Subjective experiences of recognizing and not recognizing paintings and words

Kaitlyn M. Fallow & D. Stephen Lindsay

2022

Subjective Experiences of Recognizing and Not Recognizing Paintings and Words

Kaitlyn M. Fallow & D. Stephen Lindsay

Department of Psychology, University Of Victoria

P.O. Box 1700 STN CSC

Victoria, British Columbia, Canada, V8W 2Y2

(250) 721-6165

kmfallow@uvic.ca

Author note: This research was supported by the Natural Sciences and Engineering Research Council of Canada through a Discovery Grant to D.S.L. and Canada Graduate Scholarships to K.M.F. We thank Jeffrey P. Toth for providing the painting stimuli used in the experiments described here. Painting images and trial-level data for these experiments are available at osf.io/ekduv.

Abstract

In our prior research, average recognition memory response bias tended to be conservative when stimuli were paintings, whereas bias for common English words tended to be liberal or neutral. Efforts to understand the mechanism(s) underlying this materials-based bias effect (MBBE) have yielded new questions but no definitive answers. Here we report a set of studies exploring the possibility that participants respond more conservatively to paintings because they expect the novel, visually rich paintings to evoke a strong, detailed memory experience at test, whereas the more familiar, visually similar words are not expected to produce this kind of vivid recollection as often. In three studies using variations of the remember/know procedure, we found that correctly recognized paintings were more often reported as "remembered" than were recognized words. There were also parallel materials-based differences in the reported bases for "new" responses. But we did not observe the expected relationships between response bias and these subjective reports. We discuss the implications of these results for accounts of the MBBE, and the more general issue of the role of stimulus materials in recognition memory response bias.

Keywords: Recognition memory, remember/know judgments, response bias

Public significance statement: Participants' recognition memory decisions—and the bases upon which they reported making these decisions—were markedly different for paintings and words. Understanding how memory and memory-related decisions are affected by the nature of the materials is important for the development of general theories of how memory works.

Subjective Experiences of Recognizing and Not Recognizing Paintings and Words

The remarkable capacity of human memory for pictures has long been of interest to psychology researchers (Kobayashi, 1986; Nickerson, 1965; Paivio & Csapo, 1973; Shepard, 1967). Much of the memory literature consists of studies using words and other verbal (or verbal-like, e.g., pseudowords) materials, and the similarities and differences in memory between these and other stimuli are important to understanding the limitations of our theories. With respect to words and pictures, much work has focused on the differences in memory accuracy for the two stimulus types: Pictures tend to be better remembered than words, a pattern often referred to as the "picture superiority effect" or PSE.

The mechanism(s) underlying PSEs remain debated (Ensor, Bancroft, et al., 2019; Ensor, Surprenant, et al., 2019). It is also unclear to what extent PSEs reflect quantitative differences in underlying processes versus more qualitative differences in which kinds of processing are engaged for different materials. Further, it is possible for differences in memory performance to arise from factors other than memory *per se,* such as decision-making. The contribution of stimulus/materials type to these aspects of task performance is even less well understood. For the purpose of developing theories that can account for how memory-related decisions are made in the real world in all its multimodal complexity, studying memory with verbal materials alone is unlikely to be sufficient. Understanding how, why, and under what conditions memory, metamemory, and related decision-making processes differ across materials such as words and pictures is an essential piece of the puzzle.

Here we will discuss a set of materials-based differences in recognition memory and some of our efforts to understand the underlying mechanism(s). This line of inquiry originated with an unexpected finding by Lindsay and Kantner (2011) that recognition memory response bias was consistently conservative when stimuli were images of paintings. By contrast, bias for words tested under the same conditions tended to be neutral (when words were the only stimulus type) or liberal (when words were intermixed with paintings at study and test, i.e., when materials were manipulated within-subjects). We refer to this pattern of findings as the materials-based bias effect (MBBE).

Several of the themes woven throughout Bill Hockley’s work are relevant to and have influenced our thinking about the MBBE project. Much of his work has taken up Newell's (1973) still-relevant caution against playing twenty questions with nature, emphasizing the limitations of a narrow focus on understanding laboratory effects at the expense of identifying core principles of memory and cognition (see also van Rooij & Baggio, 2021). We have no doubt sometimes fallen into this trap in our efforts with the MBBE. But Hockley and colleagues have also successfully leveraged the exploration of various “effects” toward the more central goals of better understanding memory and urging researchers to develop questions more likely to help us get there (e.g., Aßfalg et al., 2017; Jamieson et al., 2016; Neath et al., 2021; Vokey & Hockley, 2012). For example, he and coauthors have reported evidence that stimulus-based mirror effects (Neath et al., 2021) and PSEs (Ensor, Bancroft, et al., 2019; Hockley & Bancroft, 2011) are better thought of as patterns that arise from multiple interacting phenomena than as distinct “effects” in need of a single explanation. We see this as an admirable model, and think the natural scientific instinct to ask “what’s going on here?” when an unexpected effect arises—which motivated much of the work presented in this paper—can serve as a jumping-off point for thinking more deeply about how memory works and understanding the limitations of existing theories. We try to keep these principles in mind in our discussion of the MBBE.

The general procedure in all of the MBBE experiments was a standard old/new, study/test recognition memory paradigm. Participants studied 50-100 randomly ordered items one at a time and then completed a recognition test comprising all studied items and an equal number of new ones, the items being paintings, words, or equal numbers of both intermixed. The paintings and words in these experiments differed on a number of dimensions with potential relevance to memory and associated decision-making processes. Word stimuli were 4- to 8-letter English nouns of moderate-to-high frequency (e.g., spring, airport, deer) that the English-speaking undergraduate participants in our studies had encountered many times before. The paintings, by contrast, were selected for their novelty to the average undergraduate psychology student. The set comprises portraits and landscapes (and a few more abstract pieces) with a variety of styles and themes by various artists, some household names (e.g., van Gogh, Rembrandt) and others less well known (e.g., Gustave Caillebot, Winslow Homer). Extremely famous paintings (e.g., van Gogh's *Starry Night*) were deliberately excluded.

In addition to being novel to participants, the paintings were colourful, visually rich, and thematically complex, with many depicting scenes that cannot be easily labeled. The words, on the other hand, were presented in the same black font in the centre of a white background, so physical variation among these stimuli was limited. Conceptual variability among the words was fairly high, but they were all were nouns, most of which were concrete (although some could be used in multiple senses).

The recognition memory response bias difference between paintings and words might be attributed to differences in pre-experimental familiarity, complexity, and/or distinctiveness (Hunt, 2006). We have conducted many studies in an effort to better understand the MBBE (Fallow & Lindsay, 2021; Lindsay et al., 2015), most focusing on one of these dimensions. For example, we have explored the possibility that participants expect to remember the novel, vivid paintings in these studies better than the words and therefore hold the former to a higher standard, requiring more mnemonic evidence before they are willing to endorse them as "old." This mechanism—akin to the distinctiveness heuristic (Schacter et al., 1999)—could produce conservative responding regardless of whether memory is in fact better for paintings than words.[1]

Support for this "subjective memorability" account of the MBBE has been mixed. After studying a series of words and paintings (just before starting the subsequent recognition test), participants (*N* = 221, across 4 studies) did, on average, report that they expected to recognize a higher percentage of the paintings than of the words (Lindsay et al., 2015). The usual MBBE was observed on the recognition test, with participants showing a more conservative response bias on average for paintings than words. But across four experiments, there was no significant correlation between these two tendencies. Larger differences in the extent to which participants expected to remember paintings better than words did not consistently correspond with more conservative responding to paintings than words, as one would expect were subjective memorability driving the MBBE.

Despite this result, we did not entirely abandon the general idea underlying the subjective memorability hypothesis: that differing expectations about the memory experiences associated with paintings and words might contribute to response bias differences. The "differing expectations" aspect of this idea was consistently supported in the four experiments described above, and there are various reasons the design may not have been ideal for addressing the question of whether such differences underlie the MBBE. For example, there is evidence that

[1]Recognition memory sensitivity (d′) generally tends to be better for paintings than words in our experiments, a picture superiority effect. But we still see the MBBE when this pattern is eliminated or reversed (Fallow & Lindsay, 2021; Lindsay et al., 2015).

subjective memorability estimates obtained prior to a memory test can differ substantially from those obtained during the test (e.g., Guttentag & Carroll, 1998). It is possible subjective memorability as experienced throughout the recognition test is more relevant to understanding response bias differences than memorability estimates made retrospectively following the study phase. It is also possible that expectations regarding the *quality* of memory for different materials types contribute to response bias differences in ways that are not necessarily captured by asking about the quantity of different materials participants expect to remember.

We speculated that participants may have differing expectations and/or experiences regarding the quality of memories of paintings versus words, and this leads them to adopt different criteria for the two materials. The paintings in these studies are colourful, evocative, visually complex images that participants have never seen before, whereas the words are familiar, neutral English nouns presented in plain black font. Perhaps participants expect the paintings will more often elicit rich, vivid memory experiences than the words. They may notice such experiences throughout the study phase (e.g., when a painting reminds them of a previously studied painting) and/or early in the test phase, and adopt a strategy of responding “old” to paintings only when this kind of striking recollection is evoked. Such a tendency could produce a more conservative bias for paintings, regardless of whether they are in fact better remembered than words.

Here we report three experiments using versions of the remember/know procedure (Tulving, 1989) to explore possible materials-based differences in subjective memory experiences in the MBBE. In each study, subjects studied scans of paintings randomly intermixed with words and later completed a recognition memory test on which they reported on their phenomenological experience. We hypothesized that participants would, on average, report remembering more frequently for paintings than for words. We also predicted that there would be a correlation between (a) the extent to which participants reported recollective experiences more often for paintings than words and (b) the corresponding difference in response bias.

Method

Participants

This project was approved by the University of Victoria's Human Research Ethics Board (protocol 08-06-391b). Participants in all three experiments were undergraduate students at the University of Victoria who completed the study for optional bonus credit in psychology courses in 2013 or 2014. There were 93 participants in Experiment 1a, 118 in Experiment 1b, and 47 in Experiment 2. Sample sizes were determined by practical constraints on data collection (i.e., they were not planned based on power analyses or other currently recommended practices). Demographic data were not collected, but participants were drawn from a pool comprised primarily of 18- to 25-year-olds who self-identified as female and Caucasian (~70-80% of the pool falls into each of these categories), and there were no demographic criteria for participation.

Materials

The experiments were run on desktop PCs using E-Prime (Schneider et al., 2002a, 2002b). Word stimuli were 195 four- to eight-letter English nouns of moderate to high frequency obtained from the MRC psycholinguistic database (bit.ly/mrc1981; Coltheart, 1981). There were 204 images of paintings depicting various subjects and themes (see osf.io/ekduv). All paintings were from a larger collection assembled by Jeffrey P. Toth. Bitmap images ranged in size from 130-500 pixels in width and 270-500 pixels in height and the E-Prime program was set to display images at up to 75% of the screen size (with exact display size depending on image dimensions).

The study phase in all three experiments comprised 96 items (48 paintings & 48 words, randomly intermixed) bookended by three primacy and three recency buffers. The test list included all studied items and an additional 96 new/non-studied items (48 paintings & 48 words). Both study and test lists were randomly generated for each participant, such that a given item's status (studied or new) and position in the study and/or test list varied across participants[2].

Procedure

Participants completed the experiment individually with an experimenter present. Study instructions were the same for all experiments and conditions. Participants were told they were about to view a series of paintings and words, and were instructed to study each item and try to remember it for a later memory test. They were told that words and paintings would be equally important on the test and asked to pay equal attention to the two types of materials. Study phase stimuli were presented one at a time in the centre of a white background, and appeared for 1 s each with a 900-ms inter-stimulus interval (ISI; this included a 500-ms black central fixation cross). Following the study phase, there was a 5-min unrelated distractor task in which participants were asked to write down as many countries as they could.

The test was similar across all experiments and conditions. Participants were told they would see a series of words and paintings, some studied in the previous phase and others not, and that they would be asked whether each item was old/studied or new/not studied. Old/new decisions were made by pressing "1" (studied) or "0" (not studied) on the keyboard, and participants responded at their own pace to each item. Where applicable, remember/know decisions were made using the number keys (1-4 depending on the condition; see below). At the

[2] As one reviewer noted, this approach to randomization means that individual items did not necessarily end up assigned to the "studied" and "new" condition equally often. At the experiment level, the majority (77%) of items were assigned to the "studied" condition for 40-60% of participants (range: 20%-80%).

end of the test phase, participants were asked a few additional questions about their experience (e.g., what percentage of test items they thought were studied/not studied, how well they thought they did). They were then fully debriefed regarding the purpose of the experiment.

Remember/Know Procedure

Experiments and conditions differed in whether participants were asked for an additional response following the old/new decision, and in the options for these additional responses. In Experiments 1a and 1b, participants were pseudo-randomly assigned to one of three conditions: (1) a standard remember/know (RK [standard]) condition, in which participants were asked for subjective judgments following "studied" responses on the recognition test (*N*s = 31 & 40 in Exps. 1a & 1b, respectively); (2) an extended RK condition (RK [both]) in which participants were asked for subjective judgments following all responses (*N*s = 32 & 39); and (3) a control condition, in which participants completed a standard old/new recognition test without making any additional judgments (*N*s = 30 & 39).

Participants in the standard and RK (both) conditions viewed detailed instructions regarding the response options at the beginning of the test phase (see supplemental material), but brief descriptions of the options appeared for reference whenever the relevant response was being made at test. In both of these conditions, the response options following a "studied" decision were (1) "you have an experience of REMEMBERING. You remember something about the item or about studying it," (2) "you have an experience of KNOWING. You just know this was an item you studied," and (3) "you have an experience of GUESSING. You guessed your answer." In the extended RK condition, the response options following a "not studied" decision were (1) "you have an experience of REMEMBERING. You remember something that made you reject this item", (2) "you have an experience of NO FAMILIARITY. You feel that you just have NO FAMILIARITY with this item", and (3) "you have an experience of GUESSING. You guessed your answer." Participants made these responses using the 1-3 number keys and the item remained onscreen throughout.

All participants in Experiment 2 (*N* = 47) completed a modified version of the extended RK condition from Experiments 1a and 1b. A fourth response option followed "not studied" responses, specifically "You have an experience of thinking you WOULD HAVE REMEMBERED. You don't remember this and feel you WOULD had it been presented." These four response options for "new"/"not studied" responses are based on previous literature exploring the subjective bases of nonrecognition (e.g., Ghetti, 2003).

Analysis Details

Prior to analyses, some participants' data were excluded due to experimenter or computer errors, failure to understand the instructions, or memory-related disorders (5 in Exp. 1a, 3 in Exp. 1b, 1 in Exp 2). For the remaining participants (trial-level data available at osf.io/ekduv) we calculated hit rates (HRs) and false alarm rates (FARs) from old/new responses, separately for each materials type. False alarm rates of 0 (3 each in Exps. 1a & 1b, 1 in Exp. 2; all for paintings except 1 participant in Exp. 1a) were replaced as per Macmillan and Kaplan (1985; $0.5/n_{\text{new}}$) to enable estimation of equal variance signal detection measures of sensitivity (d′; $z_{\text{HR}} - z_{\text{FAR}}$) and response bias (c; $-0.5*[z_{\text{HR}} + z_{\text{FAR}}]$). We had not set any *a priori* exclusion criteria, but to keep the current analyses consistent with our more recent work we have excluded any subjects with d′ values below 0.2 for one or both materials types. This led to the exclusion of 17 further subjects (5 from Exp. 1a, 9 from Exp. 1b, 3 from Exp. 2).

For the conditions including a remember/know component, we calculated the proportions of responses of each type (i.e., hits, FAs, etc.) that each participant assigned to each remember/

know category, separately for paintings and words. The proportions of "remember" responses were used in paired two-sided t-tests with alpha = .05 to explore materials-based differences of interest. To evaluate the relationship between differences in recollection and the MBBE, we calculated difference scores for recollection ($\text{recollection}_{\text{paintings}} - \text{recollection}_{\text{words}}$) and response bias ($c_{\text{paintings}} - c_{\text{words}}$) and tested the correlation between them. We estimated two versions of the recollection score: (1) a corrected index that accounts for the proportion of incorrect "remember" responses (i.e., $pR_{\text{hits}} - pR_{\text{FAs}}$; e.g., Yonelinas, Aly, Wang, & Koen, 2010), and (2) the raw proportion of correct "old" responses (i.e., hits) reported as based on remembering (pR_{hits})[3]. Corrected recollection estimates could not be calculated for participants with FA rates of 0 for either materials type, so these participants were excluded from this version of the analysis.

Results

Signal Detection Measures

The usual pattern of conservative responding to paintings, and more conservative responding to paintings than words, was observed in all experiments and conditions (Figure 1). This included the control condition, which was included primarily to ensure this pattern—previously observed in several studies using a 6-point confidence-weighted scale—held up with a binary studied/not studied response scale. The remainder of this paper will discuss only the RK conditions. Means (with SDs) for response bias (c) as well as hit rates, false alarm rates, and sensitivity (d′) are available in the supplemental material (Table S1).

[3] Thanks to an anonymous reviewer for pointing us to the sum-difference model of the remember/know paradigm and associated SDT-based analytic techniques described by Rotello et al. (2004), which we hope to explore in future work.

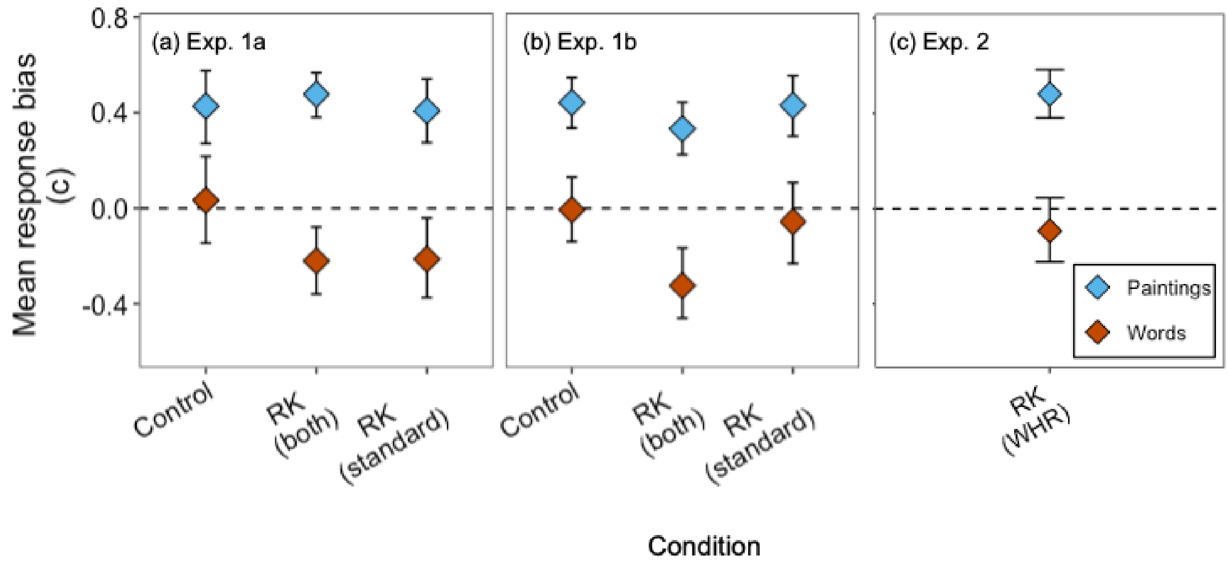

*Figure 1***.** Mean recognition memory response bias (c) by materials type (paintings in blue [lighter points], words in red-orange) and condition (control: binary old/new judgments only; RK [standard]: remember/ know judgments made following "old" responses only; RK [both]: remember/know judgments made following all responses; RK [WHR]: same as previous, with additional "would have remembered" option for "new" responses) in three remember/know experiments. Error bars are 95% bootstrap confidence intervals for the individual means.

Remember/Know Responses

Remember/know response proportions are shown in Figures 3 and 4. Of central interest were materials-based differences in the prevalence of remembering-related responses. Directionally, participants attributed a greater proportion of correct "old" responses (i.e., hits) to "remembering" for painting than word stimuli in all 3 experiments (Fig. 2a). This difference was significant in Experiments 1a, $t(54) = 3.97$, $p < .001$, mean difference = 0.11 [95 % CI: 0.05, 0.16], and 1b, $t(73) = 4.72$, $p < .001$, 0.14 [0.08, 0.19], but not in Experiment 2 ($t(43) = 1.95$, $p = .058$, 0.08 [-0.003, 0.16]). There was no consistent pattern of materials-based differences for false alarms (Fig. 2b), but these proportions should be interpreted with caution given the low numbers of false alarms on which they are based for paintings.

Participants on average attributed a greater proportion of correct rejections to having remembered something that led them to reject the item for paintings than words (Fig. 2c). This was significant in all experiments (Exp. 1a: $t(28) = 4.29$, $p < .001$, mean difference = 0.14 [95% CI: 0.07, 0.20]; Exp. 1b: $t(37) = 4.78$, $p <.001$, 0.14 [0.08, 0.20]; Exp. 2: $t(43) = 2.87$, $p = .006$, 0.07 [0.02, 0.12]). In Experiment 2, participants also chose the "would have remembered" option more often for paintings than words, $t(43) = 6.82$, $p < .001$, 0.19 [0.13, 0.24]. All of these materials-based differences were also observed for misses (Fig. 2d); that is, participants attributed a higher proportion of their incorrect "new" responses to remembering-related experiences for paintings than words (for the R response, Exp. 1a: $t(28) = 3.65$, $p = .001$, 0.12 [0.05, 0.19]; Exp 1b: $t(37) = 3.86$, $p <.001$, 0.10 [0.05, 0.15]; Exp. 2 : $t(43) = 2.56$, $p = .01$, 0.06 [0.01, 0.11]; for the WHR response, Exp. 2: $t(43) = 3.53$, $p = .001$, 0.12 [0.05, 0.19]).

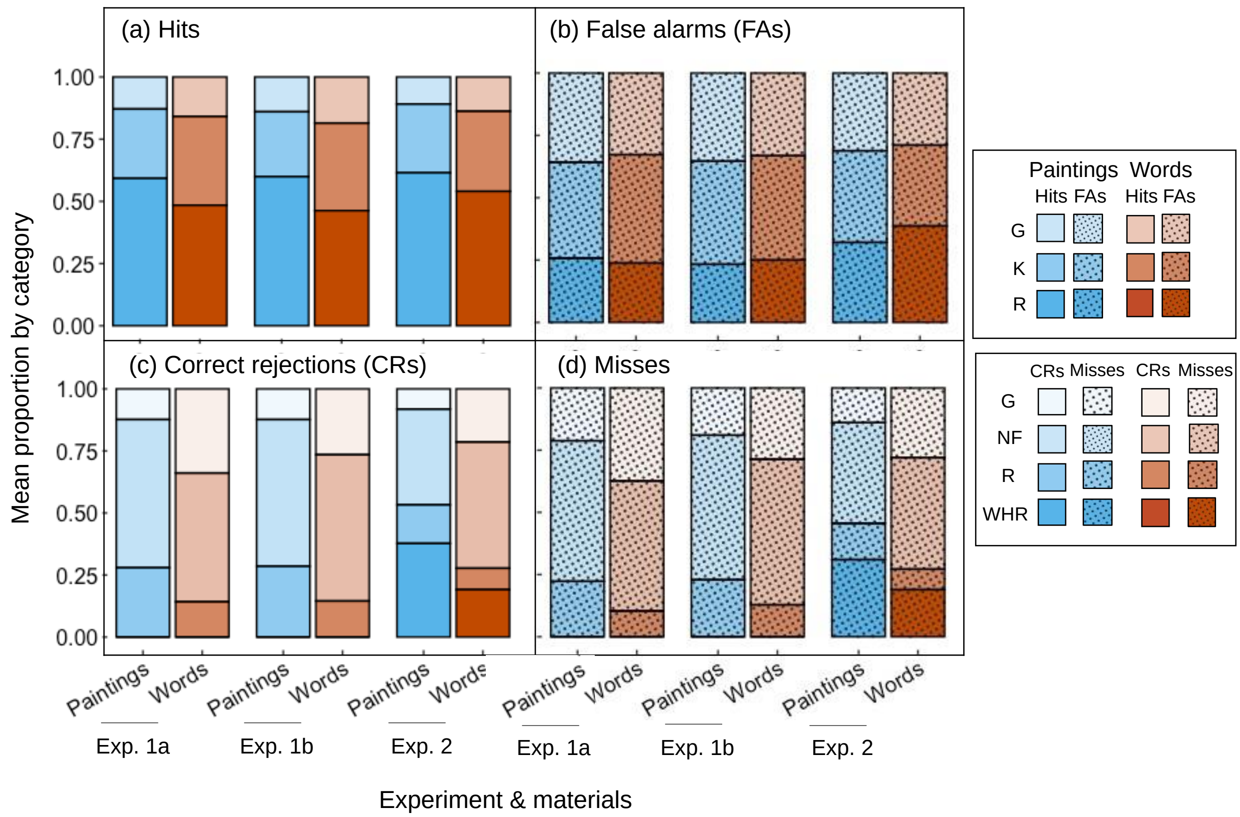

Figure 2. Mean proportions of hits, false alarms (FAs), correct rejections (CRs), and misses that were given each of the possible remember/know (RK) responses, by experiment and materials type. For “studied” responses (hits & FAs) the RK options were remember (R), know (K), and guess (G), and for “new” responses the options were remember (R), not familiar (NF), guess (G), and would-have-remembered (WHR; Exp. 2 only). Note that for Exps. 1a and 1b, which had 3 conditions, these values are based on participants from the two RK conditions (in the case of hits & FAs) or from the RK (both) condition only (in the case of CRs & misses).

Figure 3 shows a scatterplot of the materials-based differences in corrected recollection estimates versus materials-based differences in response bias. There was no significant correlation between $recollection_{paintings} – recollection_{words}$ and $c_{paintings} – c_{words}$ in Experiments 1a, *r*(52) = -.003[95% CI: -0.27, 0.26], *p* = .98, or 1b, *r*(69) =.16 [-0.08, 0.38], *p* = .19. This correlation was small but significant in Experiment 2, *r*(41) =.36 [0.07, 0.60], *p* = .02. That this sole significant correlation should be interpreted with caution is emphasized by the corresponding results using an uncorrected recollection measure (i.e., pR_{hits}); in this case, none of the correlations were significant (Exp. 1a: *r*(53) = -.03[95% CI: -.29, .24], *p* = .83; Exp 1b: *r*(72) = .0001 [-.23, .23], *p* = .99; Exp. 2: *r*(42) = .07 [-.22, .36], *p* = .64). Given the low statistical power within experiments here, we also explored these correlations after collapsing data across all 3 studies. These results showed a small-but-significantly positive overall correlation for corrected recollection, *r*(166) = .18 [.03, .32], *p* = .02, but not for the uncorrected measure, *r*(171) = .02[-.14, .16], *p* = .84.

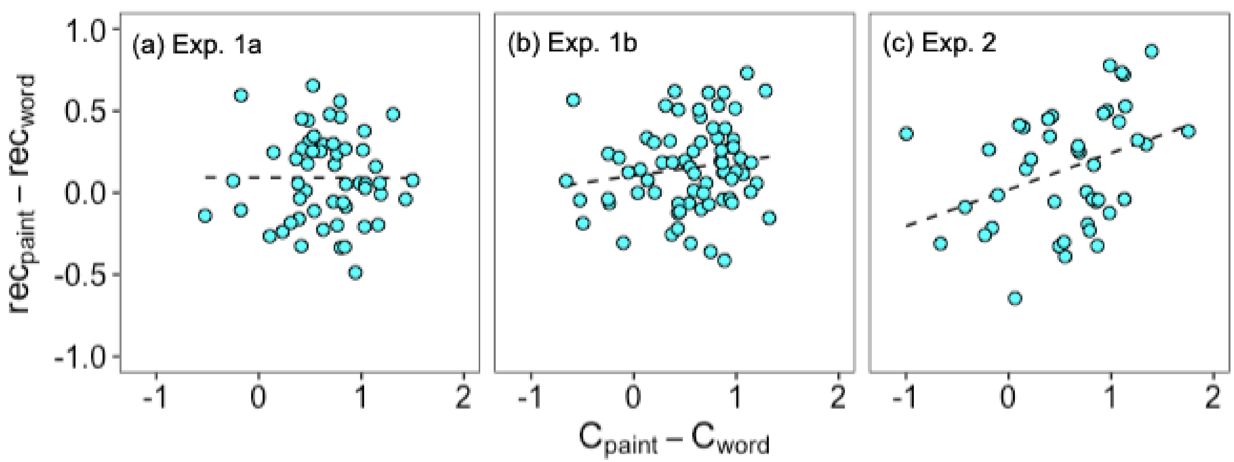

Figure 3. Scatterplot of the materials-based differences in recollection estimates ($rec_{paint} – rec_{word}$) vs. the materials-based bias effect ($c_{paint} – c_{word}$) in all three experiments, with regression line (dashed).

Discussion

We replicated the MBBE in all three experiments: Response bias was conservative for paintings, both relative to neutral bias and relative to bias for words (which was neutral or liberal). We consider that pattern, with these materials and subjects from this population, as well established (Fallow & Lindsay, 2021; Lindsay et al., 2015; Lindsay & Kantner, 2011). What is new here is that we observed materials-based differences in the subjective experience of making recognition memory decisions as captured via the remember/know procedure. When participants recognized a studied painting on the test, they attributed this to “remembering” more often than when they recognized a studied word (at least in Exps 1a & 1b; this difference was not significant in Exp. 2).

These findings are consistent with our intuition that subjects in the population tested expect to recollect studied paintings (see Bodner & Lindsay, 2003). The extent to which these rates of “remember” reports reflect a truly general materials-based expectation—that is, one that was present in previous MBBE experiments—versus a response pattern that developed over the course of the test as a result of the remember/know procedure remains to be seen. It may also be worth further exploring the possibility that the remember/know procedure itself may affect bias, as hinted at by the more liberal bias observed for words in some of the RK conditions relative to controls (Fig. 1). However, in conjunction with our previous finding that participants reported expecting to recognize more paintings than words after completing the study phase (Lindsay et al., 2015), it seems that there are materials-based differences in metamemory that may be relevant to understanding participants’ strategies and decision-making processes at test.

We undertook this work to explore the possibility that exaggerated expectations of “remembering” for paintings versus words might account for the MBBE. Our idea was that if subjects are more apt to require an experience of “remembering” to endorse a test probe as “old” if that probe is a painting rather than a word, that could lead to lower rates of “old” responses for paintings than for words (somewhat paradoxically, given the result is that the stimuli thought to be better-remembered end up producing lower observed hit rates). If so, then the proportion of “old” responses a subject classified as remembered would predict conservative bias, and the materials-based differences in these proportions would predict the existence/size of the MBBE. We examined the correlation between the MBBE and the differences in “old” responses categorized as Remembering at the subject level. Did subjects who showed the largest MBBE also show larger differences in the proportion of “old” responses characterized as based on Remembering? No. We did not find any consistent relationship between these measures (Fig. 3).

This non-significant relationship is not particularly compelling. The statistical power of these studies rests on the number of subjects, the reliability of the two measures, and the strength of the hypothesized relationship. Our sample sizes were modest and the other two parameters are not well established. Further, there may be individual differences in response strategy that are not adequately accounted for here (Miller & Kantner, 2020). Still, the results do not encourage the idea that the MBBE is largely driven by exaggerated expectations of recollection for paintings than words. To more definitively test this explanation of the MBBE may require experiments designed to manipulate “remembering:” For example, if we can eliminate or reverse the pattern of more remembering for paintings than words (e.g., with an encoding manipulation intended to boost remembering for words), can we also eliminate/reverse the MBBE?

Another novel finding here is the greater rate of Remembering as a basis for rejecting painting foils than word foils. When participants saw an item on the test that they did *not* recognize, they were more likely to attribute this to having remembered something that led them to reject the item (all experiments) or to a sense that they would have remembered the item if they had in fact seen it (Exp. 2) if the item was a painting than if it was a word. This was the case for both correct rejections and misses.

Our results suggest some interesting directions for future research. It seems the paintings more often brought to mind a memory of something that facilitated rejection, in line with a recall-to-reject mechanism (e.g., Rotello & Heit, 2000), regardless of whether the rejection was correct or not. Although such a mechanism may serve participants well in some contexts (e.g., recalling having studied *trees* when you see *tree* on a recognition test and you know items were only studied in one plurality form), it may also sometimes backfire if the memories brought to mind are less diagnostic of old/new status (e.g., recalling having studied a painting of trees when you see a painting of a single tree on the recognition test, and it is possible both were studied). This might produce uncertainty that could lead to more conservative responding.

Further exploring the pattern of remember/know judgments at the level of individual *items*, rather than just participants, may help us evaluate the above possibility. We have observed large and consistent item-level differences in response bias (i.e., across studies, response bias is consistently extremely conservative for some paintings, less so for others, with some even tending to demonstrate liberal bias (Fallow & Lindsay, 2017). Additionally, the remember/know procedure is somewhat artificial. It would be interesting to see if similar patterns emerge when participants are invited to provide more natural, unconstrained reports of the bases of their recognition decisions (e.g., Dobbins & Kantner, 2019; Selmeczy & Dobbins, 2014).

Jamieson, Mewhort, and Hockley (2016) decried a narrow focus on empirical phenomena and urged pursuit of “the goal of developing a coherent and general explanation of behavior” (see also Surprenant & Neath, 2009; van Rooij & Baggio, 2021). Bill Hockley and his coworkers have explored the interplay between recognition memory response bias (aka criterion) and subjective feelings of familiarity in the context of phenomena such as the revelation effect (Hockley & Niewiadomski, 2001), mirror effects (Hockley & Niewiadomski, 2007), context effects (Hockley et al., 2012), and directed forgetting (Montagliani & Hockley, 2019). Recently, Neath, Hockley, and Ensor (2021) presented evidence that some previously reported stimulus-based mirror effects (e.g., word frequency) may be due to confounding differences between the stimulus categories. The words and paintings in the studies reported here differed in myriad ways, making it harder to home in on a “coherent and general explanation.” But we have observed conservative recognition bias with a variety of other stimulus sets (e.g., photos of scenes, unfamiliar faces, audio clips of snippets of poetry), and neutral bias with other novel stimuli (e.g., Snodgrass & Vanderwart line drawings, nonwords). This encourages us to believe that this line of research has the potential to help elucidate a general principle of memory.

References

Aßfalg, A., Bernstein, D. M., & Hockley, W. (2017). The revelation effect: A meta-analytic test of hypotheses. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1227-6

Bodner, G., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language*, *48*(3), 563–580. https://doi.org/10.1016/S0749-596X(02)00502-8

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505. https://doi.org/10.1080/14640748108400805

Dobbins, I. G., & Kantner, J. (2019). The language of accurate recognition memory. *Cognition*, *192*(May), 103988. https://doi.org/10.1016/j.cognition.2019.05.025

Ensor, T. M., Bancroft, T. D., & Hockley, W. E. (2019). Listening to the picture-superiority effect: Evidence for the conceptual-distinctiveness account of picture superiority in recognition. *Experimental Psychology*, *66*(2), 134–153. https://doi.org/10.1027/1618-3169/a000437

Ensor, T. M., Surprenant, A. M., & Neath, I. (2019). Increasing word distinctiveness eliminates the picture superiority effect in recognition: Evidence for the physical-distinctiveness account. *Memory and Cognition*, *47*(1), 182–193. https://doi.org/10.3758/s13421-018-0858-9

Fallow, K. M., & Lindsay, D. S. (2017). *Do some images seem newer than others? Examining recognition memory response bias at the item level*. Annual meeting of the Psychonomic Society, Vancouver, BC, Canada.

Fallow, K. M., & Lindsay, D. S. (2021). Test position effects on hit and false alarm rates in recognition memory for paintings and words. *Memory & Cognition*. https://doi.org/10.3758/s13421-021-01227-5

Fallow, K. M., & Lindsay, D. S. (2022). *Subjective experiences of recognizing and not recognizing paintings and words*. Open Science Framework. https://osf.io/ekduv/

Ghetti, S. (2003). Memory for nonoccurrences: The role of metacognition. *Journal of Memory and Language*, *48*(4), 722–739. https://doi.org/10.1016/S0749-596X(03)00005-6

Guttentag, R., & Carroll, D. (1998). Memorability judgments for high- and low-frequency words. *Memory and Cognition*, *26*(5), 951–958. https://doi.org/10.3758/BF03201175

Hockley, W. E., & Bancroft, T. (2011). Extensions of the picture superiority effect in associative recognition. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *65*(4), 236–244. https://doi.org/10.1037/a0023796

Hockley, W. E., Bancroft, T. D., & Bryant, E. (2012). Associative and familiarity-based effects of environmental context on memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(2), 81–89. https://doi.org/10.1037/a0027136

Hockley, W. E., & Niewiadomski, M. W. (2001). Interrupting recognition memory: Tests of a criterion-change account of the revelation effect. *Memory and Cognition*, *29*(8), 1176–1184. https://doi.org/10.3758/BF03206387

Hockley, W. E., & Niewiadomski, M. W. (2007). Strength-based mirror effects in item and associative recognition: Evidence for within-list criterion changes. *Memory and Cognition*, *35*(4), 679–688. https://doi.org/10.3758/BF03193306

Hunt, R. R. (2006). *Distinctiveness and Memory*. https://doi.org/10.1093/acprof:oso/9780195169669.001.0001

Jamieson, R. K., Mewhort, D. J. K., & Hockley, W. E. (2016). A computational account of the production effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(2), 154–164. https://doi.org/10.1037/cep0000081

Kobayashi, S. (1986). Theoretical issues concerning superiority of pictures over words and sentences in memory. *Perceptual and Motor Skills*, *63*, 783–792. https://doi.org/10.2466/pms.1986.63.2.783

Lindsay, D. S., & Kantner, J. (2011). A search for influences of feedback on recognition of music, poetry, and art. In P. A. Higham & J. P. Leboe (Eds.), *Constructions of remembering and metacognition: Essays in honor of Bruce Whittlesea* (pp. 137–154). Palgrave Macmillan UK. https://doi.org/10.1057/9780230305281_11

Lindsay, D. S., Kantner, J., & Fallow, K. M. (2015). Recognition memory response bias is conservative for paintings and we don't know why. In *Remembering: Attributions, processes, and control in human memory: Essays in honor of Larry Jacoby* (pp. 213–229).

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*(1), 185–199. https://doi.org/10.1037/0033-2909.98.1.185

Miller, M. B., & Kantner, J. (2020). Not all people are cut out for strategic criterion shifting. *Current Directions in Psychological Science, 29*(1), 9–15. https://doi.org/10.1177/0963721419872747

Montagliani, A., & Hockley, W. E. (2019). Item-based directed forgetting for categorized lists: Forgetting of words that were not presented. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *73*(3), 135–143. https://doi.org/10.1037/cep0000177

Neath, I., Hockley, W. E., & Ensor, T. M. (2021). Stimulus-based mirror effects revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000901

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In *Visual Information Processing* (pp. 283–308). https://doi.org/10.1021/bi047419c

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, *19*(2), 155–160.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*(2), 176–206. https://doi.org/10.1016/0010-0285(73)90032-7

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory and Cognition*, *28*(6), 907–922. https://doi.org/Doi 10.3758/Bf03209339

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-Difference Theory of Remembering and Knowing: A Two-Dimensional Signal-Detection Model. *Psychological Review*, *111*(3), 588–616. https://doi.org/10.1037/0033-295X.111.3.588

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distintiveness heuristic. *Journal of Memory and Language*, *40*, 1–24.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-Prime reference guide*. Psychology Software Tools, Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-Prime: User's guide*. Psychology Software Tools, Inc.

Selmeczy, D., & Dobbins, I. G. (2014). Relating the content and confidence of recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 66–85. https://doi.org/10.1037/a0034059

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 156–163. https://doi.org/10.1016/S0022-5371(67)80067-7

Surprenant, A. M., & Neath, I. (2009). *Principles of Memory*. Psychology Press.

Tulving, E. (1989). Remembering and knowing the past. *American Scientist*, *77*(4), 361–367. http://www.jstor.org/stable/27855835

van Rooij, I., & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*, *16*(4), 682–697.

Vokey, J. R., & Hockley, W. E. (2012). Unmasking a shady mirror effect: Recognition of normal versus obscured faces. *Quarterly Journal of Experimental Psychology*, *65*(4), 739–759. https://doi.org/10.1080/17470218.2011.628399

Yonelinas, A. P., Aly, M., Wang, W.-C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*(11), 1178–1194. https://doi.org/10.1002/hipo.20864