI = [.02, .44]). Participants' estimates on the other quiz topics were not influenced by condition, $b = 1.33$, $SE = 1.70$, $p = .44$. The initial country capital quiz scores did not differ between those in the Ambiguity ($M = 4.42$, $SD = 1.62$) and No Ambiguity conditions ($M = 4.28$, $SD = 1.35$), $t(351) = 0.86$, $p = .39$.

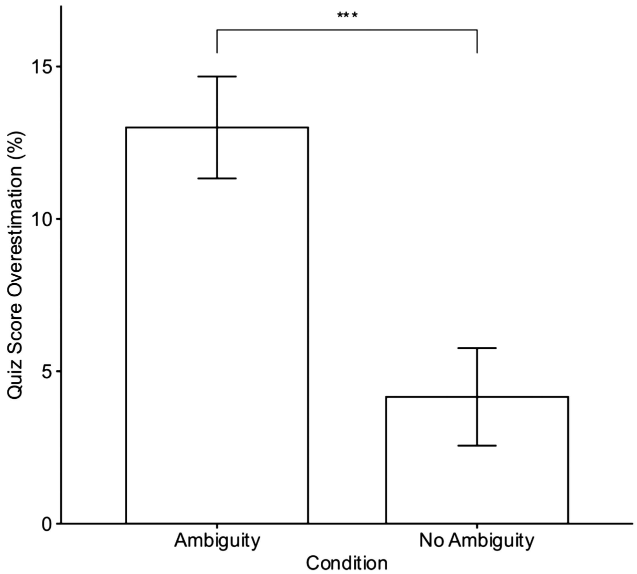
Critically, taking into account actual performance on the follow-up country capital quiz, those in the Ambiguity condition overestimated their scores ($M = 13.00\%$, $SD = 22.12\%$) more than those in the No Ambiguity condition ($M = 4.16\%$, $SD = 21.35\%$), $t(351) = 3.82$, $p < .001$, Cohen's $d = .41$ (95% CI = [.20, .62]), see Figure 6. For the corporate logos, which participants had not received outside help, but did make predictions, the quiz scores did not differ by condition, $t(351) = -.57$, $p = .57$. This suggests that simply making predictions does not explain participants overconfidence for the country capitals.[4]

Finally, a linear regression predicted participants' overconfidence using condition, their score on the initial quiz, and the demographic variables of age and gender. The effect of experimental condition remained significant ($\beta = -.40$, $SE = .10$, $p < .001$) when controlling for these other factors. Quiz scores ($\beta = .01$, $SE = .05$, $p = .84$) and age ($\beta = .07$, $SE = .05$, $p = .14$) did not significantly predict overconfidence. However, gender independently predicted overconfidence for the follow-up quiz ($\beta = -.30$, $SE = .10$, $p = .005$; overall fit: $R^2 = .07$).

These results illustrate how ambiguity regarding credit-taking leads to overconfidence. Note, these results cannot be explained simply by the number of questions participants answered. As shown in Supplementary Study 2, even when participants answer three times as many questions, they still predict lower future performance. Instead, when participants can plausibly take credit for team success, they do. These results help interpret the results from Study 3. In Study 3, participants never tested their own ability when observing performance, but critically, they are also not a member of the team. This outside perspective prevents "taking credit" for the correct responses of the AI agents.

[4] To test for baseline knowledge differences we examined the scores of the periodic table quiz, a topic for which participants had not made quiz score predictions. There was a small difference in performance, but it was the No Ambiguity condition that answered more correctly ($M = 33.90\%$, $SD = 19.47\%$) than the Ambiguity condition ($M = 29.18\%$, $SD = 19.05\%$), $t(351) = -2.30$, $p = .02$, Cohen's $d = -.24$ (95% CI = [−.45, −.03]). This difference in performance suggests that the overestimation by those in the Ambiguity condition may be a conservative estimate, because they performed worse on an independent topic, but did not show a difference in predicted scores for other independent topics.

Figure 6
Quiz Score Overestimations (Predicted Score–Predicted Score) by Condition in Study 5. Error Bars, Mean ± Standard Error



General Discussion

Using a new experimental paradigm, we showed that being engaged in a transactive memory system leads to increased assessments of one's own abilities. Study 1 demonstrated that people believe they know more when algorithms assist them and examined the accuracy of this heightened confidence, showing that participants who received assistance overestimated performance. Study 2 extended these results to show that a similar effect occurs when receiving help from other people. Study 3 demonstrated that passively observing the algorithms performing a task was not enough to elicit this miscalibration—participants needed to be actively contributing to the team in order to confuse team performance for their own. Studies 4a and 4b, showed that this tendency toward overconfidence persisted even when the teammates were not helpful, or even harmful. Study 5 showed how performance ambiguity in the team context enables people to interpret the team's outcomes in a self-serving manner and become overconfident in their own abilities.

These studies reaffirm earlier work that utilizing outsourced knowledge leads people to overestimate what they know. Here we extend the previous work on search engines (Fisher et al., 2015) to show that the illusion of knowledge generalizes to other technological tools and even social contexts. In the current paradigm, participants did not actively seek outside help and at times the help was completely unhelpful; yet we still observed boosts in confidence. Previous research had focused on domains with well-established and familiar sources of assistance and knowledge, but the current set of studies suggests outsourced knowledge illusions may be much more widespread, occurring in many applied contexts. Furthermore, we offer initial evidence that performance ambiguity, not source confusion, is the cause of miscalibration within a transactive memory network.

Participants exploit ambiguity to take credit for team success and assign blame for team failure. This aligns with our findings from Supplementary Study 2: the difficulty of the questions that the algorithms correctly answer did not affect the bias. When the algorithms answer the hard questions, people attribute the success to themselves and thus overestimate their ability to answer hard questions. When the algorithms answer the easy questions, people think "I would have gotten those, so these bots did not help me," and thus overestimate performance. Further, when the algorithms provide little or no assistance (Study 4), participants still inflate their estimates of their own ability. This could be viewed as a manifestation of the headwinds/tailwinds effect (Davidai & Gilovich, 2016), in which people are more sensitive to impediments to performance than facilitators; people fail to notice the wind at their backs speeding them up, but more readily notice the wind in their face slowing them down. Similarly, when the algorithms are particularly helpful people may take that assistance for granted, but nonetheless blame poorly performing algorithms for a lack of team success.

These results provide a potential explanation for egocentric biases in team contexts (Kruger, & Savitsky, 2009; Ross & Sicoly, 1979). People frequently overclaim the amount they have personally contributed to a group effort—a dynamic that could lead to a preference for working alone. Based on the current results, this could occur because when a team performs well, people attribute the success to themselves, increasing their estimate of how well they would perform alone. However, when a team performs poorly, people can blame the teammates for the failure, again, increasing their estimate of how well they would perform alone.

Overconfidence

Do these studies demonstrate overconfidence or could participants' increased self-assessments after working with a team be accurate? In all of our studies, the key dependent variable asked participants to estimate how well they would perform on a country capital quiz. In order to test for overconfidence, participants in Study 1 and Study 5 completed a follow-up country capital quiz. The results of these two studies clearly demonstrate overconfidence. In all the other studies, no follow-up quiz was included, but participants made estimates for the same question: they rated how well they would perform on a country capital quiz. As reported, the accuracy rates on the initial quiz are comparable across studies, suggesting that participants have comparable knowledge of geography. The consistency in methods helps interpret the results of these of other studies. If each study involved different areas of competence, for which actual performance was uncertain, it would be less clear if participants were overconfident. However, one major advantage of the similarity between our studies is that it gives us strong reason to believe that participants are consistently rating their abilities too high.

Thus, the present findings extend our understanding of overconfidence, a topic of core interest to decision scholars (Block & Harper, 1991; Fischhoff, 1975; Fischhoff et al., 1977; Hoffrage, 2004; Moore & Healy, 2008; Soll & Klayman, 2004). While previous work has focused on overconfidence in a vacuum, people are often embedded in complex networks where they can offload cognition to tools, the environment, and other people. We have found that being part of a transactive memory system leads to greater overconfidence than when people are studied in isolation. In their definitive literature review on overconfidence effects, Moore and Healy (2008) break down overconfidence into its component parts: overestimation (overrating one's actual ability to perform a task), overplacement (overrating one's ability relative to others), and overprecision (being overly certain about the accuracy of one's beliefs). While in the context of this project we deal exclusively with overestimation, future research could explore the other types of overconfidence. In particular, it is unclear how being a member of a transactive memory system might lead to overplacement relative to other elements within that transactive memory system versus elements outside the system. While we find consistent overestimation of performance, it is worth noting that previous research has found predictable underestimation when performance is high (e.g., Burson et al., 2006). This pattern has been explained by regression to the mean (Erev et al., 1994; Moore & Healy, 2008)—when one's abilities are uncertain, error will naturally be regressive. Of course, working in a transactive memory system likely entails greater uncertainty; it is harder to know one's ability in isolation when one has only experienced a task in the context of a transactive memory system. Indeed, as Koriat's (1997) cue-utilization model suggests, when teammates or technology are responsible for the retrieval of information, it creates a paucity of internal-mnemonic cues, making the situation more ambiguous. To the extent that this is true, it may be that for particularly easy tasks, getting help exacerbates such under-confidence effects. This may be a useful topic for future investigation.

Applications

The misattribution of competence could be quite pervasive. In daily life, people are constantly relying on external aids (Intons-Peterson & Fournier, 1986)—many of which could lead to metacognitive miscalibration. This can happen in timeless ways like coordinating with others and using written text to externalize cognition (e.g., reminders). However, recent technologies accelerate our ability to rely on the tools in our environment. Instead of needing to be physically present to coordinate with others, we can now do so anywhere, anytime. Further, we can now receive the assistance of technologies with unprecedented capabilities. Consider the abilities enabled by just a single company, Google: 3-D computer models to help navigate (Google Maps), machine learning to support communication (Google Translate), unlimited storage to aid episodic memory (Google Photos), and intricate algorithms help find information (Google Search). As these technologies become increasingly interwoven to our daily lives, it is important to consider the unintended consequences they may produce. Indeed, some of these consequences already been documented: for example, relying on GPS maps leads to worse spatial knowledge (Hejtmánek et al., 2018; Ishikawa et al., 2008).

The current set of studies suggest our reliance on outside technologies prevent us from testing our own abilities. How well do I know a foreign language? How much of my vacation do I remember? These questions become more difficult to answer when we can access tools that automatically fill in the gaps in our knowledge. The relative contribution of one's own abilities vs. the help from the technology remains ambiguous. While the tools we rely on are beneficial in many respects, they can foster overconfidence in our own abilities. In cases where the technology is always available, the negative consequences of this illusion may not be experienced. However, when it is necessary to perform a task without any outside help (e.g., an exam, job interview), people may have difficulty assessing their true competence.

Future Directions

Transactive memory systems can develop over long periods of time. In romantic partnerships, relational intimacy improves joint recall—those who speak in terms of "we" instead of "I" remember better together than alone (Harris et al., 2011). Perhaps additional experience dissolves boundaries in a transactive network, making the assessment of one's own ability even more difficult. Alternatively, sustained reliance could create an acute awareness of one's own limitations. The influence of experience on metacognitive accuracy remains a topic for future research.

Another important task for future research is determining the features of cognition-augmenting tools that predict illusions of competence. In some ways, the teammates used in the current studies are quite different than Internet search (e.g., limited usefulness, non-interactive), but in other ways, they are quite similar (e.g., fast, helpful). Future work could continue to explore other sorts of technology and map the predictors of metacognitive inaccuracy.

While in the context of our studies the increased confidence engendered by receiving assistance leads to miscalibration, it is worth noting that there may exist other contexts in which this is not the case. For example, in some educational contexts, collaborative learning may not only increase confidence, but also increase subsequent performance (e.g., Laal & Ghodsi, 2012), and in such circumstances, confidence boosts may actually increase accuracy. Future work should explore calibration in collaborative contexts which support learning to determine how such dynamics play out.

As technologies that enhance our own cognitive capacities and our ability to seek help from others continue to gain prevalence, understanding what we know and do not know about their effects on human cognition is increasingly important. To the extent that we use augmenting technologies to help us understand the effect of augmenting technologies on cognition, we may think we know more about the topic than we actually do.

References

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied*, *17*(1), 18–32. https://doi.org/10.1037/a0022086

Alba, J. W., & Hutchinson, J. W. (2000). Knowledge calibration: What consumers know and what they think they know. *Journal of Consumer Research*, *27*(2), 123–156. https://doi.org/10.1086/314317

Aruguete, M. S., Huynh, H., Browne, B. L., Jurs, B., Flint, E., & McCutcheon, L. E. (2019). How serious is the 'carelessness' problem on Mechanical Turk? *International Journal of Social Research Methodology*, *22*(5), 441–449. https://doi.org/10.1080/13645579.2018.1563966

Barrett, L. (2011). *Beyond the brain: How body and environment shape animal and human minds*. Princeton University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *"lme4: Linear mixed-effects models using Eigen and S4," R package version 1.1-12*. https://CRAN.R-project.org/package=lme4

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*(4), 343–375. https://doi.org/10.1207/s15327663jcp0804_01

Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. *Organizational Behavior and Human Decision Processes*, *49*(2), 188–207. https://doi.org/10.1016/0749-5978(91)90048-X

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 3–5. https://doi.org/10.1177/1745691610393980

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, *90*(1), 60–77. https://doi.org/10.1037/0022-3514.90.1.60

Camerer, C. F., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review*, *89*(1), 306–318. https://doi.org/10.1257/aer.89.1.306

Charness, G., Rigotti, L., & Rustichini, A. (2007). Individual behavior and group membership. *The American Economic Review*, *97*(4), 1340–1352. https://doi.org/10.1257/aer.97.4.1340

Chmielewski, M., & Kucker, S. C. (2019). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological & Personality Science*, *11*(4), 464–473. https://doi.org/10.1177/1948550619875149

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195333213.001.0001

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19. https://doi.org/10.1093/analys/58.1.7

Davidai, S., & Gilovich, T. (2016). The headwinds/tailwinds asymmetry: An availability bias in assessments of barriers and blessings. *Journal of Personality and Social Psychology*, *111*(6), 835–851. https://doi.org/10.1037/pspa0000066

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over-and under confidence: The role of error in judgment processes. *Psychological Review*, *101*(3), 519–527. https://doi.org/10.1037/0033-295X.101.3.519

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288–299. https://doi.org/10.1037/0096-1523.1.3.288

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552–564. https://doi.org/10.1037/0096-1523.3.4.552

Fisher, M. (2021). *Who knows what? Knowledge misattribution in the Division of Cognitive Labor*. osf.io/mg3zn

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*(3), 674–687. https://doi.org/10.1037/xge0000070

Fisher, M., Smiley, A. H., & Grillo, T. L. H. (2021). Information without knowledge: The effects of Internet search on learning. *Memory*. https://doi.org/10.1080/09658211.2021.1882501

Gagnon, L. M., & Dixon, R. A. (2008). Remembering and retelling stories in individual and collaborative contexts. *Applied Cognitive Psychology*, *22*(9), 1275–1297. https://doi.org/10.1002/acp.1437

Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy*, *82*(7), 335–349. https://doi.org/10.2307/2026523

Hargis, M. B. & Oppenheimer, D. M. (2020). The promise and pitfalls of memory offloading: Investigating metacognitive awareness of augmented cognition. Manuscript under review.

Harris, C. B., Barnier, A. J., Sutton, J., & Keil, P. G. (2014). Couples as socially distributed cognitive systems: Remembering in everyday social and material contexts. *Memory Studies*, *7*(3), 285–297. https://doi.org/10.1177/1750698014530619

Harris, C. B., Keil, P. G., Sutton, J., Barnier, A. J., & McIlwain, D. J. (2011). We remember, we forget: Collaborative remembering in older couples. *Discourse Processes*, *48*(4), 267–303. https://doi.org/10.1080/0163853X.2010.541854

Harris, J. E. (1980). Memory aids people use: Two interview studies. *Memory & Cognition*, *8*(1), 31–38. https://doi.org/10.3758/BF03197549

Hejtmánek, L., Oravcová, I., Motýl, J., Horáček, J., & Fajnerová, I. (2018). Spatial knowledge impairment after GPS guided navigation: Eye-tracking study in a virtual town. *International Journal of Human-Computer Studies*, *116*, 15–24. https://doi.org/10.1016/j.ijhcs.2018.04.006

Henkel, L. A. (2014). Point-and-shoot memories: The influence of taking photos on memory for a museum tour. *Psychological Science*, *25*(2), 396–402. https://doi.org/10.1177/0956797613504438

Hoffrage, U. (2004). Overconfidence. In R. F. Pohl (Ed.), *Cognitive illusions: Fallacies and biases in thinking, judgment, and memory* (pp. 235–254). Psychology Press.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

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*(3), 267–280. https://doi.org/10.1037/0096-3445.115.3.267

Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology*, *28*(1), 74–82. https://doi.org/10.1016/j.jenvp.2007.09.002

Jonsson, A. C., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences*, *34*(4), 559–574. https://doi.org/10.1016/S0191-8869(02)00028-4

Keil, F. C., Stein, C., Webb, L., Billings, V. D., & Rozenblit, L. (2008). Discerning the division of cognitive labor: An emerging understanding of how knowledge is clustered in other minds. *Cognitive Science*, *32*(2), 259–300. https://doi.org/10.1080/03640210701863339

Kitcher, P. (1990). The division of cognitive labor. *The Journal of Philosophy*, *87*, 5–22. https://doi.org/10.2307/2026796

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Kruger, J., & Gilovich, T. (1999). "Naive cynicism" in everyday theories of responsibility assessment: On biased assumptions of bias. *Journal of Personality and Social Psychology*, *76*, 743–753. https://doi.org/10.1037/0022-3514.76.5.743

Kruger, J., & Savitsky, K. (2009). On the genesis of inflated (and deflated) judgments of responsibility. *Organizational Behavior and Human Decision Processes*, *108*(1), 143–152. https://doi.org/10.1016/j.obhdp.2008.06.002

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, *53*(4), 636–647. https://doi.org/10.1037/0022-3514.53.4.636

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *"Package 'lmerTest,'" R package version 2.0*.

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia: Social and Behavioral Sciences*, *31*, 486–490. https://doi.org/10.1016/j.sbspro.2011.12.091

Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development*, *73*(4), 1073–1084. https://doi.org/10.1111/1467-8624.00458

Mathieu, J. E., Tannenbaum, S. I., Donsbach, J. S., & Alliger, G. M. (2014). A review and integration of team composition models: Moving toward a dynamic and temporal framework. *Journal of Management*, *40*(1), 130–160. https://doi.org/10.1177/0149206313503014

Meade, M. L., Nokes, T. J., & Morrow, D. G. (2009). Expertise promotes facilitation on a collaborative memory task. *Memory*, *17*(1), 39–48. https://doi.org/10.1080/09658210802524240

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Park, S., & Budescu, D. V. (2015). Aggregating multiple probability intervals to improve calibration. *Judgment and Decision Making*, *10*(2), 130–143.

Pfeifer, P. E. (1994). Are we overconfident in the belief that probability forecasters are overconfident? *Organizational Behavior and Human Decision Processes*, *58*(2), 203–213. https://doi.org/10.1006/obhd.1994.1034

R Development Core Team. (2013). *R: A language and environment for statistical computing*.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, *37*(3), 322–336. https://doi.org/10.1037/0022-3514.37.3.322

Sloman, S. A., & Rabb, N. (2016). Your understanding is my understanding: Evidence for a community of knowledge. *Psychological Science*, *27*(11), 1451–1460. https://doi.org/10.1177/0956797616662271

Smith, A. (1776). *Inquiry into the nature and causes of the wealth of nations*. W. Strahan & T. Cadell. https://doi.org/10.1093/oseo/instance.00043218

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299–314. https://doi.org/10.1037/0278-7393.30.2.299

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

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

Vortac, O. U., Edwards, M. B., Fuller, D. K., & Manning, C. A. (1993). Automation and cognition in air traffic control: An empirical investigation. *Applied Cognitive Psychology*, *7*(7), 631–651. https://doi.org/10.1002/acp.2350070707

Ward, A. F. (2013). *One with the cloud: Why people mistake the Internet's knowledge for their own*. Harvard University.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). Springer. https://doi.org/10.1007/978-1-4612-4634-3_9

Wegner, D. M., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, *61*(6), 923–929. https://doi.org/10.1037/0022-3514.61.6.923

Appendix A

Capital Quiz From Studies 1–5

Country	Capital	Sporcle % correct	Difficulty
Spain	Madrid	91.0	Easy
Vietnam	Hanoi	71.1	Intermediate
Marshall Islands	Majuro	36.1	Difficult
Mexico	Mexico City	92.4	Easy
Latvia	Riga	71.8	Intermediate
Tonga	Nuku'alofa	36.2	Difficult
England	London	93.0	Easy
Ecuador	Quito	72.3	Intermediate
Solomon Islands	Honiara	37.3	Difficult
France	Paris	93.1	Easy
Venezuela	Caracas	72.3	Intermediate
Myanmar	Naypyidaw	38.6	Difficult
United States	Washington, DC	94.4	Easy
Kenya	Nairobi	72.4	Intermediate
Burundi	Bujumbura	42.1	Difficult

Note. The order of the countries always matched the order listed above. Retrieved September 12, 2017.

Appendix B

Dependent Measures From Studies 1 and 5

If you were to take a quiz about the following topics on your own, what percent of the questions would you answer correctly without using any outside sources?

- Country Capitals (different countries than previously seen) [0–100]
- US Presidents [0–100]
- Corporate Logos [0–100]
- NFL Teams [0–100]

(*Appendices continue*)

- Body Parts [0–100]
- Pokemon [0–100]
- Harry Potter [0–100]
- English Vocabulary [0–100]
- Pixar Movies [0–100]
- Math [0–100]

Appendix C

Capital Quiz 2 From Studies 1 and 5

Country	Capital	Sporcle % correct	Difficulty
Russia	Moscow	91.0	Easy
Uruguay	Montevideo	72.17	Intermediate
Gambia	Banjul	42.5	Difficult
Germany	Berlin	90.6	Easy
Bolivia	La Paz	72.9	Intermediate
Niger	Niamey	43.0	Difficult
Japan	Tokyo	90.5	Easy
Guatemala	Guatemala City	70.6	Intermediate
Central African Republic	Bangui	43.1	Difficult
Italy	Rome	89.8	Easy
Croatia	Zagreb	69.9	Intermediate
Malawi	Lilongwe	43.4	Difficult
Ireland	Dublin	89.6	Easy
Bosnia and Herzegovina	Sarajevo	69.6	Intermediate
Chad	N'Djamena	43.8	Difficult

Note. The order of the countries always matched the order listed above.

Appendix D

Periodic Table Quiz From Studies 1 and 5

Symbol	Element	Sporcle % correct	Difficulty
H	Hydrogen	94.6	Easy
Se	Selenium	52.3	Intermediate
Ds	Darmstadtium	25.8	Difficult
He	Helium	91.9	Easy
W	Tungsten	51.5	Intermediate
Gd	Gadolinium	25.2	Difficult
O	Oxygen	91.8	Easy
Ga	Gallium	50.6	Intermediate
Dy	Dysprosium	24.5	Difficult
C	Carbon	88.3	Easy
V	Vanadium	49.9	Intermediate
Mt	Meitnerium	24.3	Difficult
N	Nitrogen	85.8	Easy
Rb	Rubidium	48.9	Intermediate
Rg	Roentgenium	24.3	Difficult

Note. The order of the elements always matched the order listed above.

(*Appendices continue*)

Appendix E

Corporate Logo Quiz From Studies 1 and 5

Corporation	Sporcle % correct	Difficulty
Nike	98.7	Easy
Dream Works	74.7	Intermediate
Guinness	40.1	Difficult
Shell	97.0	Easy
American Airlines	71.6	Intermediate
Marriott	39.2	Difficult
Facebook	96.6	Easy
John Deere	69.7	Intermediate
Safeway	31.0	Difficult
Apple	96.1	Easy
Mazda	69.1	Intermediate
Anheuser-Busch	27.4	Difficult
McDonald's	95.8	Easy
Atari	64.4	Intermediate
Boeing	14.6	Difficult

Note. The order of the logos always matched the order listed above.

Received May 6, 2020
Revision received January 8, 2021
Accepted January 9, 2021 ■