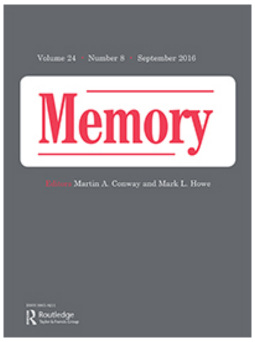

Memory

ISSN: 0965-8211 (Print) 1464-0686 (Online) Journal homepage: http://www.tandfonline.com/loi/pmem20

Conflict and metacognitive control: the mismatch-monitoring hypothesis of how others' knowledge states affect recall

Scott H. Fraundorf & Aaron S. Benjamin

MEMORY, 2016
VOL. 24, NO. 8, 1108–1122
http://dx.doi.org/10.1080/09658211.2015.1069853

Routledge
Taylor & Francis Group

Conflict and metacognitive control: the *mismatch-monitoring* hypothesis of how others' knowledge states affect recall

Scott H. Fraundorf[a] and Aaron S. Benjamin[b]

[a]Department of Psychology and Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260, USA;
[b]Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA

ABSTRACT

Information about others' success in remembering is frequently available. For example, students taking an exam may assess its difficulty by monitoring when others turn in their exams. In two experiments, we investigated how rememberers use this information to guide recall. Participants studied paired associates, some semantically related (and thus easier to retrieve) and some unrelated (and thus harder). During a subsequent cued recall test, participants viewed fictive information about an opponent's accuracy on each item. In Experiment 1, participants responded to each cue once before seeing the opponent's performance and once afterwards. Participants reconsidered their responses least often when the opponent's accuracy *matched* the item difficulty (easy items the opponent recalled, hard items the opponent forgot) and most often when the opponent's accuracy and the item difficulty *mismatched*. When participants responded only after seeing the opponent's performance (Experiment 2), the same *mismatch* conditions that led to reconsideration even produced superior recall. These results suggest that rememberers monitor whether others' knowledge states accord or conflict with their own experience, and that this information shifts how they interrogate their memory and what they recall.

ARTICLE HISTORY
Received 6 April 2015
Accepted 1 July 2015

KEYWORDS
Metacognition; metamemory; recall; cued recall; collaborative inhibition

A common scenario students face—and perhaps dread—is finding that their experience taking an exam does not accord with their peers'. Quickly finishing an exam that other students are still labouring over can induce a fear of having overlooked an important step to a problem, while finishing after other students may engender worries about one's general mastery of the material.

Such experiences suggest that others' mnemonic performance, and perhaps its similarity or dissimilarity to our own recall performance, are used in monitoring and guiding interactions with one's own memory—for instance, by suggesting whether or not one's initial guess or strategy is likely to be correct, or whether a particular exam item is likely to be a "trick" question. Researchers have frequently investigated how rememberers make metacognitive judgements about their *own* knowledge and use these judgements to control their study and retrieval (Benjamin, 2005, 2008; Fiechter, Benjamin, & Unsworth, in press; Finley, Tullis, & Benjamin, 2010; Nelson & Narens, 1990). Far less work has examined how rememberers incorporate *others*' knowledge states, such as whether or not someone else could recall a particular piece of information, into this process (but see Tullis & Benjamin, in press). Yet such information is commonly available: a student may hear that course material is easy or difficult to remember, a patient may be cautioned that a medicine regimen is complex to administer, a tourist asking for directions may be told that a route is simple or hard to follow, and an eyewitness may discuss with others whether or not their memories of an event were clear (Gabbert, Memon, & Wright, 2007; Wright, Memon, Skagerberg, & Gabbert, 2011).

In the present study, we contrasted two hypotheses about how rememberers attend to and use information about others' mnemonic performance. One possibility is that individuals primarily seek information about the general difficulty of memoranda and use that information to modulate the effort applied to retrieval. For instance, one might invest more care and effort into querying one's memory about items incorrectly remembered by others, regardless of whether those items seemed easy or challenging to oneself. Alternatively, rememberers may be more concerned with whether others' mnemonic states accord or conflict with their own, as the presence or absence of a match state can inform whether the strategies and products of one's own recall are likely to be accurate. For example, if an item appears difficult, but others answered it correctly (a mismatch), a search for simpler answers may be appropriate. These two hypotheses differ critically in what rememberers are thought to attend to: Do they attend to the external indicator of

CONTACT Scott H. Fraundorf scottfraundorf@gmail.com

difficulty itself, or do they attend to the interaction between that information and their own assessment of an item? Across two experiments, we evaluated these competing hypotheses on two measures: how information about another person's performance changes *what* participants report and whether those reports are more *accurate*.

Metamnemonic control of recall

What people report when asked to recall is not simply a product of what comes to mind; it also reflects metacognitive monitoring and control processes (Goldsmith & Koriat, 2008; Koriat, Goldsmith, & Halamish, 2008). For instance, learners have at least some metamnemonic awareness of the accuracy of their memory, allowing them to report items that are more likely to be correct and withhold those less likely to be correct (Koriat & Goldsmith, 1996). In addition to this control of *whether* to make a mnemonic report, learners appear able to control *how* to perform a memory search: cued recall performance increases when the test reminds the learner of the strength of each cue–target relationship (Higham & Tam, 2005) or of the study strategy with which the cue–target pair was encoded (Halamish, Goldsmith, & Jacoby, 2012). These results suggesting that information about the target allows learners to interrogate their memory in a different, more productive way.

Metacognitive control of recall could be further informed by awareness of *others*' knowledge states. For instance, knowing that a problem is usually answered accurately could imply that a complex strategy is unnecessary. Conversely, knowing that a seemingly easy exam question is frequently missed might suggest that it is a difficult problem for which a rapidly retrieved response might not be correct, and knowing that other eyewitness disagreed about the appearance of a particular person (e.g., a crime suspect) implies the person does not have any highly distinctive features. As noted above, indicators of others' memory performance are likely a source of information that is frequently available when monitoring recall.

Social influences on memory

Indeed, there is evidence that people use others' memory decisions to guide their own responding (Harris, Paterson, & Kemp, 2008; Rajaram, 2011). This can have both benefits and costs. Experience remembering in a group can enhance one's later memory as an individual (Basden, Basden, & Henry, 2000; Rajaram & Pereira-Pasarin, 2007). But, it can also introduce false memories (Roediger, Meade, & Bergman, 2001), and people working in a group typically recall less information than the sum of what the members could have recalled on their own (*collaborative inhibition*; e.g., Weldon & Bellinger, 1997).

The bulk of this work, however, has focused on how rememberers are influenced by the specific *products* of others' retrieval, such as knowing specifically how another eyewitness described a crime suspect's appearance. Far less work has considered how rememberers consider others' retrieval *processes*, such as knowing whether the other eyewitness could or could not recall what the perpetrator looked like. Such information might engender more beneficial social influences on memory. First, it would be less likely to introduce specific false memories: learning only *whether* others recalled something and not specifically *what* they output prevents any erroneous details in others' output from contaminating one's own memory. Second, it might be less disruptive to one's own retrieval strategies: one account of why remembering as a group often impairs memory is that exposure to the products of others' recall disrupts one's own retrieval similar to part-list cuing effects (Wright & Klumpp, 2004); however, this disruption would not occur if rememberers are not exposed to the specific items that others recalled. Thus, it is likely that there could be different, and perhaps more beneficial, effects when rememberers learn about others' mnemonic processes (e.g., whether or not someone remembered a particular detail, or whether it was easy or hard for them to remember) without being presented with specifically what they recalled. To date, however, few studies have examined the effects of information about others' mnemonic processes, and there is a need for such investigations given that rememberers often have access to such information.

Even knowledge about others' performance might not influence recall, however, if rememberers' initial recall attempts were already as comprehensive and accurate as possible. But, there is evidence that memory can improve across repeated retrieval attempts (*hypermnesia*; Erdelyi, 2010; Payne, 1987), especially when subsequent retrieval attempts are made in a different manner. For example, recalling a story from a different perspective contributes information not recalled in the first attempt (Anderson & Pichert, 1978). And, a second response to a general knowledge question typically samples different information; thus, averaging the two estimates often improves judgement accuracy, especially when the second estimate is dissimilar to the first (Fraundorf & Benjamin, 2014; Herzog & Hertwig, 2009; Hourihan & Benjamin, 2010; Vul & Pashler, 2008). The benefits of additional retrieval attempts suggest that information about others' memory performance could indeed improve recall if it motivates further consideration of an item, especially if this further processing is different from what participants would otherwise perform.

How does knowledge about others influence recall?

It is likely, then, that awareness of others' knowledge states could influence and even benefit memory retrieval. But what aspects of this information do individuals attend to, and how is it incorporated into metacognitive monitoring? In the present study, we compare two contrasting predictions—motivated by theories of metacognitive control in other domains—about which aspects of others' mnemonic performance guide recall.

One possibility, which we term the *difficulty-cue hypothesis*, is that the effect of external metacognitive information is to increase retrieval effort whenever there is a sign of apparent difficulty. One common account (e.g., Thiede & Dunlosky, 1999) of metamnemonic decisions is that learners devote more effort and attention to more difficult items. Supporting this claim, learners perform better when expecting a difficult test (Balota & Neely, 1980) and often spend more time studying more difficult items (Son & Metcalfe, 2000; Tullis & Benjamin, 2011). Analogously, participants might use another individual's memory performance to assess whether an item will present an easy or difficult retrieval experience and then reconsider difficult items. For example, knowing that an exam was difficult for others might cause a student to devote more effort and care to taking it, regardless of how she would have performed otherwise. This hypothesis predicts that information about others' performance exerts a main effect on recall strategy and success, independent of one's own experience retrieving the item. Some evidence does suggest that participants respond to others' general level of mnemonic performance: Reysen (2003) found that participants in a free recall task recalled more items when other, fictive "group members" also recalled more.

But, signs of difficulty may only matter to metacognitive control to the extent that they disagree with rememberers' own assessment of a memory task. This possibility is suggested by the literature on error monitoring, which indicates that, across a variety of cognitive tasks, the degree of conflict is an important cue to when greater cognitive control should be implemented (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Under what we term the *mismatch-monitoring hypothesis*, what rememberers monitor is not others' knowledge states themselves, but whether those states affirm or conflict with what rememberers already bring to bear on retrieval: metacognitive awareness of their own knowledge state. When external information accords with one's own retrieval experience and assessment of the memorandum, it provides additional evidence supporting one's approach to the task or problem and discourages reconsideration. By contrast, conflict between internal and external information suggests that one's own assessment or solution strategy might be inaccurate and needs changing. Continuing the test-taking example, a student taking a seemingly difficult test may feel reassured when other students also struggle, but may feel compelled to adopt a new approach if other test-takers appear to be completing the test easily. Conversely, a student who initially assesses a test as easy may reconsider if other students appear to struggle with it. The mismatch-monitoring hypothesis predicts external information to interact with one's own assessment of an item: external information encourages reconsideration when it mismatches one's own perception and discourages it when it accords with existing perceptions. Analogous interactions have been observed in some other studies. Fazio and Marsh (2009) tested participants' memory for perceptual details of feedback received in answering world knowledge questions. Feedback was best remembered when it conflicted with participants' self-reported confidence in their response—feedback that a high confidence response was wrong (see also Butterfield & Metcalfe, 2001) or that a low confidence guess was correct—and least remembered when it matched participants' experience. In general, however, it is unclear whether the influence of others' performance is qualified by its match to participants' own retrieval processes because item difficulty is rarely manipulated independently of others' performance.

Finally, of course, a third possibility is that people simply do not attend to others' mnemonic performance or do not incorporate it into their metacognitive control.

Overview of experiments

We pit these hypotheses against each other in two experiments in which participants studied paired associates for a later cued recall task and, during the test, received fictive information about another person's performance on each item. Before the task began, participants read a cover story informing them that they would compete against a previous participant who had seen some of the same paired associates. During the subsequent test phase, each cue was accompanied by information that the opponent had either responded correctly to the cue, responded incorrectly, or did not see the pair. (That is, the information provided was the *accuracy* of the fictive opponent's memory, not the specific word purportedly typed in by the opponent.) This information was never based on an actual previous participant, but was instead experimentally manipulated.

As reviewed above, one way participants might use information about the opponent is to consider whether it matches their own assessment of the item. To test this hypothesis, we included a second manipulation to influence participants' *own* study and recall: cue relatedness. The degree of semantic relatedness between a cue word on a target word is a strong influence not only on actual memory but on metacognition: prior work has consistently found that semantically related word pairs are both objectively remembered better and subjectively judged as better learned (Dunlosky & Matvey, 2001; Koriat, 1997). Thus, differing levels of cue relatedness should give rise to different experiences at test. Words pairs that are moderately associated should present a relatively easy retrieval experience for participants. For example, candidate answers might come to mind more quickly for these items, participants might be more certain about the candidate answers, and participants might even explicitly classify them as *easy* (or *related*) items. Normatively unrelated word pairs, by contrast, should on average present a more difficult retrieval experience and may be explicitly categorised by participants as *difficult* or *unrelated*. (Analyses of participants' accuracy in both experiments, presented below, confirmed that the unrelated pairs were indeed more difficult for participants to retrieve.) An advantage of using a fictive opponent is that this manipulation of cue relatedness can be implemented fully independently of others' apparent mnemonic performance.

Crucially, the difficulty-cue and mismatch-monitoring hypotheses make contrasting predictions about the role of participants' own metacognitive state in consideration of the opponent's performance. The difficulty-cue hypothesis proposes that participants increase their retrieval effort whenever others demonstrate difficulty.[1] Thus, seeing that the opponent made an error should have a main effect that is qualitatively similar across levels of cue relatedness. In contrast, the mismatch-monitoring hypothesis predicts that what is important is not whether the opponent made an error per se, but whether the opponent's performance matches the metacognitive state brought on by one's own processing of the item. This hypothesis predicts a disordinal interaction: the opponent's performance should *most* motivate a change in retrieval strategy when *inconsistent* with one's own perception of the item (seemingly easy-to-retrieve items that the opponent answered incorrectly and seemingly hard-to-retrieve items that the opponent answered correctly) and *least* motivate a change when *consistent* with one's perception (seemingly easy items the opponent answered correctly and seemingly hard items the opponent answered incorrectly).

We considered the effects of the fictive opponent's performance and cue relatedness on two aspects of responding. In Experiment 1, we explicitly tested whether information about others' memory leads participants to reconsider their retrieval strategy. Participants responded once to each item before seeing the fictive opponent's performance and once afterwards; the crucial measure was whether participants *switched* their response after viewing the opponent's performance. Having established that such information indeed influences the probability of reconsideration, we then turned in Experiment 2 to a task tailored to assessing whether those changes in retrieval behaviour could affect mnemonic *accuracy*.

Experiment 1

In Experiment 1, we investigated whether seeing the opponent's performance on a particular item would lead participants to reconsider the item and change their response on a subsequent second attempt, which we term a *switch*. In particular, we compared whether the opponent's performance would have a main effect on the rate of switching, as predicted by the difficulty-cue hypothesis, or would interact with a manipulation likely to affect the participant's own ease of retrieval (i.e., cue relatedness), as predicted by the mismatch-monitoring hypothesis.

Method

Participants

In both experiments, participants were students at the University of Illinois between the ages of 18 and 30 participating in partial fulfilment of a course requirement. Twenty-four individuals participated in Experiment 1.

Materials

Seventy-two target English nouns were selected from the MRC Psycholinguistic Database (Coltheart, 1981). Each target was paired with an associated cue to create a *related* condition, such as *EXPERT-DOCTOR*. Related cues had a forward associative strength of .02 to .03 with their targets and 0 with all other words in the experiment. Cues and targets were then randomly re-paired to create an *unrelated* condition for each target, such as *VINE-DOCTOR*.

These items were divided into two study-test cycles of 36 items each. Participants completed 1 study-test cycle immediately followed by the other study-test cycle; thus, each participant was tested on a total of 72 items. Within each cycle, an equal number of pairs were related and unrelated, and an equal number of each of those were designated as items that the fictive opponent got *correct*, got *incorrect*, or *didn't see*, yielding a 3 (opponent's performance) $\times$ 2 (cue relatedness) design. The assignment of items to conditions was counterbalanced across participants using a Latin Square design.

Procedure

We conducted the experiment on a computer running MATLAB and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

A cover story first established the fictive opponent. Participants read that they would either try to exceed the performance of a previous participant or set a score that a later participant would try to beat. A *Please wait …* screen with an hourglass icon then appeared for 5000 ms. Finally, all participants were informed they were *Partner B* and would compete against a previous participant. We used the competition frame for two reasons. First, participants working collaboratively in a group may feel compelled to maintain group cohesion by conforming to others' overall level of mnemonic performance (Reysen, 2003),

[1] Another variant of the difficulty-cue hypothesis is that, faced with external signs of difficulty, remembers would "give up" and not even attempt difficult items (Son & Metcalfe, 2000). Although inconsistent with evidence that learners often devote more effort to items perceived as difficult (Balota & Neely, 1980; Son & Metcalfe, 2000; Tullis & Benjamin, 2011), this version of the difficulty-cue hypothesis also predicts a main effect of the external cue. Thus, it still differs from the mismatch-monitoring hypothesis, which critically predicts a disordinal interaction between the external cue and one's own retrieval experience.

but conformity should be less likely to mask participants' true level of recall when the goal is to outperform others. Second, it provides a cover for displaying a fictive individual's memory performance (i.e., tracking the performance of the opponent against whom the participant is competing) with lesser demand characteristics than would come from explicitly instructing participants to use this information. Participants were told that they had the option of answering differently in their second response, but could also just type the same response as before. (Indeed, as will be seen, participants switched responses on only a minority of trials, indicating that any demand characteristics to switch responses were not overwhelming.)

During the study phase, cue–target pairs appeared one at a time in a randomised order for 4000 ms each, with a 750 ms inter-stimulus interval between pairs. No information about the opponent's performance appeared during study.

Participants were then tested on each item in a re-randomised order. For each item, the cue was displayed and the participant typed an initial response before seeing any information about the opponent's performance. Participants were instructed that if they did not know an item, they could leave it blank; that is, a free-report procedure was used. After the participant entered an initial response, the opponent's performance appeared in the centre of the screen. This information was either *Your opponent GOT THIS PAIR CORRECT* (in green text), *Your opponent GOT THIS PAIR INCORRECT* (in red), or *Your opponent DID NOT SEE THIS PAIR* (in gray). After a 1250 ms delay, participants typed a second, final response while the rest of the display remained on-screen. Participants had to retype the response even if it was unchanged. After participants entered the second response, the next test item was presented. No feedback was provided about the accuracy of either response. There was a 750 ms interval between test items.

At the end of the entire experiment (i.e., after the conclusion of the second study-test cycle), participants were explicitly instructed that some pairs consisted of two related words and some of two unrelated words. (No reference to the distinction between related pairs and unrelated pairs was presented to the participant prior to this point.) Two manipulation checks were then presented. First, participants typed numbers from 0 to 100 to rate what percentage of the related pairs and of the unrelated pairs they thought that they remembered. These responses were used to obtain further evidence, beyond the prior literature (Dunlosky & Matvey, 2001; Koriat, 1997), that participants subjectively viewed the unrelated items as more difficult (in addition to differences in objective accuracy). Second, participants rated on the same scale what percentage of each pair type they thought that the opponent remembered. This phase was used to confirm that participants expected the opponent would remember more related pairs than unrelated pairs.

Finally, a structured debriefing assessed whether participants suspected the opponent did not actually exist. Only one participant gave any such response; this participant was included in the analyses reported here, but excluding this participant did not qualitatively change any results.

Results

Analytic strategy

The proportion of trials on which participants switched responses—the primary measure of interest in Experiment 1—was generally low. Low proportions pose a problem for analysis of variance (ANOVA) models, which assume a normally distributed dependent variable, because proportions, especially those far from .5, are not normally distributed. Consequently, we analysed the data using multi-level logit models (Baayen, Davidson, & Bates, 2008; Jaeger, 2008), which avoid this issue because they are not based on a proportion computed over trials within a particular cell of the design (Jaeger, 2008). Rather, these models analyse the log odds of a particular response at the level of individual trials (e.g., whether the participant did or did not switch responses on a particular trial). Multi-level models can also incorporate multiple random effects, variables for which the observed categories were sampled from a larger population. Analyses of human memory typically consider only one random effect: participants. However, experimental items are also randomly sampled from a population of possible items (Clark, 1973), which can vary substantially in their memorability (Freeman, Heathcote, Chalmers, & Hockley, 2010), and multi-level models can simultaneously account for both participant and item variability. All models were fit by Laplace estimation using the *glmer()* function of the *lme4* R package (Bates, Maechler, & Bolker, 2011).

Manipulation checks

In the manipulation check phase, participants rated themselves as remembering more related word pairs ($M =$ 59.33%, SD = 27.50) than unrelated pairs ($M =$ 36.96%, SD = 23.68); this difference was reliable, $t_{(23)} = 5.95$, $p < .01$, 95% confidence interval of the difference: [14.60; 30.16]. Combined with the differences between pair types in objective accuracy, reported below, this result indicates that our manipulation of participants' own recall was successful: As expected, the unrelated pairs produced a retrieval experience that was on average both objectively and subjectively more difficult for participants.

Further, participants also judged the opponent as recalling more related ($M =$ 53.33% recalled, SD = 21.55) than unrelated word pairs ($M =$ 42.92%, SD = 21.76); this difference was also reliable, $t_{(23)} = 2.90$, $p < .01$, 95% confidence interval: [3.00; 17.84]. This confirms that participants indeed (falsely) believed that the opponent would be more likely to recall related pairs and could potentially treat the opponent's performance as diagnostic of item difficulty.

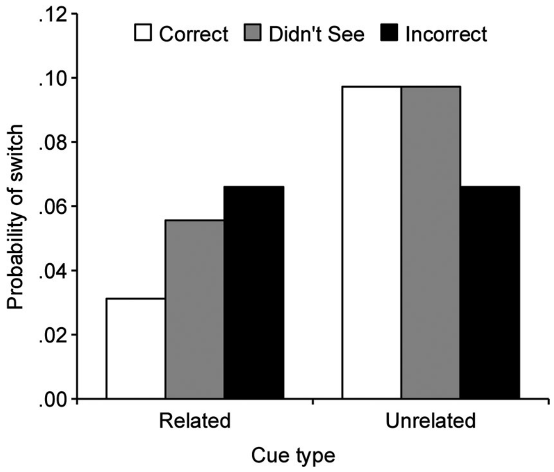

Figure 1. Probability of switched responses in Experiment 1 as a function of cue relatedness and opponent's performance.

Switch rate

Figure 1 displays the mean rate at which participants switched responses[2] as a function of cue relatedness and opponent's performance. Note that, at the time participants had to decide whether to switch their responses, they had not received any feedback explicitly indicating the correctness of the first answer (indeed, they *never* received such feedback), and, more generally, could not be certain whether their first response attempt was correct. Thus, although participants presumably would not want to switch from a correct initial response, participants sometimes switched even when their first response was correct. Consequently, we include all trials, including both accurate and inaccurate first responses, in the switch rates presented in Figure 1.

We modelled the odds of switching using a multi-level logit model. The model included participants and target words as random effects to account for variability among participants in how often they switched and variability among items in how often they elicited switching. The fixed effects of theoretical interest (cue relatedness, opponent's performance, and their interaction) were coded using mean-centred effects coding, which corresponds to the main effects and interactions analysed in an ANOVA (Cohen, Cohen, West, & Aiken, 2002). The three types of information about the opponent were compared using two contrasts that represent planned comparisons. The first contrast compared switching given *Opponent got this pair correct* feedback to the baseline rate of switching (i.e., the mean across all three conditions); the second compared the *Opponent got this pair incorrect* condition to the mean.

Table 1. Fixed-effect estimates (top) and variance estimates (bottom) for multi-level logit model of response switches in Experiment 1 (model N = 1728, log-likelihood: −380.4).

Fixed effect	β	SE	Wald z	p
Intercept	−3.26	0.29	−11.06	<.01
Related cue	−0.70	0.21	−3.23	<.01
Opponent was correct	−0.42	0.32	−1.33	.18
Opponent was incorrect	0.08	0.29	0.26	.80
Related cue × opponent correct	−1.47	0.64	−2.31	<.05
Related cue × opponent incorrect	1.40	0.58	2.43	<.05

Note: SE = standard error.

In a multi-level model, variability in an effect (e.g., the strength of the cue relatedness effect) across participants or across items can be modelled with a random slope of that effect by participants or by items. However, no random slopes for any of the effects improved the fit of the model in likelihood ratio tests (all $ps > .40$). Consequently, we report the model with only random intercepts.

Table 1 displays fixed-effect parameter estimates[3] from the model. Cue relatedness had a main effect on switching: the odds of a switch were 2.02 times greater (95% CI: [1.33, 3.05]) for unrelated pairs (M = 8.68%) than for related pairs (M = 4.98%).

Crucially, the opponent's performance did not have a reliable main effect on the switch rate. Participants did not simply switch whenever they saw the opponent made an error. Instead, opponent's performance and cue relatedness reliably interacted: participants switched their answers less in conditions where the opponent's performance was likely to match their own retrieval experience and switched more in conditions where the opponent's performance was likely to mismatch their experience.

Specifically, feedback that the opponent answered an item correctly *discouraged* switches for easy items but *encouraged* it for hard items. For related (easy-to-retrieve) items, participants switched responses less when the opponent purportedly answered the item correctly (M = 2.78%) as compared to the average rate among all related pairs (M = 4.97%). But for unrelated (hard-to-retrieve) items, participants switched *more* when the opponent answered correctly (M = 9.72%) as compared to the average for all unrelated pairs (M = 8.68%). This interaction is statistically reflected in an odds ratio: the odds of switching when given *Correct* feedback were 4.34 times higher (95% CI: [1.25, 15.09]) with an unrelated cue than with a related cue.

Feedback that the opponent answered incorrectly had the opposite effect, encouraging switches for *easy* items and discouraging it for *hard* items. For related, easy items, participants switched more when the opponent answered the item incorrectly (M = 6.60%) as compared to the baseline (M = 4.97%); for unrelated, hard items, participants switched less when the opponent answered incorrectly

[2] Because Experiment 1 used a free-report procedure in which participants did not have to enter a response to every cue, switches could be either between an omission and a reported answer or between two different reports. Although we did not have a large enough sample of each subtype of switch to conduct appropriate inferential statistics on individual subtypes, the pattern of means within each subtype was qualitatively the same as the pattern reported in the text for all switches combined. In Experiment 2, we turned to a forced-report procedure, which required participants to respond to every cue, and established that the information about the opponent's performance had effects that are not limited to how often participants make a report.

[3] To facilitate interpretation, we report point estimates and confidence intervals around the *odds* by transforming them from the *log-odds* model parameters.

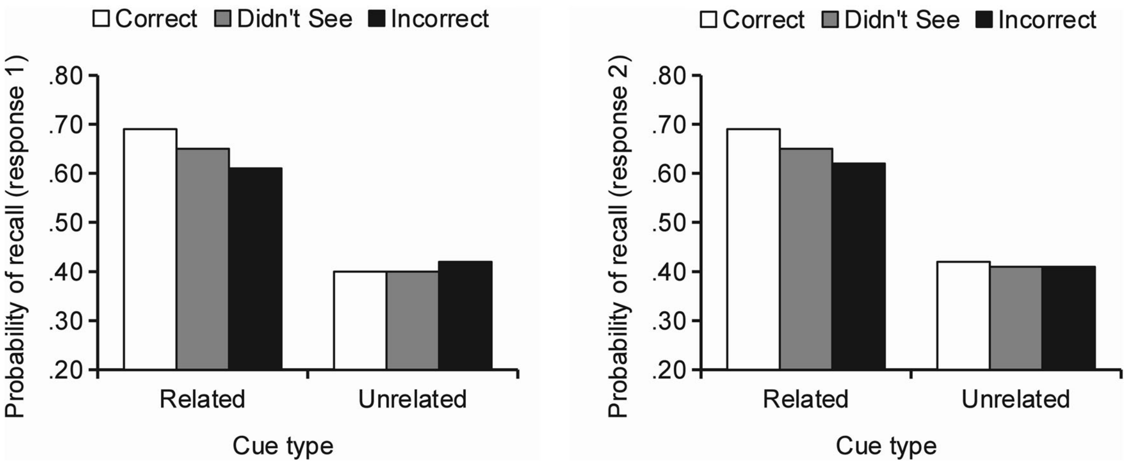

Figure 2. Cued recall accuracy in Experiment 1 for first responses (left panel) and second responses (right panel) as a function of cue relatedness and opponent's performance.

($M = 6.60\%$) as compared to baseline ($M = 8.68\%$). The odds ratio indicated that the odds of switching given *Incorrect* feedback were 4.07 times greater (95% CI: [1.31, 12.62]) with a related cue as opposed to an unrelated cue.

Final accuracy

The primary purpose of Experiment 1 was investigating response switches as a measure of participants' searching and reporting behaviour. However, each response (first and second) can also be evaluated for whether it was the correct target for the cue. Table 2 displays the mean accuracy of the first and second responses. Accuracy did not approach ceiling for either response.

Of primary interest was the second response,[4] made after feedback about the opponent, which provides a measure of whether the information about the opponent influenced the *accuracy* of recall. The right panel of Figure 2 displays mean accuracy of the second, final response in Experiment 1. (For reference, the left panel displays the accuracy of the first response, made before seeing feedback about the opponent.) We fit a multi-level logit model to the accuracy of the final response. For this model, a random slope of cue relatedness improved the fit of the model, $\chi^2_{(2)} = 7.24$, $p < .05$, indicating that relatedness benefited some participants more than others, and was incorporated into the final model. Other random slopes did not improve model fit ($ps > .30$) and were omitted.

Table 3 displays parameter estimates for the final model. The odds of correct recall were 3.40 times greater (95% CI: [2.47, 4.67]) given a related ($M = 65.28\%$) than an unrelated cue ($M = 41.32\%$), confirming that related pairs on average presented an easier retrieval experience. Accuracy did not approach ceiling across all items ($M = 53.30\%$), nor within just the comparatively easy related items ($M =$ 65.28%), thus permitting ample opportunity for the switches to benefit or impair recall accuracy. However, the opponent's performance had no main effect on accuracy, nor did it interact with cue relatedness.

Table 2. Mean accuracy of first responses and second responses in Experiment 1.

	First Responses		Second Responses	
Measure	*M* (%)	*SE*	*M* (%)	*SE*
Correct responses	52.66	4.32	53.30	4.30
Incorrect responses (total)	47.34	4.32	46.70	4.30
Omissions	26.22	3.85	28.13	4.03
Inaccurate guesses	21.12	4.00	18.58	3.26

Note: SE = standard error.

Table 3. Fixed-effect estimates for multi-level logit model of final cued recall accuracy in Experiment 1 (model $N = 1728$, log-likelihood = −1015).

Fixed effect	*β*	*SE*	Wald *z*	*p*
Intercept	0.20	0.22	0.87	.38
Related cue	1.22	0.16	7.51	<.01
Opponent was correct	0.23	0.16	1.45	.15
Opponent was incorrect	−0.19	0.16	−1.21	.23
Related cue vs. opponent correct	0.38	0.31	1.21	.23
Related cue vs. opponent incorrect	−0.37	0.31	−1.19	.23

Note: SE = standard error.

Discussion

In Experiment 1, participants made one cued recall response and then saw fictive information about an opponent's accuracy in recalling the item in question. We contrasted two hypotheses about how this information might lead participants to change their answer in a second recall response. The difficulty-cue hypothesis was not supported: there was no main effect that suggested

[4] Another potential dependent measure is participants' initial recall accuracy. However, no effect of the feedback about the opponent's performance would be expected in this measure because the first attempt was made prior to seeing information about the opponent's performance on that trial. Indeed, a model of initial recall accuracy revealed only a main effect of cued relatedness, $p < .001$, and no reliable effects involving the opponent's performance.

that an erroneous response by another person would lead participants to reconsider any item.

Rather, consistent with the mismatch-monitoring hypothesis, external information interacted with item difficulty. We expected related word pairs to be relatively easy for participants to retrieve; this expectation was supported by measures both of objective difficulty (retrieval accuracy) and of subjective difficulty (self-reported performance). Thus, for these easy items, an error by the opponent would typically conflict with the participants' own perception of the item and was observed to increase the probability that participants switched their response. Conversely, a correct response by the opponent on these pairs generally accorded with the participant's own metacognitive perception and decreased the probability of a switch. Unrelated pairs, by contrast, were typically more difficult to retrieve; for these items, a correct response by the opponent (a mismatch) produced more switches, and an incorrect response (a match) produced fewer switches. That is, participants appeared to consider whether another individual's mnemonic process matched or mismatched their own.

This pattern also suggests the results of Experiment 1 were not an artefact of the competitive nature of the task. In principle, seeing the opponent err might have increased or decreased participants' overall motivation to respond correctly based on the opportunity to take the lead in the purported competition. The disordinal interaction we observed renders such possibilities unlikely. The opponent's performance did not have any main effect on behaviour but instead interacted with item difficulty, favouring a metacognitive rather than a motivational explanation.

Although the pattern of switches across conditions was that predicted by the mismatch-monitoring hypothesis, the overall rate of switches was somewhat low. Low overall rates of response switching have also been observed in other experiments (Van Zandt & Maldonado-Molina, 2004). However, differences across conditions were large relative to the overall switch rates: for example, for related pairs, switch rates more than *doubled* if participants were told the opponent answered incorrectly than if the opponent answered correctly. These relations are captured by the logit analysis, which accounts for the fact that the means and variances of proportions are related (i.e., a five percentage point change is relatively larger when the initial percentage is low; Jaeger, 2008).

Thus, the evidence suggests that participants reconsidered their response more in mismatch conditions and less in match conditions. One possible mechanism for this effect is that participants who chose to reconsider searched their memories using more cues or different cues (or differently applied those cues to filter existing retrieval products). For example, if a participant easily generated one response to *EXPERT*-______ but then saw that the opponent got the pair incorrect, the participant might retrieve other, additional types of experts or reconsider those types of experts that the participant had retrieved but dismissed. Conversely, a participant who confidently retrieved a response to *EXPERT*-______ and then saw that the opponent also answered the item correctly would have little reason to consider other possible responses. This explanation is consistent with the finding that making a second, different judgement can tap knowledge that would otherwise go unused (Van Zandt & Maldonado-Molina, 2004; Vul & Pashler, 2008) and that cued recall can be facilitated by metacognitive control over which cues are submitted to memory (Finley & Benjamin, 2012; Halamish, Goldsmith, & Jacoby, 2012; Higham & Tam, 2005).

In other paradigms, a second recall attempt dissimilar to the first can benefit performance (Anderson & Pichert, 1978; Fraundorf & Benjamin, 2014; Herzog & Hertwig, 2009; Hourihan & Benjamin, 2010; Vul & Pashler, 2008). Thus, conditions that motivate reconsideration might improve recall—and conversely, conditions that discourage reconsideration might impair it. Although the opponent's performance was not observed to have a robust effect on final accuracy in Experiment 1, there was other evidence that reconsideration was a useful strategy. We coded each response switch as *beneficial* if it was from an incorrect to a correct response, as *harmful* if it was from a correct response to an incorrect one, and as *neutral* it was from one incorrect answer to another incorrect answer (which neither increases nor decreases accuracy). Beneficial switches (21.19% of switches) occurred at twice the rate of harmful switches (11.86% of switches). That is, reconsideration improved accuracy more often than it harmed it, and thus conditions that promote reconsideration should yield a net benefit to recall.[5] The lack of a statistically significant effect on final accuracy in Experiment 1, then, is likely to simply reflect the low overall rate of switches.

This interpretation suggests that, under other task conditions, the mismatch conditions would actually influence final recall accuracy. We tested this prediction in Experiment 2.

Experiment 2

Experiment 1 established that conditions of metacognitive match versus mismatch influence the probability that participants reconsider a recall response. In Experiment 2, we tested whether those conditions could also influence recall accuracy. Consequently, we turned from an experimental

[5] Further evidence that reconsideration is a strategy that can be used when an initial retrieval attempt is unsuccessful comes from an individual differences analysis: Participants who had lower initial accuracy rates were more apt to switch, $r = -0.53$, $p < .01$. Although individual differences were not the focus of our current investigation, differences among individuals in their switching rates or switching strategies may be an interesting avenue for future investigation.

procedure designed to characterise response switches to a procedure designed to maximise our ability to detect changes in recall.

In particular, external sources of information about memory difficulty might influence recall performance more strongly when participants could incorporate the information into their initial response. Participants may be less willing to reconsider when they have already overtly produced one response (Van Zandt & Maldonado-Molina, 2004); moreover, being tested may itself affect subsequent memory (Roediger & Karpicke, 2006). Thus, in Experiment 2, we presented the opponent's performance *before* the participant ever attempted to recall the item. Participants first viewed the opponent's performance and only afterwards saw the cue and made a single recall response. This task allows the opponent's performance to be considered from the outset and could encourage greater use of this cue in interrogating one's memory.

Gains in accuracy might also be masked if participants retrieve correct answers but do not choose to report them (Goldsmith & Koriat, 1994). Thus, in Experiment 2 we adopted a forced-report procedure in which participants were required to answer every item and could not leave them blank.

Method

Participants

Thirty individuals participated in Experiment 2. Only one participant indicated in the debriefing that they suspected the opponent was not real; again, we included this participant in the analysis, but excluding this participant did not alter the pattern of results.

Procedure

The materials and procedure were identical to Experiment 1 with three exceptions. The first was that, during the test phase, the opponent's performance was displayed first, in the same format as in Experiment 1. After a 1250 ms delay, participants made only a single response while the opponent's performance remained on the screen. The second change was that participants were required to respond to every item; the experiment did not accept a blank response. Finally, Experiment 2 omitted the phase in which participants estimated the opponent's performance.

Results

The measure of interest in Experiment 2 was the accuracy of the single response that participants made to each item. Figure 3 displays accuracy in each experimental condition.

We modelled accuracy using a mixed-effects logit model. A random slope for relatedness by items reliably improved the model in a likelihood ratio test, $\chi^2_{(2)} = 14.49$, $p < .01$, indicating that the cue relatedness effect was greater for some pairs than others. No other random slopes improved the model fit (all *ps* > .20).

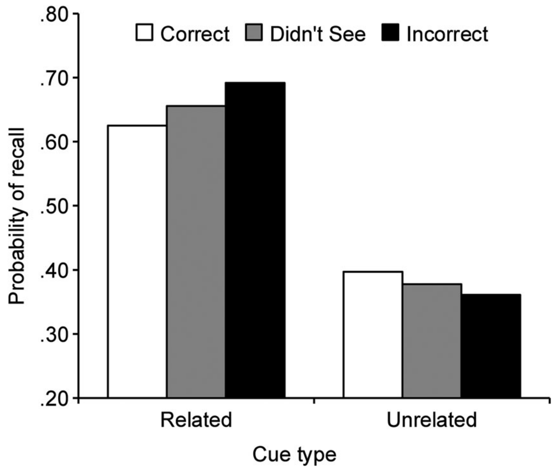

Figure 3. Cued recall accuracy in Experiment 2 as a function of cue relatedness and opponent's performance.

Table 4 displays parameter estimates from the model. As in Experiment 1, participants recalled more related pairs ($M = 65.74\%$) than unrelated pairs ($M = 37.87\%$), with the odds of correct recall being 4.66 times greater (95% CI: [3.52, 6.17]) for related pairs. Neither pair type approached floor or ceiling recall performance, potentially allowing accuracy for each to vary as a function of the opponent's performance.

The interaction between cue relatedness and the opponent's performance observed in response switching in Experiment 1 now also obtained in accuracy. Feedback that the opponent answered an item incorrectly increased accuracy for easy items but decreased accuracy for hard items: For related, easy pairs, participants were more accurate when given the mismatching feedback that the opponent answered incorrectly ($M = 69.17\%$ accurate) as compared to the baseline rate for all related pairs ($M = 65.74\%$), but for unrelated, difficult pairs, recall was less accurate when given the matching feedback that the opponent answered incorrectly ($M = 36.11\%$) as compared to the baseline for all unrelated pairs ($M = 37.87\%$). Thus, the odds of correct recall given *Incorrect* feedback were 1.79 times (95% CI: [1.01, 3.19]) higher with a related cue what they were with an unrelated cue.

Conversely, feedback that the opponent answered the item correctly decreased accuracy for easy pairs but improved it for hard pairs. For related, easy items, recall

Table 4. Fixed-effect estimates for multi-level logit model of cued recall accuracy in Experiment 2 (model $N = 2160$, log-likelihood = −1214).

Fixed effect	β	*SE*	Wald *z*	*p*
Intercept	0.14	0.23	0.60	.55
Related cue	1.54	0.14	10.76	<.01
Opponent was correct	−0.08	0.15	−0.55	.58
Opponent was incorrect	0.10	0.15	0.71	.48
Related cue × opponent correct	−0.56	0.29	−1.94	.05
Related cue × opponent incorrect	0.58	0.29	1.99	<.05

Note: SE = standard error.

was less accurate when the opponent answered incorrectly ($M = 62.50\%$) as compared to the baseline ($M = 65.74\%$); for unrelated, hard items, recall was more accurate when the opponent answered correctly ($M = 39.72\%$) as compared to the baseline ($M = 37.87\%$). Thus, the odds of answering correctly when given *Correct* versus other feedback were 1.75 times greater (95% CI: [0.99, 3.09]) with a related cue than an unrelated cue, although this latter difference was only marginally significant in Experiment 2.

Discussion

Experiment 2 tested the hypothesis that recall accuracy could be improved by reconsideration prompted by the mismatch of an opponent's performance with one's own perception of an item. This hypothesis was supported: recall accuracy increased in conditions in which the opponent's performance mismatched participants' likely experience retrieving an item—for easy-to-retrieve (related) pairs, when the opponent answered incorrectly, and for hard-to-retrieve (unrelated) pairs, when the opponent answered correctly. These are the same conditions that increased switching in Experiment 1. Conversely, accuracy decreased when the opponent's performance was likely to match participants' own experience; that is, the same conditions that discouraged switching in Experiment 1. This pattern suggests that the reconsideration brought on by information about others' performance can actually influence recall performance—sometimes to the rememberer's benefit—perhaps especially when task conditions encourage consideration of such information.

These differences in accuracy cannot be explained simply by participants' willingness to respond. The forced-report procedure in Experiment 2 disallowed omissions and required a response on every single trial. Rather, information about another person's mnemonic performance actually influenced what participants reported when they made their reports.

Experiment 2 also provides evidence against an alternative account of participants' use of the information about the opponent. One other way that participants in both experiments might have used the feedback about the opponent is to respond with a word related to the target when they saw that the opponent had answered correctly and with a word unrelated to the target when they saw that the opponent was incorrect (under a naïve theory that related word pairs are more apt to be answered correctly). Such a strategy, however, could not explain the result of Experiment 2: If "opponent correct" feedback led participants to guess words related to the cue, then it should have impaired performance on unrelated cue–target pairs (for which a related word is never the answer), but in fact it *benefited* performance on such pairs; conversely, if "opponent incorrect" feedback led participants to guess unrelated words, it should have impaired recall on related pairs, but it benefited them. Thus, participants did not simply use the opponent's performance as a guide to guess related or unrelated words. Rather, the evidence above suggests they used it in conjunction with their own assessment of the item to decide whether to reconsider an item.

General discussion

Rememberers often know whether others succeeded or failed at retrieving certain information, but little prior work has investigated how such information might be incorporated into metacognitive control of recall. We contrasted two hypotheses suggested by metacognition in other domains. The difficulty-cue hypothesis proposes that rememberers look for indicators of item difficulty, such as others' failure to retrieve it, and reconsider those items that appear difficult. This hypothesis predicts a main effect of others' performance on recall behaviour. By comparison, the mismatch-monitoring hypothesis proposes that rememberers attend to whether others' mnemonic performance matches or mismatches their own perception of an item. According to this hypothesis, rememberers devote additional reconsideration to items for which external indicators of difficulty mismatch one's own internal assessment of the item, and they devote less reconsideration to items for which external cues accord with one's existing assessment. This hypothesis thus predicts that others' performance would interact with the participant's own perception of item difficulty.

We assessed these hypotheses in two cued recall experiments by manipulating the performance of a fictive opponent independently of the associative strength of word pairs. Less strongly associated pairs are both objectively harder to remember and subjectively viewed as such (Dunlosky & Matvey, 2001; Koriat, 1997;), so we expected that unrelated pairs would present a more difficult retrieval experience for participants (and perhaps even be explicitly categorised by participants as *difficult* or *unrelated*). Consistent with this expectation, in both experiments, unrelated pairs were objectively less apt to be remembered, and a manipulation check in Experiment 1 confirmed that participants also subjectively thought they remembered fewer of the unrelated pairs. The critical question was whether this manipulation of implied difficulty would interact with the information about the opponent, as predicted by the mismatch-monitoring hypothesis. In both experiments, it did. In Experiment 1, participants switched their response more when others' mnemonic performance conflicted with item difficulty. Items with related cues were expected to be generally easy to recall; for example, candidate answers might come to mind more quickly, participants might be more confident in the candidate answers, and the cue–target relationship might be explicitly recognised by participants who have a (correct) naïve theory that related pairs are easier to remember. For these easy items, participants switched more when the opponent purportedly answered incorrectly and less if the opponent answered correctly. But for items with unrelated cues, which were generally more difficult to retrieve, participants switched more if the opponent supposedly answered correctly and less if the opponent answered incorrectly. Experiment 2 further revealed that, under some task conditions, reconsideration based on others' performance even influences recall accuracy, sometimes allowing participants to recall more than they would otherwise. Experiment 2 also indicated that the reconsideration induced by mismatch conditions can influence *what* is reported and not only *whether* a report is made.

The effects of others' knowledge on recall

The first major finding of the present study is that, in controlling their recall, individuals monitor whether others' knowledge states match or mismatch their own perceptions of an item. In Experiment 1, information about a fictive opponent's performance, presented after an initial recall attempt, influenced the rate at which participants changed their answer on a second attempt. In particular, participants switched responses most often when the opponent's performance *mismatched* participants' own presumed level of difficulty retrieving the item: for easy-to-retrieve items, when the fictive opponent was wrong, and for hard-to-retrieve items, when the opponent was correct.

This pattern of results is inconsistent with the predictions of the difficulty-cue hypothesis, which predicted that participants would be most apt to reconsider items that the opponent answered incorrectly (and thus that appeared difficult), regardless of cue type. One variant of the difficulty-cue hypothesis that cannot be fully ruled out is that participants do just attend to signs of difficulty, rather than a metacognitive mismatch, but that those signs of difficulty have non-monotonic effects: Difficulty generally increases the probability of reconsideration, but if an item appears *too* difficult, participants simply give up on it entirely. Such an account could explain the low probability of reconsideration in Experiment 1 for unrelated pairs that the opponent answered incorrectly. However, the fact that the same interaction emerged even under forced-report conditions in Experiment 2, in which "giving up" was not an option, renders such a possibility unlikely.

Rather, the interaction of item difficulty and others' knowledge suggests that what rememberers monitor is whether others' recall behaviour conflicts with their own. In general, a greater degree of conflict may be a cue to deploy additional cognitive control (Botvinick et al., 2001). In the present case, rememberers had multiple sources of information to integrate in controlling their recall; these sources included both their own metacognitive experience recalling the item (which could involve an explicit categorisation of the items as *related* versus *unrelated* and/or more implicit measures such as feeling of knowing) and others' mnemonic performance. A discrepancy between these two cues might signal that one has taken an erroneous approach to the item and that different or additional knowledge should be retrieved and applied. Consequently, when participants' own metacognitive experience conflicted with the opponent's performance, they more often reconsidered their answer, leading them to change their response (Experiment 1) or output a response that was more likely to be correct (Experiment 2). Conversely, a match between these two cues provided further evidence in support of participant's existing approach and discouraged reconsideration.

Why can reconsideration brought about by conflict produce more accurate recall? Rememberers often do not sample all relevant knowledge in their first retrieval attempt, so a second guess, particularly one distinct from the first, can improve accuracy (Anderson & Pichert, 1978; Fraundorf & Benjamin, 2014; Herzog & Hertwig, 2009; Hourihan & Benjamin, 2010; Van Zandt & Maldonado-Molina, 2004; Vul & Pashler, 2008). The benefits of exposure to a conflicting perspective is also consistent with the importance of dissent and opposing perspectives in group decision-making, which can stimulate divergent thinking and creativity even at the individual level (Nemeth & Nemeth-Brown, 2003; Nemeth, Personnaz, Personnaz, & Goncalo, 2004). In each of these cases, exposure to conflicting information stimulates a reconsideration or search for additional information that benefits the ultimate decision.

The disordinal interaction between the opponent's performance and item difficulty suggests that, contrary to the difficulty-cue hypothesis, participants did not directly use the opponent's performance to determine the general level of effort to devote to an item. Participants did not just reconsider any item that another person answered incorrectly; indeed, for difficult items, it was actually *successful* retrieval by the opponent that led participants to reconsider their answer. Indeed, recent work (Ariel, Dunlosky, & Bailey, 2009; Metcalfe & Kornell, 2005) suggests that item difficulty alone is often insufficient to account for metacognitive behaviour. Rather, what rememberers attend to about another's mnemonic processes are whether they conflict or accord with their own.

When can social information benefit memory?

Our second major finding is that conditions of metacognitive conflict can lead participants to report more accurate answers in recall. Conversely, low-conflict (match) conditions actually impair recall relative to having no information about others' recall. Both of these effects on recall were observed only in Experiment 2, in which participants responded only after first seeing the opponent's performance and could not omit responses, suggesting that information about others' performance may be more influential in some task conditions than in others.

The fact that information about another person sometimes benefited memory appears to contrast with the finding that group recall performance is frequently *worse* than what the group members could achieve individually (e.g., Weldon & Bellinger, 1997). But, in those experiments, participants were exposed to the actual *products* of their partner's recall (e.g., specific words that were recalled), rather than just the success or failure of the partner's retrieval *process*. Those specific memory reports may disrupt recall similarly to part-list cuing (Wright & Klumpp, 2004).

Rather, the present results support the emerging perspective (Rajaram, 2011) that, under some circumstances, memory benefits from information about how others responded. Remembering as part of a group can enhance later individual memory by allowing individuals to learn veridical information from others (Basden et al., 2000) and filtering out false memories based on feedback (Rajaram & Pereira-Pasarin, 2007). Even *initial* memory judgements sometimes benefit from others' reports (Jaeger, Lauris, Selmeczy, & Dobbins, 2011). Collaborating with others may be particularly important for individuals who have more constrained memory abilities, such as older adults, or who are experienced at collaborating, such as long-term married couples (Dixon, 1996, 1999).

The present study extends these results by showing that it can also be important to know the *accuracy* of another individual's retrieval process, rather than the specific details they remembered (see also Reysen, 2003). Such information is probably frequently available, perhaps especially so in educational settings (e.g., hearing that a test is "easy" or "hard", or being cautioned that other students frequently find a particular concept challenging), and the present study suggests it can be productively used to control retrieval. Although this knowledge can sometimes be harmful, it is beneficial when metacognitive conflict encourages further consideration, consistent with the benefits of opposing perspectives on decision-making and creativity (Nemeth et al., 2004; Nemeth & Nemeth-Brown, 2003;).

What cues to mismatch?

Our primary hypotheses in this study concerned the conditions under which participants would be most apt to reconsider a cued recall response; these conditions turned out to be those in which the opponent's performance mismatched the participant's likely metacognitive state. A related, but unanswered, question is exactly which cues participants used to detect this mismatch. Other research (e.g., Kelley & Jacoby, 1996) has suggested that metacognitive decisions can be made either on the basis of subjective experience or on the basis of a naïve theory (or the combination thereof; Fraundorf & Benjamin, 2014). Either of these bases could have been used in the present experiment to detect the mismatch state. Participants could have used their subjective experience with the item on the cued recall task—how fluently possible responses came to mind, or how confident they felt in those responses—and compared that to the opponent's performance. Or, they could have categorised the pairs as *related* and *unrelated*, and then, on the basis of a (justified) naïve theory that related pairs should be easier to recall, considered whether their inference about the item difficulty matched or mismatched the opponent's performance. (The fact that some pairs were related and others unrelated was not explicitly presented to participants until the manipulation check phase after the final test phase, but it is possible that some participants independently noticed and made use of this fact.) In the present study, our objective was to examine how a mismatch between one's own metacognitive experience and others' mnemonic performance affected recall behaviour once it had been detected, but one avenue for future research is to determine whether participants use either or both of these bases when deciding whether another person's mnemonic performance accords with their own.

Another potential direction for future research is to determine which forms of information about others' memory can be adeptly used in recall. In the present experiments, information about others' knowledge states was operationalised as feedback about whether a specific individual responded correctly or incorrectly (or did not see the item) on a cued recall test. But, as in the real-world scenarios reviewed earlier, information about others' memory sometimes arrives in a slightly different form. For instance, one may receive an indicator of others' performance aggregated across multiple people, as when a student sees that two-thirds of the class has already completed a test but not the other third. Individuals may also encounter others' expressions of subjective difficulty in addition to, or instead of, objective information about recall success or failure; for example, a tourist may be cautioned that a route is "difficult" to remember, rather than objectively seeing another person fail to remember the route. It would be useful for future research to explore whether participants are equally sensitive to all of these forms of information and whether they are used differently, especially because information about others' memory is often seen in a wealth of different formats outside of the laboratory.

The influence on others' knowledge states on cognition

Although little work has investigated the influence of others' influence knowledge states on memory, research has revealed that individuals frequently make use of such cues in other cognitive domains. For example, in language comprehension, listeners interpret and remember utterances differently if the speaker seemed to have difficulty producing them (Arnold, Hudson Kam, & Tanenhaus, 2007; Barr, 2003; Fraundorf & Watson, 2011). Listeners may even consider the specifics of what a speaker knows: a phrase that could plausibly refer to either of two objects can be interpreted unambiguously when listeners know that the speaker could only see one object (Hanna, Tanenhaus, & Trueswell, 2003; but see Keysar, Barr, Balin, & Brauner, 2000). Even young infants are surprised when an actor reaches for a hidden object she should not know about (Onishi & Baillargeon, 2005).

One question meriting further investigation is the level of detail at which people model others' mnemonic performance. In language processing, it has been debated whether speakers and listeners consider their interlocutor only at the level of a generic individual (e.g., Brown & Dell, 1987) or consider a particular individual's knowledge (Hanna et al., 2003; Lockridge & Brennan, 2002) and abilities (Arnold et al., 2007). The present experiment required only generic knowledge that associated pairs are easier to remember than unassociated pairs. But other results suggest more detailed modelling of an individual's memory. Participants' willingness to accept eyewitness testimony is influenced by characteristics of the witness (e.g., police officers are viewed as more reliable than children) and by the witness's relation to the memorandum, such as how long it was studied (Gabbert et al., 2007; Wright et al., 2011). Rememberers may track the reliability of an informant even in the absence of an explicit cue to reliability: Recognition memory decisions are influenced more heavily by a generally accurate informant than a generally inaccurate one (Jaeger et al., 2011). However, the level of detail at which others are modelled remains to be tested in other tasks and as a function of other informant characteristics such as domain expertise.

Conclusion

Rememberers use others' knowledge in controlling their recall. In particular, they monitor whether others' performance matches or mismatches their own. When their own experience retrieving an item matches someone else's, they less frequently reconsider their response, but when their own experience conflicts with someone else's, they more frequently reconsider. This process of reconsideration involves not merely changes in whether participants make a mnemonic report, but in what answers they ultimately report. Under some task conditions, it even enhances recall accuracy.

More broadly, these findings support a view in which memory retrieval is not a passive process but one actively guided by metacognition. The selection of cues and retrieval strategies depends on rememberers' expectations about items and their monitoring of their own cognition, and information about others' mnemonic performance is an important constraint on this process.

Acknowledgements

We thank Jason Finley for stimulus materials, Kirstin Shafer and Erika Schmit for data collection, and members of the University of Illinois Human Memory and Cognition Lab for their comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the United States National Science Foundation [grant number GRFP2007053221]; the United States National Institutes of Health [grant number 5T32HD055272], [grant number R01AG026263] to Aaron S. Benjamin.

References

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unavailable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*, 1–12. doi:10.1016/S0022-5371(78)90485-1

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, *138*, 432–447. doi:10.1037/a0015928

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say *thee uh-* you're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 914–930. doi:10.1037/0278-7393.33.5.914

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005

Balota, D., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 576–587. doi:10.1037/0278-7393.6.5.576

Barr, D. (2003). Paralinguistic correlates of conceptual structure. *Psychonomic Bulletin & Review*, *10*, 462–467. doi:10.3758/BF03196507

Basden, B. H., Basden, D. R., & Henry, S. (2000). The costs and benefits of collaborative remembering. *Applied Cognitive Psychology*, *14*, 497–507. doi:10.1002/1099-0720(200011/12)

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using s4 classes (version 0.999375-42) [Computer software]. Retrieved from http://CRAN.R-project.org/package=lme4

Benjamin, A. S. (2005). Response speeding mediates the contribution of cue familiarity and target retrievability to metamnemonic judgments. *Psychonomic Bulletin & Review*, *12*, 874–879. doi:10.3758/BF03196779

Benjamin, A. S. (2008). Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation: Vol. 48. Skill and strategy in memory use* (pp. 175–123). San Diego, CA: Elsevier. doi:10.1016/S0079-7421(07)48005-7

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652. doi:10.1037//0033-295X.108.3.624

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*, 441–472. doi:10.1016/0010-0285(87)90015-6

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *27*, 1491–1494. doi:10.1037//0278-7393.27.6.1491

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359. doi:10.1016/S0022-5371(73)80014-3

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). Categorical or nominal independent variables. In *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed., pp. 302–353). Mahwah, NJ: Lawrence Erlbaum Associates.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*, 497–505. doi:10.1080/14640748108400805

Dixon, R. A. (1996). Collaborative memory and aging. In D. Hermann, C. McEvoy, C. Hertzog, P. Hertel, & M. K. Johnson (Eds.), *Basic and applied memory research: Vol. 1. Theory in context* (pp. 359–383). Mahwah, NJ: Lawrence Erlbaum Associates.

Dixon, R. A. (1999). Exploring cognition in interactive situations: The aging of N + 1 minds. In T. M. Hess & F. Blanchard-Fields (Eds.), *Social cognition and aging* (pp. 267–290). San Diego, CA: Academic Press. doi:10.1016/B978-012345260-3/50013-6

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1180–1191. doi:10.1037/0278-7393.27.5.1180

Erdelyi, M. H. (2010). The ups and downs of memory. *American Psychologist*, *65*, 623–633. doi:10.1037/a0020440

Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review*, *16*, 88–92. doi:10.3758/PBR.16.1.188

Fiechter, J. L., Benjamin, A. S., & Unsworth, N. (in press). The metacognitive foundations of effective remembering. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford Handbook of Metamemory*. Oxford: Oxford University Press.

Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *38*, 632–652. doi:10.1037/a0026215

Finley, J. R., Tullis, J. G., & Benjamin, A. S. (2010). Metacognitive control of learning and remembering. In M. S. Khine & I. M. Saleh (Eds.), *New science of learning: cognition, computers and collaboration in education* (pp. 109–131). New York, NY: Springer. doi:10.1007/978-1-4419-5716-0_6

Fraundorf, S. H., & Benjamin, A. S. (2014). Knowing the crowd within: Metacognitive limits on combining multiple judgments. *Journal of Memory and Language*, *71*, 17–38. doi:10.1016/j.jml.2013.10.002

Fraundorf, S. H., & Watson, D. G. (2011). The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language*, *65*, 161–175. doi:10.1016/j.jml.2011.03.004

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language*, *62*, 1–18. doi:10.1016/j.jml.2009.09.004

Gabbert, F., Memon, A., & Wright, D. B. (2007). I saw it for longer than you: The relationship between perceived encoding duration and memory conformity. *Acta Psychologica*, *124*, 319–331. doi:10.1016/j.actpsy.2006.03.009

Goldsmith, M., & Koriat, A. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, *123*, 297–315. doi:10.1037/0096-3445.123.3.297

Goldsmith, M., & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation: Vol. 48. Skill and strategy in memory use* (pp. 1–60). San Diego, CA: Elsevier. doi:10.1016/S0079-7421(07)48001-X

Halamish, V., Goldsmith, M., & Jacoby, L. L. (2012). Source-constrained recall: Front-end and back-end control of retrieval quality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1–15. doi:10.1037/a0025053

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43–61. doi:10.1016/S0749-596X(03)00022-6

Harris, C. B., Paterson, H. M., & Kemp, R. I. (2008). Collaborative recall and collective memory: What happens when we remember together? *Memory*, *16*, 213–230. doi:10.1080/09658210701811862

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*, 231–237. doi:10.1111/j.1467-9280.2009.02271.x

Higham, P. A., & Tam, H. (2005). Generation failure: Estimating metacognition in cued recall. *Journal of Memory and Language*, *52*, 595–617. doi:10.1016/j.jml.2005.01.015

Hourihan, K. L., & Benjamin, A. S. (2010). Smaller is better (when sampling from the crowd within): Low memory-span individuals benefit more from multiple opportunities from estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1068–1074. doi:10.1037/a0019694

Jaeger, A., Lauris, P., Selmeczy, D., & Dobbins, I. (2011). The costs and benefits of memory conformity. *Memory & Cognition*, *40*, 101–112. doi:10.3758/s13421-011-0130-z

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformations or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. doi:10.1016/j.jml.2007.11.007

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, *35*, 157–175. doi:10.1006/jmla.1996.0009

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*, 32–38. doi:10.1111/1467-9280.00211

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. doi:10.1037/0096-3445.126.4.349

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490–517. doi:10.1037/0033-295X.103.3.490

Koriat, A., Goldsmith, M., & Halamish, V. (2008). Control processes in voluntary remembering. In J. Byrne & H. L. Roediger, III (Eds.), *Cognitive psychology of memory* (Vol. *2*, pp. 307–324). Oxford: Elsevier.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, *9*, 550–557. doi:10.3758/BF03196312

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*, 463–477. doi:10.1016/j.jml.2004.12.001

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation: Vol. 26. Advances in research and theory* (pp. 125–173). New York, NY: Academic Press. doi:10.1016/S0079-7421(08)60053-5

Nemeth, C., & Nemeth-Brown, B. (2003). Better than individuals?: The potential benefits of dissent and diversity for group creativity. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity: Innovation through collaboration* (pp. 63–84). New York, NY: Oxford University Press.

Nemeth, C. J., Personnaz, B., Personnaz, M., & Goncalo, J. A. (2004). The liberating role of conflict in group creativity: A study in two countries. *European Journal of Social Psychology*, *34*, 365–374. doi:10.1002/ejsp.210

Onishi, K., & Baillargeon, R. (2005). Do infants understand false beliefs? *Science*, *308*, 255–258. doi:10.1126/science.1107621

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, *101*, 5–27. doi:10.1037/0033-2909.101.15

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. doi:10.1163/156856897X00357

Rajaram, S. (2011). Collaboration both hurts and helps memory: A cognitive perspective. *Current Directions in Psychological Science*, *20*, 76–81. doi:10.1177/0963721411403251

Rajaram, S., & Pereira-Pasarin, L. P. (2007). Collaboration can improve individual recognition memory: Evidence from immediate and delayed tests. *Psychonomic Bulletin & Review*, *14*, 95–100. doi:10.3758/BF03194034

Reysen, M. B. (2003). The effects of social pressure on group recall. *Memory & Cognition*, *31*, 1163–1168. doi:10.3758/BF03195799

Roediger, H. L.III, & Karpicke, J. D. (2006). The power of testing: Basic research and implications. *Perspectives on Psychological Science*, *1*, 181–210. doi:10.1111/j.1745-6916.2006.00012.x

Roediger, H. L.III, Meade, M. K., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin & Review*, *8*, 365–371. doi:10.3758/BF03196174

Son, L. K., & Metcalfe, N. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 204–221. doi:10.1037//0278-7393.26.1.204

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1024–1037.

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language*, *64*, 109–118. doi:10.1016/j.jml.2010.11.002

Tullis, J. G., & Benjamin, A. S. (in press). Cueing others' memories. *Memory & Cognition*. doi:10.3758/s13421-014-0478-y

Van Zandt, T., & Maldonado-Molina, M. M. (2004). Response reversals in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1147–1166. doi:10.1037/0278-7393.30.6.1147

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *19*, 645–647. doi:10.1111/j.1467-9280.2008.02136.x

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1160–1175. doi:10.1037/0278-7393.23.5.1160

Wright, D. B., & Klumpp, A. (2004). Collaborative inhibition is due to the product, not the process, of recalling in groups. *Psychonomic Bulletin & Review*, *11*, 1080–1083. doi:10.3758/BF03196740

Wright, D. B., Memon, A., Skagerberg, E. M., & Gabbert, F. (2011). When eyewitnesses talk. *Current Directions in Psychological Science*, *18*, 174–178. doi:10.1111/j.1467-8721.2009.01631.x