Running head: VISUAL WORKING METAMEMORY

Looking Inwards and Back: Realtime Monitoring of Visual Working Memories

Jordan W. Suchow
Department of Psychology, Harvard University
Institute of Cognitive and Brain Sciences, UC Berkeley

Daryl Fougnie
Department of Psychology, Harvard University
Department of Psychology, NYU Abu Dhabi

George A. Alvarez
Department of Psychology, Harvard University

Abstract

Confidence in our memories is influenced by many factors, including beliefs about the perceptibility or memorability of certain kinds of objects and events, as well as knowledge about our skill sets, habits, and experiences. Notoriously, our knowledge and beliefs about memory can lead us astray, causing us to be overly confident in eyewitness testimony or to overestimate the frequency of recent experiences. Here, using visual working memory as a case study, we designed two experiments that strip away all these potentially misleading cues, requiring observers to make confidence judgments by directly assessing the quality of their memory representations. We show that individuals can monitor the status of information in working memory as it degrades over time. Our findings suggest that people have access to information reflecting the existence and quality of their working memories, and furthermore, that they can use this information to guide their behavior.

Looking Inwards and Back: Realtime Monitoring of Visual Working Memories

Metamemory is an awareness of our memories and the systems that store them. We use metamemory to determine that we are uncertain (“I can’t remember where I parked my car”), to ask for a reminder (“When’s that appointment, again?”), and to form beliefs about our ability to remember certain kinds of information (“I’m good with names”) (Flavell & Wellman, 1976). Metamemory is often studied in the context of long-term memory, where it is invoked to explain phenomena such as tip-of-the-tongue states and the feeling of knowing. Healthy individuals have a rich set of metamnemonic skills that guide learning, decision-making, and action (Metcalfe & Shimamura, 1994). Neurological diseases, such as Alzheimer’s and Korsakoff’s syndrome, adversely affect metamemory judgments, causing a mismatch between what is remembered and what is believed to be remembered (Pannu & Kaszniak, 2005).

For better or for worse, judgments of confidence in our memories are influenced by many factors. These include general knowledge in the form of beliefs about ourselves and what we find memorable, as well as more specific knowledge derived from our previous experience with the task at hand (Koriat, 1997; Schwartz, 1994; Schwartz, Aaron, & Bjork, 1997). However, the cognitive mechanisms underlying metamnemonic judgments are poorly understood.

Looking towards research in other areas of metacognition, where a variety of confidence mechanisms have been explored in detail, may provide a clue about the workings of metamemory. For example, in the case of perceptual discriminations, one simple mechanism for judging confidence is to use visual cues associated with uncertainty (e.g. faintness and blur), alone or in combination, as a proxy for confidence (Barthelmé & Mamassian, 2010). Then, when asked to identify an object that appears blurry or faint, an observer using this mechanism will report having low confidence because blurriness and faintness are stimulus features typically associated with uncertainty. Importantly, cue-based mechanisms like this one draw on *static* information about the stimulus and decision-maker, rather than directly accessing an internal representation or decision-making process. In contrast, an alternative class of mechanisms has been proposed that can compute *realtime* measures of perceptual confidence (Kepecs, Uchida, Zariwala, & Mainen, 2008). Realtime mechanisms are notable because, rather than relying on externally observed cues, they monitor internal states as they change over time (Kepecs et al., 2008). These monitored states may be those of decision variables associated with the task, or those of underlying representations that store uncertainty explicitly — e.g., as probability distributions over past states of the environment (Barthelmé & Mamassian, 2010).

Here, we use visual working memory as a case study for exploring evaluations of confidence in memory. Metamemory in visual working memory, change detection, and working memory more broadly has been demonstrated using a variety of paradigms (Bunnell, Baken, & Richards–Ward (1999); Levin, Momen, Drivdahl, & Simons, 2000; Amichetti, Stanley, White, & Wingfield, 2013; Bona, Cattaneo, Vecchi, Soto, & Silvanto, 2013; Bona & Silvanto, 2014; Vandenbroucke, Sligte, Barret, Seth, Fahrenfort, & Lamme, 2014). Here, we designed two experiments that isolate realtime mechanisms underlying the evaluation of confidence in memory.

Experiment 1 is a variant of a popular test of visual working memory in which participants view a display of colorful dots and then, shortly thereafter, report the color of a dot selected for them at random (Wilken & Ma, 2004). In our variant, however, instead of requiring participants always to report the color of a randomly-selected dot ("random" condition), they were sometimes afforded the opportunity to report the color of the object they remembered best ("best" condition). Choosing the best-remembered object requires an inward-looking comparison of the relative quality of multiple memories, and is a within-trial analogue to the opt-out procedure used extensively in studies of human and animal metacognition (Smith et al., 1995; Smith, Shields, Schull, & Washburn, 1997; Tanaka & Funahashi, 2012). To strip away nearly all the usual sources of metamnemonic information — general knowledge, stimulus-based cues, and time-based fluctuations in attention and arousal — we compared memory for an object in a display when it was chosen by the participant as their best remembered object to when that same item from the same display was chosen by the experimenter at random. This procedure enables us to isolate a form of monitoring whereby an individual tracks the status of a memory as it degrades over time.

Experiments 2a, 2b, and 2c capitalize on *directed remembering*, a process by which information is prioritized in memory for later access (c.f. directed forgetting, Muther, 1965; Bjork, Laberge, & Legrand, 1968; Williams, Hong, Kang, Carlisle, & Woodman, 2012; Williams & Woodman, 2012). We extend the phenomenon of directed remembering (Exp. 2a) beyond external cues to a kind of *self*-directed remembering (Exp. 2b) whereby participants shift the balance of maintenance according to an internally generated metamemory signal. We show that metamemory judgments reflect these trial-specific changes in the underlying memory representations caused by the redirection of maintenance (Exp 2c).

Experiment 1: Method

Logic of the task: isolating the contribution of realtime monitoring.

Participants were asked to remember the colors of a set of colorful dots, and then either to report the color of a randomly selected dot, or to make an inward-looking decision by choosing the dot they remembered best and reporting its color. Because our goal was to isolate the contribution of realtime monitoring to this decision, the experimental procedure combined multiple techniques to eliminate confounding sources of metamnemonic cues:

Stimulus-based cues. Of principal interest was whether memory would be more accurate for the best-remembered object than for a randomly selected item. However, the best-remembered object might be preferred for reasons that do not require a realtime assessment of memory quality. For example, a participant may prefer a particular color, say red, and pay more attention to it. Or perhaps they find it more memorable, preferring to report the color of red objects whenever the chance arises. This is a form of metamemory, but it does not reflect realtime monitoring. To eliminate display factors like these, we used a double-pass procedure (Burgess & Colborne, 1988; J. M. Gold, Murray, Sekuler, Bennett, & Sekuler, 2005; Green, 1964) where participants encounter each display (i.e., a particular color and arrangement of dots) twice, once in each of two sessions, separated by a few days. Displays used in the "choose the best" condition in the first session were reused in the "random" condition in the second session, and vice versa. We compare memory performance for a particular object when it was chosen as the object that was best remembered (first session) to when it was randomly selected by the experimenter (in the second session). Any advantage for the preferred object cannot depend on stimulus-based factors, which are held constant across the conditions.

Tradeoffs in encoding or maintenance. When viewing the stimulus, a participant's attention might wander due to either an explicit strategy or accidental drift, causing one object to be encoded more vigorously than another. This could also happen during maintenance, if the participant were to shift priority from one object to another after the stimulus has disappeared. Such imbalances, if known to the participant, could be used as a proxy for memory fidelity: an ignored object is unlikely to be remembered well. Because our goal is to determine whether participants can access the quality of internal representations, independent of other factors such as knowledge about which objects were given the most resources, we designed the experiment to minimize tradeoffs and then performed a separate tradeoff detection procedure once the experiment was over:

Our design made tradeoffs costly by interleaving two types of trials in random order. On half the trials, participants reported the color of the dot they remembered best, while on the other half, they reported the color of a dot selected at random by the experimenter. Interleaving the trial types encourages participants to remember all the dots, because they do not know which dot will be tested.

The above procedure mitigates the tradeoffs, but it is possible that they were still present. To determine whether tradeoffs occurred, we used an additional procedure: After reporting either the best-remembered object or a randomly-selected one, the participant was then asked to report the color of a second dot on the display, selected at random. These two reports are together used to detect tradeoffs with the detection procedure introduced in Fougnie et al. (2012), which relies on the fact that, if there are tradeoffs, there will be dependencies in the measured quality of representations of objects on a display: if one object is remembered particularly well, it comes at the expense of the others. Therefore, the detection procedure compares performance for the first-reported object in two conditions: where the absolute error for the second-reported item was (1) above or (2) below the median absolute error across all second-reports. A tradeoff is revealed by first reports being significantly more accurate when second reports are worse (above-median error) than when they are better (below-median error).

Fluctuations in attention or arousal. Attention, arousal, and effort can fluctuate from moment to moment (Kahneman, 1973). Most studies of metamemory ask for ratings or judgments about a particular memory at a particular moment, and so momentary fluctuations can affect performance and therefore contribute to metamnemonic decisions. In our task, we asked participants to make a relative judgment about the quality of simultaneously held memories, such that the confidence judgment could never depend on the overall state of attention or arousal, which would apply equally to all objects on the display. Similarly, we randomized the order of the two trial types, which prevents momentary fluctuations from systematically affecting one trial type over the other.

Implementation of the task

At the beginning of each trial, the participant fixated a small dot in the center of the screen. Then the stimulus (a set of three colorful dots) appeared for 600 ms. Next, the trial type was revealed to the participant through a display that contained a cue in each of the locations of the test objects. If it was a trial where the participant was asked to report a specific object, that object was highlighted as a filled circle among open circles. On the other hand, if it was a trial where the participant reported the best-remembered object, all the objects appeared filled in. Then a color wheel with all the possible colors appeared and the color of the best-remembered object was reported. Finally, the participant used a mouse-controlled cursor to select which object was best remembered. After this first report, the participant was asked to report the color of a second object selected at random from the two that remained. The reporting procedure was the same. Feedback was provided at the end of each trial. The feedback screen, which appeared for 1000 ms, showed the actual color (inner ring) and reported color (outer ring) for both of the tested objects (Fig. 1i). There were 200 trials in each of two sessions; half the trials probed a random object, half probed the best-remembered object, interleaved in a random order. The second session was identical to the first, with exactly the same displays, except that the condition assigned to each display was swapped. Thus the "randomly" probed objects in the second session were in fact the very same objects that, in the previous session, had been chosen as best-remembered.

Stimuli and presentation

Each dot had a radius of 0.4° of visual angle. They were arranged in a ring with a radius of 3.8° and centered on the display. The color of each dot was drawn uniformly from a circle cut out of the CIE 1976 $L^*a^*b^*$ color space, centered at $L = 54$, $a = 18$, $b = -8$, with the constraint that the magnitude of each display's mean hue vector was 0.35. This decreases grouping cues and reduces imbalances in appearance across displays. After the stimulus disappeared, there was a 900 ms retention interval. Stimuli were rendered by MATLAB with the Psychophysics toolbox (Brainard, 1997; Pelli, 1997), and presented on a 1920 × 1200 LCD screen at 60 Hz, 38 pixel/cm, positioned 60 cm from the participant.

Participants

Twelve people between the ages of 18 and 31 participated. They all had normal or corrected-to-normal visual acuity. The protocol, approved by the Committee on the Use of Human Subjects in Research under the Institutional Review Board for the Faculty of Arts & Sciences at Harvard University, was carried out in accordance with the provisions of the World Medical Association Declaration of Helinski.

Data analysis

To quantify memory performance, for each participant and condition we separately fit a variable-precision model to the data (Fougnie, Suchow, & Alvarez, 2012; van den Berg, Shin, Chou, George, & Ma, 2012). This model supposes that each object on the display is either remembered or forgotten, and that the quality with which objects are remembered can vary. To ensure that our findings do not depend on this particular choice of model, we also tested others. We considered a simpler fixed-precision model that did not allow memory quality to vary (see Online Supplemental Material) as well as an extension to it that allows for the possibility that the participant will "swap" items, erroneously reporting an item that was not the target (Bays, Catalao, & Husain, 2009). Analysis was performed with MemToolbox 1.0.0 (Suchow, Brady, Fougnie, & Alvarez, 2013). Analysis scripts and data are available as Online Supplementary Material.

Experiment 1: Results

We found that participants can use realtime monitoring to make metamnemonic judgments. Fig. 2 shows estimates of guessing rate (left panel) and precision (middle panel), averaged across participants. Individual participant results are shown (right panel) for items that were chosen as the best remembered (circles) versus those same items when they were selected at random (squares). Observers performed better in both guessing rate and memory precision when they chose to report the object, versus when the object was randomly selected. When asked to report the color of the best-remembered object, participants remembered it 92±2% (mean ± sem) of the time and with fidelity of 20.8±1°. When those same displays were presented in the second session and participants were forced to report the same object that they had previously picked, they performed worse, remembering it 71±3% of the time and with a fidelity of 23.8±2° (paired samples *t*-test, $t(11) = 7.5$, $p = 1.2\times10^{-5}$ and $t(11) = -2.5$, $p = 0.03$, respectively). This across-exposure worsening happened despite overall performance being comparable in the two rounds (difference of 0.6° in fidelity from the first to second round, paired samples *t*-test, $t(11) = 0.51$, $p = 0.62$; difference of 0.4% in guess rate, paired samples *t*-test, $t(11) = 0.14$, $p = 0.89$).

Using the tradeoff detection procedure described in Methods, we tested for tradeoffs in the encoding or maintenance of items, but found none (guess rate for above vs. below median split: 20.5 vs. 19.4%, paired samples *t*-test, $t(11) = -0.68$, $p = 0.51$; fidelity: 22.1° vs. 22.4°, $t(11) = 0.42$, $p = 0.68$;).

We performed additional analyses to determine whether our results are robust to different assumptions about the structure of visual memory representations. Specifically, we repeated the above analyses with the two-component mixture model introduced by Zhang & Luck (2008), the "swap" model introduced by Bays et al. (2009) and a 1-component model with no guessing. The results were comparable under all models (see Online Supplemental Material).

The results of Exp. 1 suggest that people have access to information reflecting the existence and quality of their working memories, and furthermore, that they can use this information to guide their behavior. However, an alternative reading of the results warrants a second look. Suppose that when participants select the object they remember best, they tend to report the first object that comes to mind and that this first-recalled object tends to be better remembered than the others. Participants like these would pass the test for metamemory described in Exp. 1 regardless of whether they had metamemory in actuality.

Exps. 2a, 2b, and 2c together provide a second, stronger test of realtime metamemory, one that circumvents the alternative reading of Exp. 1. The key is to leverage *directed remembering*, a dynamic process whereby maintenance is biased towards certain representations over others. By showing that metamemory judgments reflect these trial-specific shifts in the underlying memory representations caused by the redirection of maintenance.

Experiment 2a: Method

Logic of the task: directed remembering

Exp. 2a replicated the effect of directed remembering in visual working memory (e.g. Williams, et al., 2012). Participants were asked to remember the appearance of some objects and then to report what they remembered. Participants were sometimes given a cue early in the retention interval that signaled which object would later be tested. This gives the participant the opportunity to redirect maintenance accordingly.

Stimuli and presentation

The presented objects were cubes. Cubes were chosen because of their high complexity, which makes it possible to tax memory considerably while presenting only two objects (Alvarez & Cavanagh, 2004), which simplifies the design of Exps. 2a–c. Each cube had three visible sides, one white, one grey, and one black, viewed either from above or below, for a total of twelve possible configurations (Fig. 3). Cubes were positioned 5° to the left or right of a central fixation mark. The stimuli, adapted from Alvarez & Cavanagh (2004), were rendered by MATLAB with the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) and presented on a 1920 × 1200 LCD screen at 60 Hz, 38 pixel/cm, positioned 60 cm from the participant.

Participants

Eight people between the ages of 20 and 35 participated. They all had normal or corrected-to-normal visual acuity. The protocol, approved by the Committee on the Use of Human Subjects in Research under the Institutional Review Board for the Faculty of Arts & Sciences at Harvard University, was carried out in accordance with the provisions of the World Medical Association Declaration of Helinski.

Procedure

There were three conditions: 1, 2′, and 2. In condition 1, one object was presented on either the left or right side of fixation at random. In conditions 2′ and 2, two objects were presented, one to the left of fixation, the other to the right. The objects appeared for 300 ms and then disappeared. The retention interval was 4000 ms. In conditions 1 and 2, a cue appeared at the end of the retention interval in the location of one of the presented objects, selected at random. Critically, in condition 2′, the cue came earlier, 700 ms into the retention interval. In all conditions, the participant chose the cued object from a response screen containing all the possible objects. There were 25 trials per condition, ordered randomly.

Experiment 2a: Results

Fig. 4 shows performance on the task across the three conditions. As in previous studies of working memory, performance was considerably better with one object than with two (conditions 1 vs. 2, mean difference 0.51; paired samples *t*-test, $t(7) = 10.3$, $p = 1.74\times10^{-5}$). Critically, replicating the effect of directed remembering, performance was better with two objects when the cue came early than when it came late (condition 2 vs. 2′, mean difference 0.32; paired samples *t*-test, $t(7) = 6.17$, $p = 4.56\times10^{-4}$).

Experiment 2b: Method

Logic of the task: self-directed remembering

Exp. 2b extends directed remembering to *self*-directed remembering. Participants were sometimes given a cue to redirect maintenance to the best- or worst-remembered object. We compared the fidelity of memory after maintenance had been redirecting to a baseline where it was not.

Stimuli, presentation, and participants

All details matched those from Exp. 2a.

Procedure

Two objects appeared for 300 ms, one to the left of fixation, the other to the right of fixation, and then disappeared. The retention interval was 4000 ms. We manipulated two factors in a 3 × 2 design. The first factor was the valence of the cue — a high tone, a low tone, or a visually presented cue. Participants were instructed that when the tone was high, they were to decide which object was best remembered, press a keyboard button to select it, and then redirect maintenance to chosen object, assured that only it would be tested later. In contrast, when the tone was low, the same procedure was applied to the worst-remembered object. And when there was no tone, they maintained the visually cued object. The second factor was the timing of the cue — early (700 ms after the offset of the stimuli) or late (at the end of the retention interval). There were 25 trials per condition, ordered randomly.

Experiment 2b: Results

Fig. 5 shows performance on the task across the six conditions. Replicating the effect of directed remembering in Exp. 2a, performance was better when a randomly selected object was cued early than when it was cued late (mean difference 0.21; paired samples *t*-test, $t(7) = 3.33$, $p = 0.0126$). Critically, we also saw evidence of self-directed remembering. Performance was better when participants directed maintenance to the best-remembered object earlier than later (mean difference 0.17; paired samples *t*-test, $t(7) = 5.22$, $p = 0.0012$). Performance was also better when participants directed maintenance to the worst-remembered object earlier than later (mean difference 0.18; paired samples *t*-test, $t(7) = 2.57$, $p = 0.0368$). And, finally, performance was better when participants reported the best-remembered object than when they reported the worst-remembered object (mean difference 0.285, paired samples *t*-test, $t(15) = 7.62$, $p = 1.56\times10^{-6}$).

Experiment 2c: Method

Logic of the task: reevaluating metamemory decisions after self-directed remembering

Redirecting maintenance affects the fidelity of the prioritized and neglected memories. Critically, when maintenance is redirected to the worst-remembered object, there is sometimes a reversal of relative memory strength — what had once been the worst-remembered object is now best. Exp. 2c takes advantage of these reversals by allowing participants to revisit their decision from earlier in the trial about the relative fidelity of their memories for the objects. Participants performed a self-directed remembering task like that in Exp. 2b, but always redirected maintenance to the worst-remembered object. On 25% of trials, however, participants were told to revisit their decision and report whichever object they remembered best. If participants lack realtime metamemory, on these trials they would always make a selection that was consistent with their previous determination, choosing the other object — in their metacognitive mind, nothing had changed. If, however, participants have realtime metamemory, reversals will lead to a selection that is inconsistent with their previous determination, benefitting their performance on the task. An inconsistent choice that *benefits performance* can be due only to events that occurred during the maintenance interval, thus providing a stringent test of realtime metamemory.

Stimuli, presentation, and participants

All details matched those from Exps. 2a and 2b.

Procedure

Two objects appeared for 300 ms, one to the left of fixation, the other to the right of fixation, and then disappeared. After 700 ms, participants heard a low tone, decided which object was worst remembered, pressed a keyboard button to select it, and then redirected maintenance to chosen object. The retention interval was 4000 ms in total. On 50% of trials ("condition 1"), after the retention interval elapsed, a second low tone was played. Participants reselected the previously chosen object (i.e., the original worst-remembered object) by pressing the corresponding button on the keyboard, at which point the response screen appeared. Participants selected that object from the response screen. On 25% of trials ("condition 2"), after the retention interval elapsed, a high tone was played. Participants selected the previously unchosen object by pressing the other button on the keyboard, at which point a response screen appeared. Participants selected the previously unchosen object (i.e., the original best-remembered object) from the response screen. On 25% of trials ("condition 3"), after the retention interval elapsed, a medium tone was played. Participants now had the option to choose either object, whichever was currently best remembered, pressing the corresponding button on the keyboard and selecting the object from the response screen. The three conditions were randomly interleaved. There were 100 trials in total.

Experiment 2c: Results

Fig. 6 shows performance on the task across the three conditions. On trials where participants were allowed to revisit their decision about which object was remembered best, they reversed their decision on 91% of trials, which is significantly different from zero (one-sample *t*-test, $t(7) = 25.0$, $p = 4.18\times10^{-8}$) and from one (one-sample *t*-test, $t(7) = 2.62$, $p = 0.0342$). This led to better performance than when participants were required to report the object originally selected as the best (conditions 2 vs 3, mean difference 0.285, paired samples *t*-test, $t(7) = 5.15$, $p = 0.0013$).

Discussion

The results of these experiments suggest that realtime monitoring can be used to make judgments of confidence in working memory. In Experiment 1, we found that participants remembered an object's color more accurately when it had been chosen as the one they remembered best than when that same object, presented in the context of the same display, was selected at random by the experimenter. Even after eliminating other sources of metamnemonic information, such as stimulus-based cues and tradeoffs in encoding and maintenance, we found that observers were able to assess the quality of their memories in realtime and could use that information to guide their behavior. In Experiments 2a–c, we found that participants could track changes to the relative strength of their memories that had been altered by self-directed maintenance. Thus, the present results reveal a strategy that can monitor the fidelity of representations in visual working memory.

This form of metamemory requires access to information that indexes the quality of memories. Though it is unclear what mechanism provides realtime access to memories, research on uncertainty in decision-making may provide a clue:

A number of simple mechanisms have been proposed that might support realtime measures of confidence in perceptual judgments (Kepecs et al., 2008). These mechanisms involve accessing decision variables that contribute to a decision. For example, in a race model, where evidence simultaneously accumulates for each alternative choice (Gold & Shadlen, 2007), confidence can be estimated by measuring the difference in accumulated evidence for each alternative at the moment the choice is made. High confidence is appropriate when there is a big imbalance in accumulated evidence. Analogous mechanisms may be at play in the monitoring of visual working memories. For example, confidence in a memory could be computed by comparing the accumulated evidence for the winning decision (i.e., stimulus value) to the average of the others. Alternatively, monitoring may be accomplished through more indirect means, using a process akin to the availability heuristic. Specifically, suppose that less precise memories are more difficult to access (see Brady et al., 2013). Then, the participant can use a metamemory routine that tries to access a memory, terminating if nothing has been accessed after a fixed amount of time. Time-to-access then serves as a proxy for memory fidelity and can be used to inform confidence.

Whatever the mechanism, the present results demonstrate it is possible to access the current state of a memory and to use that information to guide behavior. The existence of realtime monitoring mechanisms has important implications not only for our understanding of metamemory, but also for theories of the representational format of visual working memory. Models need to consider the source of variation in working memories (e.g. Fougnie, et al., 2012; van den Burg et al., 2012) and provide an account for how participant's can access representational uncertainty in realtime.

Leading models of visual working memory assume that memory limits are determined purely by the availability of a limited commodity: once you run out of memory slots (Awh, Barton, & Vogel, 2007; Zhang & Luck, 2008) or memory resources (Alvarez & Cavanagh, 2004; Bays et al., 2009; Wilken & Ma, 2004), you can no longer store additional objects in memory. However, in addition to possible commodity-based limits, there is emerging evidence that it is also limited by interference, degradation, or decay that leads to the gradual loss of information over time (Fougnie, et al., 2012). This decrease in quality appears to reflect a process that operates independently across items (Fougnie, et al, 2012). Such degradation leads to substantial variability in the quality of memories across objects, with some objects remembered very well, others remembered poorly, and others may be completely forgotten (but see van den Burg et al., 2012). Consistent with this, the present results not only provide further evidence for the presence of variability in memory quality (Fougnie et al., 2012; van den Burg et al., 2012), but show that this variability cannot be explained by stimulus differences or by differential allocation of attention within or across trials.

Conclusion

Most research on metacognition has focused on perception and long-term memory, exploring how people assess uncertainty about their current perceptions and distant memories. Theories of metamemory have thus focused on how multiple sources of information influence judgments of confidence, including several static factors such as how memorable the material is, or judgments about our own abilities. Because of this, it has been difficult to assess whether and how participants have access to information that directly indexes the quality of a memory. In the present study, we developed two methods to strip away these static factors, enabling us to isolate realtime metamemory mechanisms, taking advantage of the fact that working memories appear to degrade stochastically over time. We found that observers appear to have access to the current state of their memories, and can use that information to guide their behavior in an ongoing task. These findings open the door to new explorations into the nature of the cues that enable realtime memory monitoring and into the impact of metamemory in complex cognitive processes that rely on working memory.

Acknowledgments

The authors thank Tim Brady and Patrick Cavanagh for helpful comments and to Sarah Cormiea for help with data collection. This work was supported by NIH Grant 1F32EY020706 to D.F. and NSF CAREER BCS-0953730 to G.A.A.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106–111.

Amichetti, N. M., Stanley, R. S., White, A. G., & Wingfield, A. (2013). Monitoring the capacity of working memory: Executive control and effects of listening effort. *Memory & Cognition*, 41(6), 839–849.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*(7), 622–628.

Barthelmé, S., & Mamassian, P. (2010). Flexible mechanisms underlie the evaluation of visual confidence. *Proceedings of the National Academy of Sciences, 107*(48), 20834-20839. doi: 10.1073/pnas.1007704107

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10), 1–27.

Bjork, R. A., Laberge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science*, 10, 55–56.

Bona, S., Cattaneo, Z., Vecchi, T., Soto, D., & Silvanto, J. (2013). Metacognition of Visual Short-Term Memory: Dissociation between Objective and Subjective Components of VSTM. *Frontiers in Psychology*, *4*, 62.

Bona, S. & Silvanto, J. (2014) Accuracy and Confidence of Visual Short-Term Memory Do Not Go Hand-In-Hand: Behavioral and Neural Dissociations. *PLoS ONE*, 9(3): e90808.

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual Long-Term Memory Has the Same Limit on Fidelity as Visual Working Memory. *Psychological Science, 24*(6), 981–990.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological bulletin, 109*(2), 204–223.

Bunnell, J. K., Baken, D. M., Richards–Ward, L. A. (1999). The effect of age on metamemory for working memory. *New Zealand Journal of Psychology,* 28(1), 23–29.

Burgess, A., & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *JOSA A, 5*(4), 617–627.

Flavell, J. H., & Wellman, H. M. (1976). Metamemory.

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nat Commun, 3*, 1229.

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience, 30*(1), 535–574. doi:10.1146/annurev.neuro.29.051605.113038

Gold, J. M., Murray, R. F., Sekuler, A. B., Bennett, P. J., & Sekuler, R. (2005). Visual memory decay is deterministic. *Psychological Science, 16*, 769–774.

Green, D. M. (1964). Consistency of auditory detection judgments. *Psychological Review, 71*(5), 392–407.

Kahneman, D. (1973). Attention and effort.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 455*(7210), 227–231.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370.

Levin, D. T., Momen, N., Drivdahl, S. B., & Simons, D. J. (2000). Change blindness blindness: The metacognitive error of overestimating change-detection ability. *Visual Cognition*, *7*, 397–412.

Metcalfe, J. E., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*: The MIT Press.

Muther, W. S. (1965). Erasure or partitioning in short-term memory. *Psychonomic Science*, 3, 429–430.

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology review, 15*(3), 105–130.

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

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review, 1*(3), 357–375.

Schwartz, B. L., Aaron, S. B., & Bjork, R. A. (1997). The Inferential and Experiential Bases of Metamemory. *Current Directions in Psychological Science, 6*(5), 132–137. doi: 10.2307/20182470

Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: a study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 452–460.

Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R., & Erb, L. (1995). The uncertain response in the bottlenosed dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology: General, 124*(4), 391–408.

Smith, J. D., Shields, W. E., Schull, J., & Washburn, D. A. (1997). The uncertain response in humans and animals. *Cognition, 62*(1), 75–97.

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision, 13*(10).

Tanaka, A., & Funahashi, S. (2012). Macaque monkeys exhibit behavioral signs of metamemory in an oculomotor working memory task. *Behav Brain Res, 233*(2), 256–270.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, 109*(22), 8780–8785.

Vandenbroucke, A. R. E., Sligte, I. G., Barrett, A. B., Seth, A. K., Fahrenfort, J. J., & Lamme, V. A. F. (2014). Accurate metacognition for visual sensory memory representations. *Psychological Science*, 24(4), 1–13.

Williams, M., Hong, S. W., Kang, M. S., Carlisle, N. B., & Woodman, G. F. (2013). The benefit of forgetting. *Psychonomic Bulletin & Review*, *20*(2), 348-355.

Williams, M. & Woodman, G. F. (2012). Directed forgetting and directed remembering in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1206.

Wellman, H. M. (1977). Tip of the tongue and feeling of knowing experiences: A developmental study of memory monitoring. *Child development*, 13–21.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*(12), 1120–1135.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233–235.

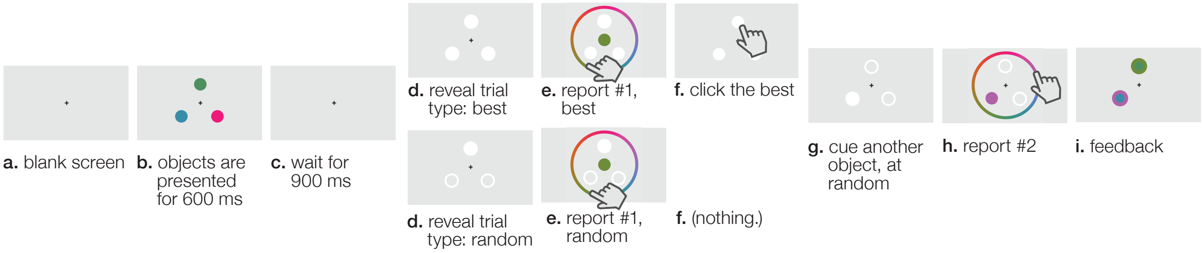

Figure 1. Timeline of one trial of the double-pass metamemory task. First, the participant sees a set of colorful dots and is asked to remember them for 900 ms (a–c). Then the type of trial is revealed: the participant will either choose, reporting the object they remember best (d, top, “best”), or mandatorily report the object highlighted for them (d, bottom, “random”). These two trial types are interleaved in random order, so that until (d) the participant does not know the trial type and thus must encode all the items. Once the trial type is revealed, the participant reports the color by selecting it from the color wheel (e). Then, a different item is selected at random (g) and the participant reports its color (h). This second report is later used in an assay of strategic or accidental tradeoffs (see Method and Online Supplemental Information). Finally, the participant receives feedback (i). These steps, which constitute one trial, are repeated hundreds of times in two rounds. In the second round, the displays used in the two conditions (“best” or “random”) are swapped, producing a double-pass procedure where, unbeknownst to the participant, in the second round the “randomly” chosen objects are in fact those chosen by the participant in the first round.

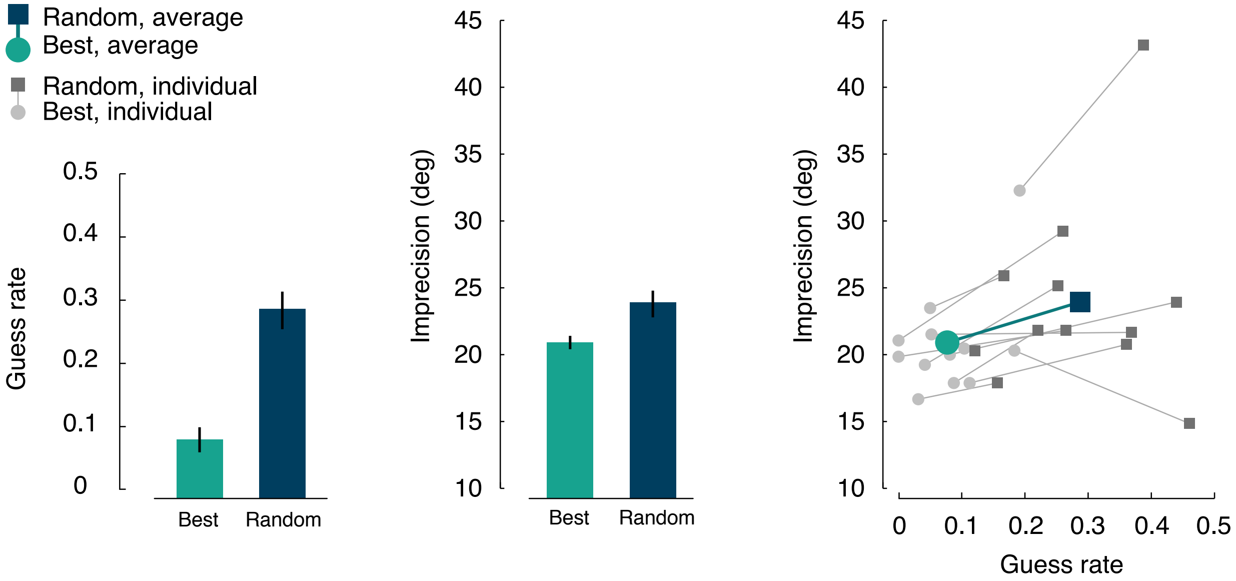

Fig. 2. The contribution of realtime monitoring to judgments of confidence in memory. We compare performance across two conditions, one where the participant selects an item as the one that was best remembered from that display (cyan circles), and another where the participant reports that same item because they were required to (blue squares). They perform better when they made the choice, which implies that participants can use realtime monitoring to guide their selections in the task, picking out the one they remember best.

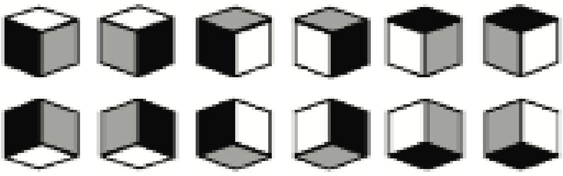

Fig. 3. Stimuli used in Exps. 2a–c. Stimuli were cubes with three visible sides in 12 possible arrangements.

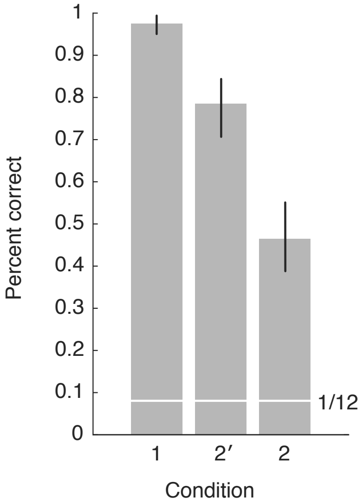

Fig. 4. Replication of directed remembering in visual working memory. In conditions 1 and 2, participants try to remember 1 or 2 objects, respectively. In condition 2′, the participant tries to remember 2 objects at first, but early in the retention interval is cued as to which object will later be tested. Participants use the cue to their advantage, redirecting maintenance to the to-be-tested object. Chance performance is 1/12. Error bars show the across-subject SEM.

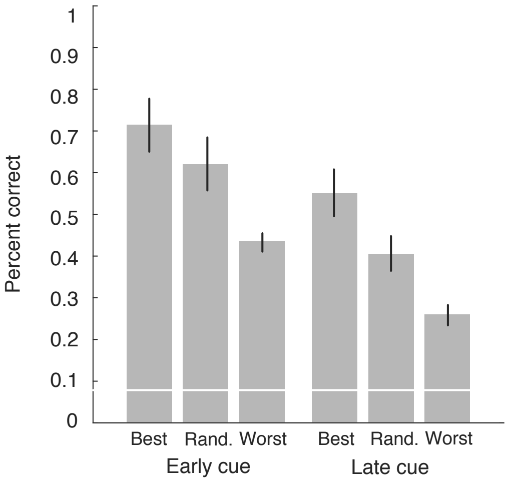

Fig. 5. Self-directed remembering in visual working memory. Participants try to remember two objects and then direct maintenance to the best-remembered object, the worst-remembered object, or a randomly selected object. The random condition is a replication of directed remembering from Exp. 2a. Participants benefit from directing maintenance in all conditions. Chance performance is 1/12. Error bars show the across-subject SEM.

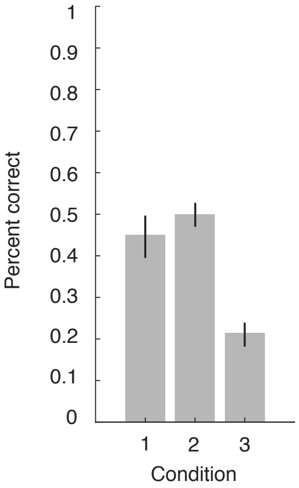

Fig. 6. Revisiting a metamemory decision after self-directed forgetting. Participants hold in mind two objects and then direct maintenance to the worst-remembered object. In condition 1, participants are tested on the object they to which they directed maintenance. In condition 3, participants are tested on the neglected object (i.e., the one that had originally been chosen as the best-remembered object). In condition 2, they are given the option to report either object, whichever is currently remembered best at the end of maintenance. On 96% of trials, they reverse their decision, opting to report the object that had initially been chosen as the worst remembered. Chance performance is 1/12. Error bars show the across-subject SEM.