1-1-2011

Likely Old or Likely New? Taking Memory Advice During Recognition Decisions

Diana Selmeczy

Recommended Citation

Selmeczy, Diana, "Likely Old or Likely New? Taking Memory Advice During Recognition Decisions" (2011). *All Theses and Dissertations (ETDs)*. 532.
https://openscholarship.wustl.edu/etd/532

WASHINGTON UNIVERSITY

Department of Psychology

Likely Old or Likely New? Taking Memory Advice During Recognition Decisions

by

Diana Selmeczy

A thesis presented to the
Graduate School of Arts and Sciences of
Washington University in
partial fulfillment of the
requirements for the
degree of Master of Arts

December 2011

Saint Louis, Missouri

Abstract

We examined the influence of external recommendations on recognition memory decisions. In contrast to prior literature that primarily focuses on the negative impacts of external influences during memory judgments, we investigated whether participants can capitalize on explicit reliable recommendations in order to improve their performance. In the first experiment, participants were given explicit external recommendations ("Likely Old" or "Likely New") that were 75% accurate for deeply and shallowly encoded test items. In the second experiment, participants were given varying levels of recommendations (65% and 85% accurate). Across both experiments we found that participants improved their performance when given external recommendations relative to when no recommendations were available. Furthermore, we found that the degree to which participants benefitted from external recommendations is, in part, dependent on metacognitive monitoring ability. Finally, corrective feedback did not seem to improve participants' ability to utilize external recommendations.

Acknowledgments

Research Support: MH073982

Table of Contents

List of Tables

List of Figures

Introduction

Recognition memory does not take place in a vacuum and it is often the case that environmental factors can signal the likely memory status of an encountered stimulus. For example, when identifying a person in the hall one might use cues such as location (are you in a place where most people tend to be familiar?), time of day (are you likely to encounter this person at this time?), or a nearby friend's explicit opinion about whether or not he/she recognizes the individual. Such an approach would be ideal because it would mean that the observer is not wasting potentially valuable information when making recognition judgments. That is, the observer would be utilizing two useful sources of information, namely, environmental cues and internal memory signals when judging memory status. However, inappropriately incorporating this information —such as under relying or over relying on external cues— can result in awkward and costly mistakes (e.g. accidentally approaching a stranger because you ignored your friend's advice that the person in the hall seems highly unfamiliar). Therefore, ideal observers should not only be sensitive to environmental factors, but also rely on them judiciously.

One extant framework that suggests that observers should be able to judiciously integrate cues into recognition judgments is the Theory of Signal Detection (Macmillan & Creelman, 2005). Indeed, this model was developed under the assumption that observers render decisions in a statistically ideal manner (Pastore, Crawley, Berens, & Skelly, 2003). In the case of recognition memory judgments it is assumed that observers estimate two likelihoods based on the memory strength of each test item, namely, the likelihood that an item of a particular strength level is from the studied pool and the likelihood that the item is from the novel pool. Unsurprisingly, the observer chooses the response with the highest likelihood, and this decision strategy is captured by the ratio of the likelihoods or the odds the item is old. In other words, the point at which the two likelihoods are equal (i.e. 1:1odds) is an optimal neutral point, and items with odds above this likelihood criterion should be called old, while items below this criterion should be called new (Figure 1A). The likelihood ratio decision model is referred to as an ideal observer model because it maximizes the long-term accuracy of the decision maker. Such a model is assumed in several different applications of Signal Detection Theory to recognition memory judgments (see Glanzer, Hilford, & Maloney, 2009).

If one assumes observers have access to such sophisticated information, the integration of environmental influences, such as external recommendations, into memory judgments is very straightforward, provided the reliability of the source of recommendations is known (c.f., Jaeger, Lauris, Selmeczy, & Dobbins, 2011). For example, imagine an observer receives an external recommendation (Likely Old or Likely New) that is 75% predictive of the upcoming test items memory status. A Likely Old cue would indicate 3:1 odds that the upcoming item will be old vs. new, while a Likely New cue would indicate a 1:3 odds that the upcoming item would be old vs. new. An ideal observer would integrate this external recommendation into their decision by simply multiplying these odds with those derived from their own internal memory evidence, namely, the likelihood ratio decision variable assumed under Signal Detection Theory. For example, if an observer receives the recommendation Likely Old and their internal memory evidence suggests that the odds the item is old are 2:1, the odds specified by the recommendation and those indicated by the memory evidence (3/1 * 2/1) yields a final or posterior odds that the item is old of 6 to 1. This strongly indicates the participant should respond old and should do so more confidently than he or she would have if the external recommendation were absent. As shown in Figure 1B, when given the recommendation Likely Old the entire likelihood decision axis is shifted, and the optimal decision criterion (odds ratio of 1:1) is shifted to the left. Returning to the prior example, if the observer instead received the recommendation Likely New for the same trial, then the posterior odds would be 2 to 3 that the item was old (1/3 * 2/1), indicating that the participant should now respond new. In this situation even though the observers' internal evidence indicates that the item is old, when evaluated in light of the external recommendation the ideal response would be that the item is in fact new (Figure 1C).

By optimally moving the decision criterion under external cueing as described above, observers maximize their long-term accuracy and elevate their performance relative to situations where no external cues are available. For example, under the case in which there are 100 studied and 100 new items the unbiased observer with no external cues in Figure 1A (d' = 1) would respond correctly on 140 trials. If, however, the actual odds of encountering old items were 3 to 1, such as under a Likely Old cue, then maintaining the same criterion location would not be ideal because the observer is not capitalizing on the disproportionate likelihood of encountering an old item (i.e. 75 old items to 25 new items). Instead the criterion should be shifted to the point on the axis where the prior odds of an item being old were 1 in 3 (Figure 1). Similarly, if the actual odds of encountering an old item were now instead 1 to 3 (i.e. 25 old items to 75 new items), such us under a Likely New cue, the criterion should be shifted to the point on the axis where the prior odds of an item being old were 3 in 1 (Figure 1). The performance difference between leaving the criterion at the midpoint across cued trials and uncued trials versus shifting between the two ideal locations for the two cueing conditions would be 139 correct responses in the former and 156 correct responses in the latter (See Footnote 1). Thus, under a Signal Detection Theory model that assumes a likelihood ratio decision axis, observers should easily be able to integrate external recommendations into recognition judgments in order to considerably improve long-term accuracy when in the presence of reliable recommenders. From this perspective, the presence of external cues in the environment is always viewed as a benefit provided their base rate validity is above chance. However, in contrast to this positive outlook on the benefits of external influences during memory judgments, most prior work has instead focused on the negative impacts that such sources can have.

Based on the pioneering social conformity work of Asch (1955), memory researchers have considered the potential negative impact of external recommendations during memory attributions. In a typical memory conformity experiment, participants are led to believe that they studied an identical set of stimuli as another participant or confederate, when in fact a subset of studied material was different. There are many variations as to how participants are later tested on these items (in groups vs. pairs, intentionally planned confederate vs. another participant as a confederate, virtual confederate vs. real confederate, etc.), but the general finding is that participants will conform to the response of the confederate (Allan & Gabbert, 2008; Axmacher, Gossen, Elger, & Fell, 2010; Betz & Skowronski, 1996; Meade & Roediger, 2002; Reysen, 2005; Roediger, Meade, & Bergman, 2001; Schneider & Watkins, 1996; Walther et al., 2002; Wright, Mathews, & Skagerberg, 2005; Wright, Self, & Justice, 2000; Wright, Gabbert, Memon, & London, 2008). Some of these memory conformity studies assess how memory on a final test is altered by false information introduced by a confederate during an earlier test (e.g. Meade & Roediger, 2002; Roediger et al., 2001; Betz & Skowronski, 1996), while other experiments specifically examine how participants' responding is influenced by trial-by-trial information suggested immediately before a recognition judgment (e.g. Schneider & Watkins, 1996; Reysen, 2005 group recognition test; Wright et al., 2000 Experiment 1). This prior research demonstrates that observers are sensitive to external influences, but much of this research uses deceptive others and focuses on its negative impacts. In these situations the observer may be best served by completely ignoring external influences.

In contrast to traditional memory conformity research, here we examine the degree to which observers can benefit from external cues from a source known to be reliable. As noted earlier, the Signal Detection model illustrated in Figure 1 anticipates that observers should be able to easily incorporate such information. However, many have questioned the degree of statistical sophistication assumed under the model and there are several findings that suggest actual performance can fall short of ideal. For example, observers do not appear to adopt ideal criteria when provided monetary payout matrices (Green & Swets, 1989; Healy & Kubovy, 1978; MacMillan and Creelman, 2005) and response rates do not change for tests that are composed purely of one item type (i.e. only targets or only novel items) relative to standard recognition tests composed of both old and new items (Cox & Dobbins, 2011). One possibility for such shortcomings in memory paradigms is that observers either lack, or considerably differ in the degree to which they are subjectively aware of small gradations in their internal memory strength representations, a skill that would fall under the more general rubric of metacognitive awareness. More specifically, the extent to which participants' changes in subjective confidence track changes in memory veracity has been referred to as metacognitive monitoring (Nelson, 1990). For example, observers differ in their ability to assess how well they have learned particular target material (judgments of learning) or their ability to predict future recognition of a general knowledge answer they could not currently recall (feeling of knowing)(Metcalfe & Dunlosky, 2009)— however, this monitoring ability may transfer poorly between domains (see Kelemen, Frost, & Weaver, 2000). Given that accurate subjective awareness of memory representations may be potentially important for ideal criteria placement and hence effective cue utilization, we also examine a measure of metacognitive monitoring to see if it accounts for any differences in the ability of observers to benefit from external recommendations. That is, we examine whether the utilization of external recommendations, at least in part, depends on metacognitive monitoring skills. Prior memory conformity research has begun to investigate metamnemonic awareness, but this research tends to focus on heuristics about when it is more or less appropriate to rely on an external source and does not examine individual differences in monitoring. These studies show that conformity increases when the confederate is highly confident (Schneider & Watkins, 1996; Wright et al., 2000), when items have lower memorability (Betz & Skowronski, 1996), when perceived encoding time is manipulated (Gabbert, Memon, & Wright, 2007), and when previously unseen items are non-salient (Walther et al., 2002).

Finally, were also interested in the possibility that the ability to use recognition recommendations might benefit from performance feedback. Prior work has suggested that trial-by-trial feedback may be necessary for accurate representations of statistical likelihoods (Turner, Van Zandt, & Brown, 2011), and a host of studies have demonstrated that feedback results in more appropriate criterion placement (Estes & Maddox, 1995; Kantner & Lindsay, 2010; Rhodes & Jacoby, 2007; Verde & Rotello, 2007). Based on this prior research, we hypothesized that feedback might result in more appropriate criterion shifts in response to the external recommendations and hence a greater improvement in performance when comparing uncued to cued recognition accuracy.

To summarize, our aims are to determine: a) whether observers can effectively incorporate predictive external recommendations into their recognition decisions, b) the role of metacognitive awareness in this skill, and c) whether feedback improves this ability. In our first experiment we manipulated levels of processing during encoding in order to assess subsequent cueing effects under varying levels of memory evidence, while in our second experiment we manipulated the levels of cue validity in order to determine if our results replicate under these conditions.

Experiment 1

Participants

Experiment 1 included 37 Washington University students (average age = 20.9, 23 females) who were paid $20 for participation. Three subjects were removed due to low performance (d'<0.19) leaving 34 subjects for analyses. All participants provided informed consent in accordance with the University's Institutional Review Board.

Materials and Procedure

Testing was self-paced with observers entering their responses via keyboard, and presentation and timing controlled via Matlab's Psychophysics Toolbox (version 3.0.8) (Brainard, 1997; Pelli, 1997). For each participant, words were randomly selected from a 1216 item pool with an average of 7.09 letters, 2.34 syllables and Kučera -Francis frequency of 8.85.

Participants completed four study/test cycles, with two tests preceded by deep encoding and two shallow encoding. The order of deep and shallow tests sequentially alternated with half the participants beginning with the shallow test condition and half the deep test condition (100 study items and 200 test items for each cycle). During shallow encoding participants indicated whether the first and last letter of each presented word was in alphabetical order, whereas during deep encoding they performed an abstract/concrete rating. Recognition testing immediately followed each study phase, with subjects indicating whether randomly intermixed old and new items were studied ("old") or novel ("new") (100 old items, 100 new items). On 120 of the test trials (60 old, 60 new) a probabilistic mnemonic cue, "Likely Old" or "Likely New", was presented one second before the probe word appeared. These cues were correct 75% of the time, with subjects correctly informed that "Cues will be correct 75% of the time. This means about 7 out of 10 times the cue will give you the correct answer and should be useful for your recognition judgment." In addition to the cued trials, there were 80 (40 old, 40 new) baseline uncued trials intermixed in the test phase, with participants notified that some portion of the probes would be presented without anticipatory cues. After each old/new recognition decision, participants provided confidence on a 6-point scale ranging from 50% (guessing) to 100% (certain), which was then immediately followed by corrective feedback for half the participants.

Results and Discussion

The order in which the two levels of processing conditions were administered did not influence accuracy (d') or criteria (C), nor did it interact with other factors. Given this, we collapsed across test order in the analyses below. Hit rates of 1 and false alarm rates of 0 were corrected using the formulas suggested by MacMillan and Creelman (2005) (1-1/(2N) for hits and 1/2N for false alarms, where N is the number of trials).

Does accuracy improve with provision of cues?

To assess potential gains in accuracy (d'), we used a 2 X 2 X 2 mixed ANOVA with repeated measures factors of Levels of Processing (deep vs. shallow targets present during test) and Cue Condition (cued vs. baseline), and a between subjects factor of Feedback (present or absent). Results revealed a main effect of Levels of Processing ($F(1,32)=174.32$, $MSe = 0.26$, $p < 0.001$) reflecting higher accuracy for deep than shallow test items. There was also a main effect of Cue Condition ($F(1,32)=36.65$, $MSe=0.06$, $p<0.001$), indicating that participants significantly improved performance on cued vs. uncued/baseline trials (Table 1). In contrast, the interaction between Feedback and Cue Condition was not significant ($F(1,32)=1.09$, $MSe=0.06$, $p=0.30$), indicating that feedback did not have an appreciable effect on accuracy. There were no significant two-way interactions and the three-way interaction also failed to reach significance.

Overall these analyses demonstrate that participants increased their accuracy on cued trials relative to baseline trials for both deeply and shallowly encoded items, and this improvement in performance was not dependent upon the provision of feedback. Thus, although they are effectively incorporating the cues into their judgments, the mechanism by which this occurs does not appear to require or benefit from feedback based learning. We further consider the inefficacy of feedback in the discussion.

Reactivity to Cues

Because the accuracy analysis demonstrates that observers are improving when cues are in the environment, it is clearly the case that these cues are being used to adjust decision standards (Table 2). Nonetheless, we wanted to verify that observers were shifting criteria more vigorously during shallow tests than during deep tests, because this pattern should result if the cues are being considered in light of the recognition evidence. That is, the cues should have more influence when the internal evidence is less discriminable (shallow tests) then when it is more discriminable (deep tests). Using C as our criteria measure, we ran a 2 X 2 X 2 mixed ANOVA with repeated measures of Levels of Processing (deep vs. shallow) and Cue Type (Likely Old vs. Likely New), and a between subjects measure of Feedback (present or absent). Critically, we found a significant interaction between Levels of Processing and Cue Type ($F(1,32)=21.34$, $MSe = 0.06$, $p < 0.001$), demonstrating that the difference in criteria across Likely Old and Likely New cue conditions was greater for shallow tests than deep tests. Additionally, the three way interaction between Levels of Processing, Cue Type, and Feedback was not significant ($F(1,32)=0.094$, $MSe = 0.06$, $p = 0.76$), suggesting that feedback did not influence the difference in criterion shifts for shallow and deep tests. Although comparing criteria under different levels of accuracy can be problematic, we are only interpreting the interaction in terms of the absolute difference in criteria between the two-cueing conditions for deep and shallow test items. In other words, we are only interpreting the absolute shift and not making direct comparisons about relative criteria placement across different accuracies. Thus, these results suggest that participants' absolute shifts in criteria are greater under conditions where memory performance is lower which confirms that the cues are being used in relation to the quality of internal recognition evidence. Furthermore, these absolute shifts in criteria are not feedback dependent, again demonstrating the fact that the use of cues does not seem to benefit from feedback based learning.

Although judgments for deep test items are less influenced by cues than shallow test items, one might argue that participants' responses just default to the cue, especially on shallow test items. If this was the case and participants were to always agree with the available cue, then we would expect a correct response rate of zero for invalidly cued items (i.e. cue Likely New for a target or Likely Old for a lure). In other words, if participants were always following the cue, they would always follow an incorrect cue and fail to respond accurately on any trial where the cue was incorrect. However, this was not the case. During both shallow and deep tests, the 95% confidence intervals for invalidly cued trial types clearly excluded zero (See Table 3).

Individual differences in efficacy of cue use

Although on average accuracy benefited from cueing, there were large individual differences in the degree of improvement. As noted in the introduction, the effective use of external cues may critically depend upon metacognitive awareness. To examine the role of metacognition we used the gamma index, which captures the correspondence between changes in subjective confidence and changes in accuracy at the trial-by-trial level for each participant (Nelson, 1984). Because gamma has a restricted range, unlike the accuracy measure d', we used the logit transformation of gamma (G*) to improve its scale properties (Benjamin & Diaz, 2008).

If metacognitive monitoring plays a role in cue utilization skill and is not a simple alternative measure of baseline observer accuracy, then hierarchical regression analysis should demonstrate that it makes a significant contribution to cued performance while baseline performance has been appropriately partialled from the data. In other words, we examined if metacognitive monitoring explains any unique variance in cued performance that is non-overlapping with baseline recognition skill. Using a hierarchical regression analysis, we examined the contribution of metacognition to cued accuracy (d') by entering feedback condition (dummy coded) and baseline (uncued) recognition accuracy as predictors in Step 1. Next, we examined whether metacognitive monitoring made a contribution beyond these factors by entering each participant's G* as an additional predictor in Step 2. Critically, G* was calculated from baseline performance and is therefore a measure of metacognitive monitoring in the complete absence of cues. Table 4 shows the results of the two hierarchical regressions that were separately conducted for the shallow and deep test lists. For the shallow test, in Step 1 baseline accuracy was a significant predictor ($b=0.60$, $t(30)=5.02$, $p<0.001$) of cued accuracy, while feedback group was not ($b=-0.04$, $t(30)=-0.47$, $p=0.64$). It is not surprising that subjects with high accuracy in the uncued condition would also have high accuracy under cueing. However, the clear absence of any contribution of the feedback variable serves to again underscore the fact that the provision of feedback has no appreciable influence on the manner in which participants use the cues (see ANOVA results above). Entering G* in Step 2 explained an additional 7.37% variance ($F(1,30)=4.72$, $p=0.04$). When the hierarchical regression was repeated for the deep test list condition, a similar pattern emerged where in Step 1 baseline accuracy was a significant predictor of cued accuracy ($b=0.77$, $t(29)=8.40$, $p<0.001$) and feedback group was not ($b=-0.16$, $t(29)=-1.28$, $p=0.21$) , and during Step 2 G* accounted for an additional 5.08% of unique variance in cued performance ($F(1,29)=6.00$, $p=0.02$).

The regression analyses demonstrate that cued performance is linked with baseline accuracy but more importantly, that after controlling for baseline performance, metacognitive monitoring is a significant predictor of cued performance gains. Again, the provision of feedback had little influence on the effective use of the external cues. These results hold for both shallowly and deeply encoding items. In summary, Experiment 1 demonstrates that participants are able to benefit from external cues for both deep and shallow test items, individual differences in cued utilization performance are, in part, related to metacognitive monitoring ability, and feedback does not improve cue utilization performance.

Experiment 2

For Experiment 2, we wanted to replicate our results from Experiment 1 and examine if participants are able to effectively differentiate between cues of differing validity (65% and 85% predictive). We again also examined whether cued performance is, in part, dependent on metacognitive monitoring ability.

Participants

Experiment 2 included 38 Washington University students (average age = 21.5, 18 females) who were paid $20 for participation. Three subjects were removed due to chance performance (d'<0.19) leaving 35 subjects for analyses. All participants provided informed consent in accordance with the University's Institutional Review Board

Materials and Procedure:

Testing was self-paced with observers entering their responses via keyboard, and presentation and timing controlled via Matlab's Psychophysics Toolbox (version 3.0.8) (Brainard, 1997; Pelli, 1997). For each participant, words were randomly selected from a 1216 item pool with an average of 7.09 letters, 2.34 syllables and Kučera -Francis frequency of 8.85.

Participants completed four study/test cycles (100 study items each) during which the encoding task was syllable-counting (1,2,3 or more syllables?). During the recognition test a total of 160 (80 old, 80 new) words were preceded by a probabilistic mnemonic cue (Likely Old or Likely New) one second before the word probe appeared. Cue predictability varied for this experiment where half the cues were 65% predictive (40 old, 40 new) and half the cues were 85% predictive (40 old, 40 new). Subjects were clearly informed of the two different cue validities. The 65% predictive cues were presented in a smaller blue font with the numbers 65 appearing next to the cue. The 85% predictive cues were presented in a larger yellow font with the numbers 85 appearing next to the cue. Instructions stated, “Cues that are 65% correct will give you the correct answer about 6 out of 10 times. Cues that are 85% correct will give you the correct answer about 8 out of 10 times. Use the cues to help increase your performance.” In addition to the cued trials, there were 40 (20 old, 20 new) baseline uncued trials intermixed in the test phase, with participants notified that some portion of the probes would be presented without anticipatory cues. Following each recognition decision subjects performed a confidence rating on a 6-point scale ranging from 50% (guessing) to 100% (certain), which was then followed by corrective feedback for half the participants.

Results and Discussion

Does accuracy improve with provision of cues?

To asses gains in accuracy (d’), we used a 3 X 2 mixed ANOVA with a repeated measures factor of Cue Condition (uncued baseline, 65% predictive cue, 85% predictive cue) and a between subjects factor of Feedback (present or absent). Results revealed a significant main effect of Cue Condition ($F(2,66)=44.13$, $MSe = 0.05$, $p <0.001$), no significant effect of feedback ($F(1,33)=1.18$, $MSe=0.45$, $p=0.28$), and no significant interaction between Cue

Condition and Feedback (F(2,68)=0.54, MSe=0.05, p=0.58). Follow up post hoc tests on the main effect of Cue Condition demonstrated that relative to baseline, there was a significant increase in performance on 85% predictive cued trials (MSe=0.04, p<0.001) and only a numeric improvement on 65% predictive cued trials (MSe=0.05, p=0.17) (Table 1). Overall, these results demonstrate that participants can benefit from the use of cues, even when two differing levels of cue predictability are intermixed. It is not necessarily surprising that performance does not significantly improve with the 65% predictive cues, since these cues are not highly accurate. However, replicating the results from Experiment 1, we find that when cues are highly predictive participants are able to improve their performance. Furthermore, we again see that cueing benefit does not depend on the provision of corrective feedback.

Reactivity to Cues

We wanted to examine if observers were shifting criteria more vigorously during highly predictive cues (85%), since this pattern should result if participants consider the relative predictability of the two cue levels. That is, since the 85% cues are accurate more often we would expect participants responding to be more highly influenced by these more accurate cues. With criteria measure C as our dependent variable we ran a 2 X 2 X 2 mixed ANOVA with repeated measures of Cue Type (Likely Old vs. Likely New) and Cue Condition (65% predictive cue, 85% predictive cue), and a between subjects factor of Feedback (absent or present). Importantly, there was a significant interaction between Cue Type and Cue Condition (F(1,33)=29.67, MSe=0.02, p<0.001), showing a greater difference in criterion shifts for 85% predictive cues than 65% predictive cues (Table 2). As before, the

3-way interaction between Cue Type, Cue Condition, and Feedback was not significant (F(1,33)=0.007, MSe=0.02, p=0.94). These results suggest that participants are in fact more influenced by high predictability cues than low predictability cues, and this relationship is not affected by feedback. Furthermore, as in Experiment 1, it is clear that participants do not just default to the cue since the 95% confidence intervals for invalidly cued trial types all excluded zero (Table 3).

Individual differences in efficacy of cue use

Although participants as a whole increased their cued performance relative to uncued baseline performance, there were once again large individual differences in cueing benefit. We wanted to replicate results from Experiment 1 and demonstrate that metacognitive monitoring contributes unique variance to cued performance above and beyond baseline performance. To examine this, we ran a separate hierarchical regression analysis on 65% predictive cued performance and 85% predictive cued performance with baseline uncued recognition accuracy (d') and feedback as predictors in Step 1, and metacognitive monitoring (G*) as a predictor in Step 2 (Table 6). Because the two different cue predictabilities were intermixed with baseline uncued trials, both analyses use the same measure for baseline recognition as well as metacognitive monitoring (which is again determined from baseline confidence reports). In Step 1 for 65% predictive cued performance, baseline accuracy was a significant predictor (b=1.04, t(32)=9.70, p<0.001) while feedback was not (b=-0.10, t(32)=-1.16, p=0.25). After controlling for baseline performance and feedback, metacognitive monitoring explained an additional 7.70% of the variance in cued performance (b=0.78, t(31)=3.74, p<0.001). Similar results were also found when using 85% predictive cued performance, where again in Step 1 baseline performance was a significant predictor (b=0.59, t(32)=4.50, p<0.001) and feedback was not (b=-0.18, t(32)=-1.73, p=0.09) , while in Step 2 metacognitive monitoring explained an additional 7.74% of the variance in cued performance (b=0.64, t(31)=2.22, p=0.03). These results demonstrate that although uncued baseline recognition skill is related to cued performance, there is additional unique variance explained by metacognitive monitoring ability. These results hold for both 65% and 85% predictive cues, and do not seem to be affected by feedback. Thus, replicating results from Experiment 1, we find that metacognitive monitoring is a significant predictor in cued performance above and beyond baseline accuracy, and corrective feedback does not seem to influence this relationship.

General Discussion

Our study examines the integration of external recommendations and internal memory evidence when provided with a reliable source of information. Prior studies examining external influences on memory generally have a confederate intentionally provide misinformation on a subset of trials, while the participant is led to believe that he/she studied the same material as the confederate. The overall finding from this memory conformity research is that people's decisions are in fact influenced by others' responses (Allan & Gabbert, 2008; Axmacher et al., 2010; Betz & Skowronski, 1996; Meade & Roediger, 2002; Reysen, 2005; Roediger et al., 2001; Schneider & Watkins, 1996; Walther et al., 2002; Wright et al., 2000; Wright et al., 2005; Wright et al., 2008). These prior studies focus on the negative aspect of conformity, mainly that participants' performance is decreased when given inaccurate external information. Such implications from memory conformity research are especially important when the goal is to minimize external influences, such as eyewitness testimony situations where the legal system wishes to preserve the original fidelity of the observer's remembrances. However, most our recognition decisions are not made in the context of the legal system or in the context of deceptive others. Generally, our goal is to maximize accuracy and, in the presence of useful sources of external information, this goal is achieved by judiciously integrating external influences with internal memory evidence.

Our study examined whether observers are able to increase performance when given predictive external recommendations and whether this skill is influenced by feedback and related to individual differences in metacognitive awareness. Under both shallow and deep encoding conditions (Experiment 1) and varying levels of cue validity (Experiment 2), participants elevated their performance on cued trials relative to baseline uncued trials, demonstrating that observers can judiciously incorporate known reliable external recommendations. When examining individual differences in cued performance, we found that after controlling for baseline recognition skill and feedback, metacognitive monitoring was a significant predictor of cued accuracy. Critically, this metamnemonic process is, in part, independent of memory retrieval or baseline recognition performance. Prior studies have assessed whether metamnemonic beliefs influence the degree of conformity (Betz & Skowronski, 1996; Gabbert et al., 2007; Schneider & Watkins, 1996; Walther et al., 2002; Wright et al., 2000), but these studies generally did not assess individual differences in metacognitive awareness or its contribution when observers are attempting to capitalize on external sources of information. Our study suggests that individual differences in monitoring may be relevant to assess in future memory conformity research as well as other studies assessing criterion shifts. Additionally, it would be interesting to determine if the ability to incorporate cues, and the role of metacognitive monitoring in this process, changes across development. For example, given prior aging research suggesting behavioral inhibition deficits in healthy older adults (Hasher & Zacks, 1988), it may be the case that older adults may tend to over rely on external recommendations as opposed to judiciously incorporating them with internal memory evidence. Furthermore, we can examine whether this skill is domain general by assessing if external recommendations have similar effects on both recognition and perceptual decisions.

The individual differences we found in the ability for observers to capitalize on external recommendations also has implications for Signal Detection models that assume a likelihood ratio decision axis. To the extent that observers actually have a decision axis akin to likelihood ratios, there are clearly considerable differences in the quality of it across observers. Additionally, the lack of any effect of feedback in the ability to utilize recommendations seems to suggest that corrective feedback does not result in updating of these representations.

We were surprised that corrective feedback did not improve the extent to which participants benefitted from external cueing, nor did it influence the degree to which metacognitive monitoring predicted cued performance. These results may seem puzzling since feedback could potentially inform participants about their subjective performance accuracy and help them respond ideally. Prior work on feedback in recognition memory suggests it does not improve recognition accuracy (Kantner & Lindsay, 2010), but that it is sometimes critical for observers to realize that a shift of the criterion may be appropriate or useful (Estes & Maddox, 1995; Rhodes & Jacoby, 2007; Verde & Rotello, 2007). The key difference between prior work using feedback and the current study, is that the in the former feedback is typically used to alert the subject to some experimental manipulation that should ideally induce a criterion shift. For example, in Rhodes and Jacoby (2007) base rates of items were correlated with screen location such that words presented on one side were more likely to be targets and should result in more liberal responding relative to words presented on the other side of the screen. In Verde and Rotello (2007), the strength of old items was manipulated through repetition, where old items on the first half of the test were strong (repeated 4 times at study) and should elicit conservative responding, while items on the second half of the test were weak (repeated only once at study) and should elicit more liberal responding. The key commonality across these studies is that the feedback appeared critical in order for the subjects to realize that responding similarly to the two locations (Rhodes & Jacoby, 2007) or similarly in the two test halves (Verde & Rotello, 2007) was not ideal because the overall distributions of targets or the average target strength differed across locations or test periods. In the current study however, the question was not whether observers would realize that the external cues were potentially useful, since this information was already provided. Instead, the key question was whether the feedback would increase the ability of the observers to optimally integrate the cues into their judgments. In this context, feedback could inform participants whether the degree of criteria shifts under cueing are appropriate to result in improved performance. Unfortunately, this was not the case. Furthermore, one may also suspect that feedback could increase cueing benefits by improving metacognitive monitoring ability. However, our current experiments suggest this is not the case since feedback actually numerically lowered monitoring scores (data not shown), perhaps suggesting limits on the plasticity of metacognitive monitoring of recognition content.

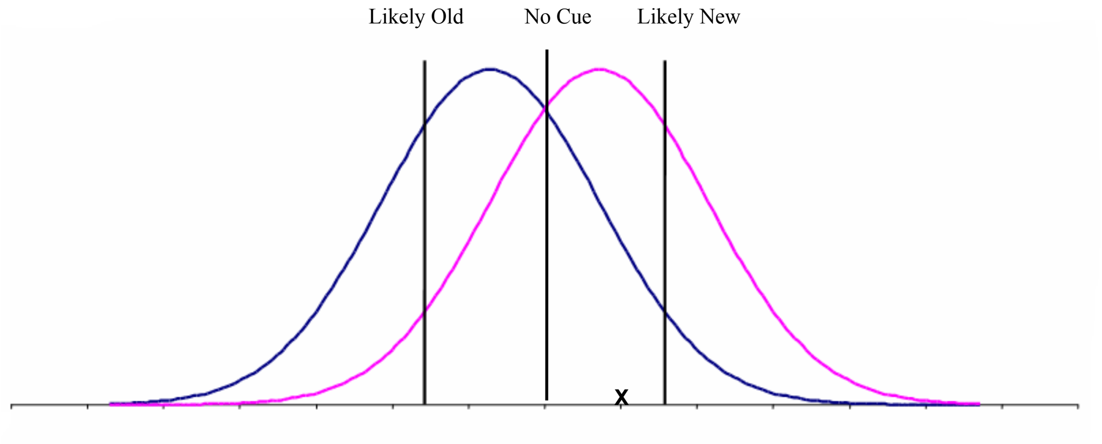
Conclusion

In summary, our current study demonstrates that people are able to improve their recognition performance when using a known, reliable source of external information. Furthermore, the ability to improve from external information is, in part, dependent upon metacognitive monitoring ability. Finally, the ability to benefit from external cues in recognition memory does not appear to be dependent on corrective feedback.

Footnote

1.) Since there is greater information available under conditions where both cues and internal evidence are available, compared to conditions in which only internal evidence is available, the ideal observer must necessarily improve performance under the former. Under the likelihood ratio Signal Detection model, this is achieved by advantageous shifting of the criterion on a trial-by-trial basis. When no external evidence is available, the hit rate and false alarm rate are weighted equally since the probability of occurrence for each item type is equal. During cued trials, we must weight response rates by the disproportionate likelihood of encountering an Old item. On trials where a 75% predictive cue reads Likely Old, Old items are presented 3 times more often than new items and we must weight the hit rate by 75% and the false alarm rate by 25%. Although an ideal observer would respond Old more often in this condition (Figure 1B), the increase in hit rate is weighted much more heavily than the increase in false alarm rate. When the cue reads Likely New the hit rate must be weighted by 25% and the false alarm rate should be weighted by 75%, since new items are presented 3 times more often than old items. Again, although an ideal observer responds New more often under this condition (Figure 1C), the increase in correct rejections is weighted much more heavily than the increase in misses. To determine the overall hit and false alarm rate under cueing, we would sum the weighted hit rates under the Likely Old and Likely New cue and do the same for the false alarms rates. Thus, ideally we observe overall increases in accuracy when cues are available, but note that discrimination ability itself is not changing. For example, during a vision test (e.g. discriminating between X's and Y's) you may have a friend with perfect vision telling you all the correct answers. The aid of your friend results in you increasing your performance, but your perceptual acuity does not change. Analogously, predictive external cues in recognition memory can improve overall performance, but discrimination ability itself remains the same.

Figure 1



A) No Cue	1:3	1:2	**1:1**	2:1	3:1
B) Likely Old Cue	**1:1**	1.5:1	3:1	6:1	9:1
C) Likely New Cue	1:9	1:6	1:3	1:1.5	**1:1**

The figure above depicts how optimal criteria location (odds of 1:1) shifts as a function of external recommendations under a likelihood ratio model of recognition memory. The x-axis represents likelihood ratios, and is computed by determining the probability density of the target distribution relative to the probability density of the lure distribution for a specific location. Notice that the response to the value indicated by an X (2:1 odds with No Cue), changes depending on the cueing condition. **A.)** Under conditions with no external cue, the ideal criteria location is in the center of the overlap of the two distributions **B.)** Under conditions with a Likely Old cue, the ideal observer would multiply the likelihood axis by cue predictability (3:1). Evidence values and ideal criteria shift to the left. **C.)** Under condition with a Likely New Cue, the ideal observer would multiply likelihood axis by cue predictability (1:3). Evidence values and ideal criteria shift to the right.

Table 1 Experiment 1 and 2 accuracy (d’) under uncued baseline and cued conditions (standard deviations in parenthesis).

		Baseline	Cued
Exp 1	Shallow	0.91 (0.38)	1.17 (0.36)
	Deep	2.09 (0.66)	2.32 (0.62)
Exp 2	65% Predictive	1.20 (0.40)	1.30 (0.48)
	85% Predictive		1.65 (0.39)

Table 2 Experiment 1 and 2 criteria (C) under Likely Old and Likely New cues (standard deviations in parenthesis).

		Likely Old Cue	Likely New Cue
Exp 1	Shallow	-0.37 (0.40)	0.51 (0.34)
	Deep	-0.47 (0.38)	0.007 (0.35)
Exp 2	65% Predictive	-0.27 (0.34)	0.27 (0.26)
	85% Predictive	-0.42 (0.43)	0.41 (0.30)

Table 3 Experiment 1 and 2 hit and correct rejections rates under invalid cues (95% confidence intervals in parenthesis).

		Invalid HR	Invalid CR
Exp 1	Shallow	0.47 (0.41-0.53)	0.52 (0.45-0.58)
	Deep	0.83 (0.79-0.87)	0.70 (0.65-0.75)
Exp 2	65% Predictive	0.63 (0.58-0.68)	0.60 (0.55-0.66)
	85% Predictive	0.56 (0.50-0.62)	0.55 (0.49-0.62)

Table 4 Experiment 1 hierarchical regression analysis with cued accuracy (d') as the dependent variable for shallow and deep encoding.

Shallow Encoding						
	Variable	**B**	**Std Error of B**	**P**	R^2	ΔR^2
Step 1	Uncued Recognition (d')	0.60	0.12	<0.001		
	Feedback	-0.04	0.09	0.64	0.47	
Step 2	Metacognitive monitoring (logit gamma)	0.95	0.43	0.04	0.55	0.07
Deep Encoding						
	Variable	**B**	**Std Error of B**	**P**	R^2	ΔR^2
Step 1	Uncued Recognition (d')	0.77	0.09	<0.001		
	Feedback	-0.16	0.12	0.21	0.71	
Step 2	Metacognitive monitoring (logit gamma)	0.69	0.28	0.02	0.76	0.05

Table 5 Experiment 2 hierarchical regression analysis with cued accuracy (d’) as the dependent variable for 65% predictive cues and 85% predictive cues .

65% Predictive Cues						
	Variable	**B**	**Std Error of B**	**P**	R^2	ΔR^2
Step 1	Uncued Recognition (d’)	1.04	0.12	<0.001		
	Feedback	-0.10	0.08	0.25	0.75	
Step 2	Metacognitive monitoring (logit gamma)	0.79	0.21	<0.001	0.82	0.07
85% Predictive Cues						
	Variable	**B**	**Std Error of B**	**P**	R^2	ΔR^2
Step 1	Uncued Recognition (d’)	0.59	0.13	<0.001		
	Feedback	-0.18	0.10	0.09	0.43	
Step 2	Metacognitive monitoring (logit gamma)	0.64	0.29	0.03	0.51	0.08

References

Allan, K., & Gabbert, F. (2008). I still think it was a banana: Memorable 'lies' and forgettable 'truths.' *Acta psychologica*, *127*(2), 299–308.

Asch, S. E. (1955). Opinions and Social Pressure. *Scientific American*, *193*(5), 31–35.

Axmacher, N., Gossen, A., Elger, C. E., & Fell, J. (2010). Graded effects of social conformity on recognition memory *PloS one*, *5*(2), e9270. doi:10.1371/journal.pone.0009270

Benjamin, A. S., & Diaz, M. (2008). Measurement of Relative Metamnemonic Accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 73–94). Psychology Press.

Betz, A., & Skowronski, J. (1996). Shared realities: Social influence and stimulus memory. *Social Cognition*, *14*(2), 113–140.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Cox, J. C., & Dobbins, I. G. (2011). The striking similarities between standard, distractor-free, and target-free recognition *Memory & cognition*, *39*(6), 925–940. doi:10.3758/s13421-011-0090-3

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition *Journal of Experimental Psychology Learning, Memory, and Cognition*, *21*(5), 1075–1095.

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*(3), 319–331. doi:10.1016/j.actpsy.2006.03.009

Glanzer, M., Hilford, A., & Maloney, L. T. (2009). Likelihood ratio decisions in memory:

Three implied regularities. *Psychonomic bulletin & review*, *16*(3), 431–455. doi:10.3758/PBR.16.3.431

Green, D. M., & Swets, J. A. (1989). *Signal Detection Theory and Psychophysics* (p. 521). New York: Wiley.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 22, pp. 193–225). San Diego, Ca: Academic Press.

Healy, A., & Kubovy, M. (1978). The effects of payoffs and prior probabilities on indices of performance and cutoff location in recognition memory. *Memory & cognition*, *6*(5), 544–553.

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

Kantner, J., & Lindsay, D. S. (2010). Can corrective feedback improve recognition memory *Memory & cognition*, *38*(4), 389–406. doi:10.3758/MC.38.4.389

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: evidence against a general metacognitive ability *Memory & cognition*, *28*(1), 92–107.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory*. A user's guide (2nd ed.). Lawrence Erlbaum.

Meade, M. L., & Roediger, H. (2002). Explorations in the social contagion of memory *Memory & cognition*, *30*(7), 995–1009.

Metcalfe, J, & Dunlosky, J. (2009). Metamemory. In H. Roediger & J. H. Byrne (Eds.), *Learning and memory: a comprehensive reference.* (Vol. 2, pp. 349–362). Oxford: Elsevier.

Nelson, T. (1990). *Metamemory: A theoretical framework and new findings*. (G. Bower, Ed.) *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125–173). San Diego, CA: Academic Press.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological bulletin*, *95*(1), 109–133.

Pastore, R. E., Crawley, E. J., Berens, M. S., & Skelly, M. A. (2003). “Nonparametric” A' and other modern misconceptions about signal detection theory *Psychonomic bulletin & review*, *10*(3), 556–569.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Reysen, M. (2005). The effects of conformity on recognition judgements. *Memory (Hove, England)*, *13*(1), 87–94. doi:10.1080/09658210344000602

Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: effects of base rate, awareness, and feedback *Journal of Experimental Psychology Learning, Memory, and Cognition*, *33*(2), 305–320. doi:10.1037/0278-7393.33.2.305

Roediger, H. L., Meade, M L, & Bergman, E. T. (2001). Social contagion of memory *Psychonomic bulletin & review*, *8*(2), 365–371.

Schneider, D., & Watkins, J. M. (1996). Response conformity in recognition testing. *Psychonomic bulletin & review*, *3*, 481–485.

Turner, B., Van Zandt, T., & Brown, S. (2011). A Dynamic Stimulus-Driven Model of Signal Detection. *Psychological review*, 1–102.

Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory *Memory & cognition*, *35*(2), 254–262.

Walther, E., Bless, H., Strack, F., Rackstraw, P., Wagner, D., & Werth, L. (2002). Conformity effects in memory as a function of group size, dissenters and uncertainty. *Applied Cognitive Psychology*, *16*(7), 793–810. doi:10.1002/acp.828

Wright, D. B., Mathews, S. A., & Skagerberg, E. M. (2005). Social Recognition Memory: The Effect of Other People's Responses for Previously Seen and Unseen Items *Journal of Experimental Psychology: Applied*, *11*(3), 200–209. doi:10.1037/1076-898X.11.3.200

Wright, D. B., Self, G., & Justice, C. (2000). Memory conformity: exploring misinformation effects when presented by another person *British journal of psychology (London, England : 1953)*, *91 (Pt 2)*, 189–202.

Wright, D. B., Gabbert, F., Memon, A., & London, K. (2008). Changing the criterion for memory conformity in free recall and recognition. *Memory (Hove, England)*, *16*(2), 137–148. doi:10.1080/09658210701836174