

Preprint: This version may differ in non-substantive ways from the published version.

Response Bias in Recognition Memory as a Cognitive Trait

Justin Kantner and D. Stephen Lindsay

University of Victoria, Victoria, BC, Canada

Abstract

According to signal detection theory, old/new recognition decisions can be affected by response bias, a general proclivity to respond “old” or “new.” In recognition experiments, response bias is usually analyzed at a group level, but substantial individual differences in bias can underlie group means. These differences suggest that, independent of any experimental manipulation, some people require more memory evidence than others before they are willing to call an item “old.” In four experiments, we investigated the possibility that recognition response bias is a partial function of a trait-like predisposition. Bias was highly correlated across two recognition study/test cycles using words as stimuli that were separated by 10 minutes (Experiment 1). A nearly identical correlation was observed when the tasks were separated by one week (Experiment 2). Correlations remained significant even when the stimuli differed sharply between the first and second study/test cycle (Experiment 3). No relationship was detected between bias and response strategies in two general knowledge tests (Experiments 2 and 4), but bias predicted frequency of false recall in the Deese-Roediger-McDermott (DRM) paradigm (Experiment 4). This evidence of trait-like stability suggests an entirely different aspect of response bias than that studied by examining its reaction to task variables, one for which complete theories of recognition memory may need to account.

Keywords: memory, recognition memory, response bias, individual differences, trait

Response Bias in Recognition Memory as a Cognitive Trait

Suppose that two participants studied an identical list of commonplace items and each took an identical yes/no recognition test containing an equal number of items that were and were not on the study list. Suppose further that one participant achieved a hit rate (H) of .90 and a false alarm rate (FA) of .40 while the other achieved a H of .60 and a FA of .10. Although these two participants were equally accurate (75% correct), their performances were strikingly different. This difference would be characterized as one of response bias: The first participant tended to favor “old” responses and was liberally biased, while the second favored “new” responses and was conservatively biased.

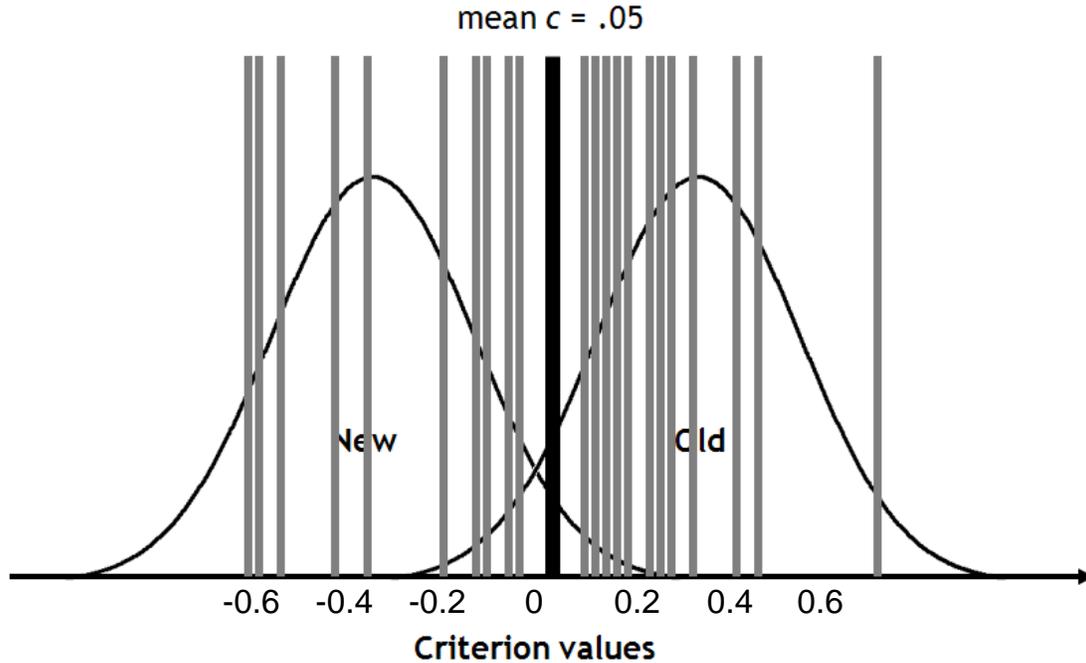
Recent years have seen an accelerating interest in response bias as a measure revealing important strategic influences on recognition memory. It is well established that participants adjust bias according to instructional motivation (e.g., Egan, 1958), payoff schedules that encourage old or new responses (e.g., Healy & Kubovy, 1978; Van Zandt, 2000), and the proportion of old test items (e.g., Van Zandt, 2000). Further work has revealed stimulus-specific properties that affect bias, such as the emotional content of test items (e.g., Dougal & Rotello, 2007) and their subjective memorability (e.g., Brown, Lewis, & Monk, 1977). Another class of studies has investigated the extent to which participants can adjust criterion during the course of a test when conditions such as the memory strength of probes or target-distractor similarity are changed midstream (e.g., Benjamin & Bawa, 2004; Cox & Shiffrin, 2012; Singer, 2009; for a review see Hockley, 2011).

The present experiments were designed to examine recognition response bias from a different perspective. The objective was to ask whether bias is strictly a function

of prevailing experimental conditions or inheres to a degree in individual recognizers as a cognitive trait. Often, average response bias on recognition memory tests is neutral, unless manipulations are employed to push bias in a liberal or conservative direction. However, substantial individual differences in bias often underlie group means. Figure 1 illustrates an example of this phenomenon from an experiment reported by Kantner and Lindsay (2010), which involved a standard item recognition task. In the control condition of that experiment, the mean response bias (bold line) was statistically neutral, but estimates of criteria from the 23 individual participants (gray lines) showed considerable variability. Why would participants in an experiment containing no biasing instructions or incentives differ to such a degree on this measure? While this variability might simply have resulted from measurement error, an intriguing possibility is that such variability reflects bias proclivities within individuals that are independent of the parameters of the recognition task. In the current experiments, we tested the hypothesis that the level of memory evidence participants require before committing to an "old" decision is a cognitive trait, and the bias they display on a recognition test a manifestation of that trait.

Figure 1

The spread of individual criterion values in Kantner and Lindsay (2010), Experiment 1, control condition. The curves representing new and old strength distributions are purely illustrative, not derived from the data.



In referring to recognition response bias as a potential trait, we mean to suggest a characterization of bias as an aspect of cognition that typifies an individual. We use the label *cognitive trait* to distinguish response bias from a personality trait, although we do not imply that cognitive and personality traits are independent constructs. Response bias might be thought of as a cognitive trait in the same sense as recognition sensitivity: all experimental conditions being equal, one would not expect sensitivity to the difference between old and new items to vary haphazardly within an individual from one test to another. Rather, sensitivity on the two tests would be expected to have a predictive relationship. We used this metric to assess the within-individual stability of response bias in the present experiments. The notion of bias as a trait also implies that it holds

behavioral consequences that extend beyond the domain of recognition memory; we conducted some initial tests of this possibility in Experiments 2 and 4.

Although response bias is not generally characterized as representing a trait in the recognition literature, studies have identified a number of populations exhibiting a more liberal response bias than appropriate comparison groups, such as the elderly (e.g., Huh, Kramer, Gazzaley, & Delis, 2006), patients with Alzheimer's disease (e.g., Beth, Budson, Waring, & Ally, 2009), schizophrenia patients (e.g., Moritz, Woodward, Jelinek, & Klinge, 2008), dementia patients (e.g., Woodard, Axelrod, Mordecai, & Shannon, 2004), individuals with mental retardation (Carlin, Toglia, Wakeford, Jakway, Sullivan, & Hasel, 2008), and panic disorder patients (Windmann & Kruger, 1998). The association of liberal response bias and particular populations is consistent with the idea that groups of individuals may be differentiated from one another on the basis of response bias without a specific experimental intervention, which is, in turn, consistent with the notion of response bias as a cognitive trait. Relatedly, a small literature on the "yea-saying bias" suggests that some individuals are predisposed to respond affirmatively to questions, a phenomenon demonstrated by young children and individuals with impaired cognitive development (e.g., Couch & Keniston, 1960).

At least two studies have examined the correlation of response bias and cognitive or personality traits within an individual; significant relationships between recognition response bias and established traits suggest that bias also possesses trait-like qualities. Because frontal brain regions are often implicated in criterion setting (e.g., Kramer et al., 2005), Huh et al. (2006) correlated response bias on a recognition test with performance on four measures of executive function in 293 35- to 89-year-old adults. Of these,

inhibition (measured via a Stroop task) was the only significant predictor of response bias ($r = .31$, estimated from the reported beta values); Huh et al. declared the analysis inconclusive. In a study of 28 undergraduates, Gillespie and Eysenck (1980) found that introverts used a more conservative recognition criterion than extraverts and were described as exercising greater “response cautiousness.” This finding is consistent with the possibility that response bias may arise from a trait corresponding to a required level of evidence before action is taken, a trait that, like introversion/extraversion, is stable within an individual.

A number of studies have investigated individual differences in false memory proneness via the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995). Given that a liberal recognition bias is associated with increased endorsement of test probes that were not studied, evidence that DRM false recognition has trait-like qualities could suggest the same for response bias. DRM performance has been correlated with a number of individual difference measures (e.g., age, working memory, frequency of dissociative experiences; for a review see Gallo, 2010), and some experiments have identified populations with particularly high rates of DRM errors (e.g., Geraerts, Lindsay, Merckelbach, Jelicic, Raymaekers, Arnold, & Schooler, 2009). These findings suggest that some individuals are inherently more prone than others to accept memories as true even when memory evidence is weak, making them especially vulnerable to false memories.

Two studies have assessed the within-individual stability of DRM false recognition. Salthouse and Siedlecki (2007) found reliable stability within a single test but not across separate tests differing in stimulus type, and false recognition of critical

lures was uncorrelated with a host of cognitive and personality measures in two experiments. However, Blair, Lenton, and Hastie (2002) found high levels of reliability in tests of the same DRM lists given two weeks apart, indicating that false recognition does not vary unpredictably within an individual.

Two further findings from the DRM literature are suggestive with respect to trait response bias. Blair et al. (2002) reported a significant correlation between critical and non-critical false alarms on the first test, a result that hints at a relationship between general recognition bias and DRM false memories (although that correlation was not significant in the second test). Relatedly, Qin, Ogle, and Goodman (2008) found that response bias calculated from the non-critical DRM trials was significantly (if weakly) predictive of susceptibility to adopting fictitious childhood events as autobiographical. These results are consistent with the possibility that response bias is a trait that generalizes to tasks outside of recognition memory.

Measurement of Response Bias

The measurement of response bias raises complex theoretical and statistical questions relevant to any recognition memory experiment, and the optimal method for estimating bias has been a matter of extensive debate. There are many options for calculating bias, and each is tied to model-based assumptions that may or may not hold for a given dataset (see Rotello & Macmillan, 2008). In testing the trait-like stability of bias, then, it is important to establish that any evidence of within-individual consistency is not tied to a particular index. Therefore, in addition to our primary bias measure (c), we calculated consistency using four other well-known estimates of bias (c_a , $\ln[\beta]$, B'' , and FA) in each of the current experiments. A brief description of these measures follows.

We describe and illustrate our results based on the bias estimate c (Macmillan, 1993), a simple and widely used measure given as the opposite of half the sum of the z -converted hit and false alarm rates. Positive values of c indicate a criterion to the right of the intersection of the old- and new-item distributions and a bias to respond “new;” negative values indicate a criterion to the left of the intersection of the distributions and a bias to respond “old.” One criticism of c is the assumption inherent in its calculation that the two distributions have equal variance; evidence from Receiver Operating Characteristic (ROC) curves suggests that the variance is often greater for the old than for the new distribution (e.g., Ratcliff, Sheu, & Gronlund, 1992). An alternative to c that addresses this shortcoming is c_a , which produces estimates of response criterion at multiple levels of confidence that take into account the relative variances of the old- and new-item distributions (Macmillan & Creelman, 2005). When the two variances truly are equal, c will be equivalent to c_a at the middle (neutral) confidence level; to the extent that the variances differ, c will deviate from middle c_a . We calculated c_a from confidence ratings collected at test and report the middle c_a values obtained in each of the current experiments.

The additional bias measures computed in the present experiments were chosen to represent a range of approaches to estimating bias. $\ln(\beta)$ is the natural log of the ratio of the target and lure likelihoods at a given (criterial) point on the strength of evidence axis. B'' is a prominent non-parametric bias statistic, although its status as a “model free” estimate has been questioned (Macmillan & Creelman, 1996). Calculations and discussion of $\ln(\beta)$ and B'' can be found in Stanislaw and Todorov (1999). Finally, the

use of the false alarm rate as a bias index simply estimates bias with no reference to the hit rate.

An additional consideration in the measurement of response bias is the means of estimating recognition sensitivity; changes in bias can be difficult to disentangle from changes in sensitivity, particularly when both are indexed by hit and false alarm rates. For example, if a participant completes two recognition tests and achieves the same false alarm rate on each but a higher hit rate on the second test, the resulting increase in d' and decrease in c across tests leaves unclear whether sensitivity has increased, bias has become more liberal, or both. To help minimize the statistical confounding of bias and recognition sensitivity, we used the sensitivity estimate A_z , the area under the ROC curve (Verde, Macmillan, & Rotello, 2006). Both A_z and c_a were computed using Lewis O. Harvey's RSCORE program (Harvey, 2005).

Experiment 1

If response bias represents a cognitive trait, it should remain consistent within an individual across time. Therefore, an important first step in establishing response bias as trait-like is to determine whether a given subject will show the same level of bias on two different recognition tests. Experiment 1 tested this possibility in a straightforward manner: Each subject took two recognition tests (each preceded by its own study list) that were separated by a filled 10-minute interval. The measure of interest was the correlation between bias on Test 1 and bias on Test 2.

Method

Participants. In each of the present experiments, University of Victoria students participated for optional bonus credit in an undergraduate psychology course. There were 41 participants in Experiment 1.

Materials. The stimuli were 192 4- to 8-letter medium- to high-frequency English nouns drawn from the MRC psycholinguistic database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm; Coltheart, 1981). Study and test lists were created via random selection from the 192-word pool for each participant. Forty-eight randomly selected words composed Study List 1. Test List 1 consisted of the 48 words from Study List 1 and 48 non-studied words. Study and Test List 2 were populated in the same way, with no words repeated from the first study/test cycle. Each study list included 3 primacy and 3 recency buffers. Study and test lists were presented in a randomized order. All of the current experiments were conducted with E-Prime software.

Procedure. Study items were presented for 1 s each with a blank 1-s interstimulus interval (ISI). Participants were instructed to remember each word for a subsequent memory test. Upon completion of the study list, participants received test instructions informing them that they would see another list of words, that some of these words had appeared in the preceding study list and some had not, and that their task was to indicate whether or not each item had been studied. Recognition judgments were made on a 6-point, confidence-graded scale (1 = Definitely Not Studied, 2 = Probably Not Studied, 3 = Maybe Not Studied, 4 = Maybe Studied, 5 = Probably Studied, 6 = Definitely Studied). Responses were non-speeded. A 1-s intertrial interval (ITI) separated test trials.

After the first study-test cycle, participants spent 8 minutes writing down the names of as many countries as they could. The procedure for the second study/test cycle was identical to that of the first, with the exception of a minor instructional modification to inform participants that no words would be repeated from the first cycle.

Results and Discussion

In this and each subsequent experiment, recognition rating data were converted to hits and false alarms by scoring responses of 4, 5, and 6 as hits on target trials and as false alarms on lure trials. Occasional false alarm rates of 0 and hit rates of 1 were replaced according to Macmillan and Kaplan (1985). Across experiments, such replacements were made for 0.65% of scores.

Table 1

Recognition means in Experiment 1.

	H		FA		c		A_z	
	M	SD	M	SD	M	SD	M	SD
Test 1	.70	.13	.26	.13	.07	.38	.79	.09
Test 2	.74	.17	.27	.17	-.02	.47	.80	.14

Note. H = hit rate, FA = false alarm rate, M = mean, SD = standard deviation.

The means of interest are displayed in Table 1. Bias was roughly neutral and did not differ significantly between Test 1 and Test 2, $t(40) = 1.59$, $p = .12$. Recognition sensitivity was statistically equivalent across tests, $t < 1$.

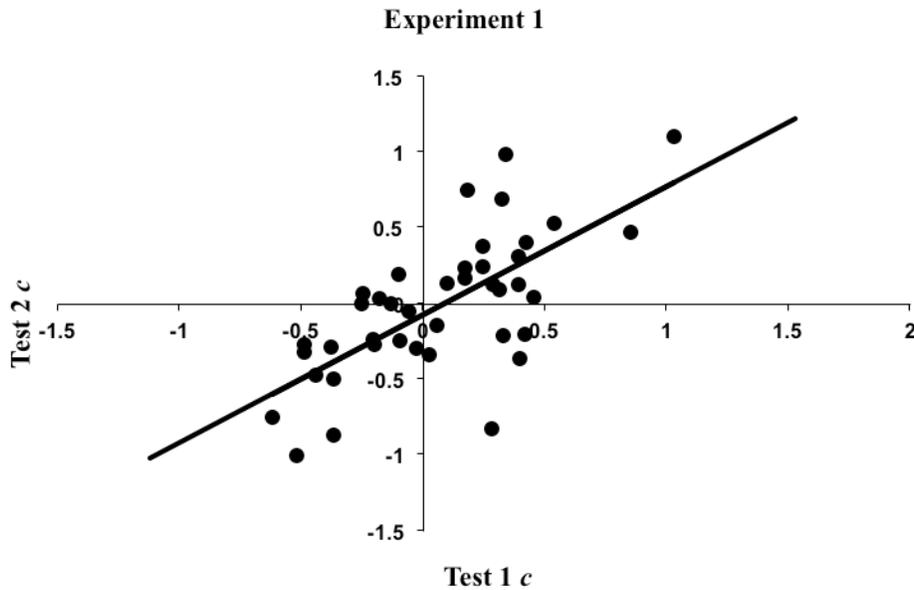
Bias varied greatly at the level of the individual, ranging from extremely conservative to extremely liberal. The highest value of c in a single test was 1.10 ($H = .44$, $FA = .02$); the lowest value was -1.01 ($H = .92$, $FA = .73$). The question was whether these values were predictive of bias across the two recognition tests.

Five bias statistics (c , c_w , $\ln[\beta]$, B'' , and FA) were used to calculate Test 1-Test 2 bias correlations in Experiments 1-3. To minimize the influence of outliers on observed correlations, any bias scores more than 3 SD from the mean for a given measure in a given experiment were removed prior to correlational analysis; across all measures and experiments, this cutoff resulted in the removal of 0.73% of bias scores, or 1.47% of data points in correlational analyses. Most of the bias scores removed were $\ln[\beta]$ (53%) or FA (29%) values. All correlations were calculated with Pearson's r statistic and controlled for sensitivity (A_z) on Test 1 and Test 2.

Test 1 c is plotted against Test 2 c for each participant in Figure 2. As is clear from inspection of the figure, the direction and magnitude of individuals' bias tended to be consistent from Test 1 to Test 2. Overall, there was a strong positive correlation between bias on the first and second tests, $r(37) = .67$, $p < .001$. Correlations based on the supplementary bias measures are displayed in Table 2. These analyses indicated positive, highly significant bias relationships across tests regardless of the measure used, suggesting that the within-individual stability of bias in Experiment 1 was not an artifact of the properties of any particular estimate.

Figure 2

Correlation of recognition bias at Test 1 and Test 2 in Experiment 1.



To establish a benchmark against which to compare inter-test bias correlations, the split-half reliability of c within a single test was measured. The within-test correlations were .69 and .78 on Tests 1 and 2, respectively, for a mean within-test correlation of .73. Thus, the level of stability in bias across tests in Experiment 1 was similar to that observed within a single test, an indication that a delay of 10 minutes and a separate study/test cycle had very little effect on participants' response bias.

Table 2.

Correlations of recognition memory response bias between Test 1 and Test 2 as a function of bias measure in Experiments 1-3.

	c	c_a	$\ln(\beta)$	B''	FA
Experiment 1	.67***	.58***	.55***	.63***	.60***
Experiment 2	.73***	.73***	.47**	.60***	.71***
Experiment 3: PP	.31	.30	.33	.32	.40*
Experiment 3: PW	.64***	.67***	.51**	.57**	.60***
Experiment 3: WP	.45**	.41*	.68***	.31	.75***
Experiment 3: WW	.81***	.80***	.70***	.76***	.72***

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

Experiment 2

The finding of bias consistency when 10 minutes separated two recognition study/test cycles shows that variability in bias across participants is not solely the result of measurement error. Experiment 2 was designed to provide a stronger test of lasting consistency in bias by separating two study/test cycles by one week.

The second goal of Experiment 2 was to investigate a second dimension of trait-like stability in response bias: its transfer to non-recognition memory tasks. The idea is that if response bias is the manifestation of an “evidence requirement” trait, it should correlate with performance on other tasks in which an evidence requirement might influence responses.

This possibility was tested with two such tasks in Experiment 2. The first was a DRM list recall task. Given the decreased caution exercised by liberal recognizers in accepting words as having been encountered previously, the prediction was that such participants would be more likely to commit false recall of critical DRM lures than participants exhibiting a conservative recognition bias. The second non-recognition measure was grain size in estimating answers to general knowledge questions (Goldsmith, Koriat, & Weinberg-Eliezer, 2002). Participants were asked questions (e.g., “What year did Elvis Presley die?”) and responded with numerical ranges that they believed were likely to contain the exact answer. Fine-grained answers (e.g., “1973-1978”) are less likely to be accurate but are more informative than coarse-grained answers (e.g., “1950-1990”). The grain size with which one answers a question is understood to reflect preference for accuracy versus informativeness in responding (Ackerman & Goldsmith, 2008). We predicted that participants exhibiting a more conservative recognition bias would tend to use wider ranges than liberal recognizers on the basis that recognition response bias is a reflection of a “required evidence” trait: Conservative recognizers were hypothesized to require more evidence of their knowledge of a topic before committing to a narrow range answer.

Method

Participants. There were 46 participants in Experiment 2.

Materials. The stimuli used in the recognition portions of the experiment were identical to those used in Experiment 1. The stimuli used in the DRM task were the doctor, window, rough, bread, anger, sweet, couch, and smell lists from Stadler,

Roediger, and McDermott (1999). Each list contained 15 words in decreasing order of semantic relatedness to the category prototype (Roediger & McDermott, 1995).

The general knowledge task included 50 trivia-style questions, each with an exact numerical answer, written by the first author and a research assistant. All questions selected for use in Experiment 2 called for answers in the form of specific years. Each question began with the words “In what year” and referred to a historical, political, scientific, or pop cultural event from the last 200 years.

Procedure. Participants took part in two sessions at the same time of day exactly one week apart. Session 1 consisted of a recognition study/test cycle and either the general knowledge or DRM task. Session 2 consisted of a second recognition study/test cycle (using different words from Session 1) and whichever of the general knowledge and DRM tasks was not included in Session 1. The assignment of the non-recognition tasks to Sessions 1 and 2 was random for each participant, as was the order of the two tasks within each session.

The procedure for the recognition phases was identical to that of Experiment 1. The procedure for the DRM and general knowledge tasks was as follows.

DRM task. Participants were informed that on each trial they would see a list of words presented one-at-a-time on the computer screen, that they were to read each word aloud, and that they would subsequently be asked to write down as many words from the list as they could recall within a 2-minute time limit. Words were presented for 2 s each with a 1-s ISI separating the words. The ordering of the eight lists was random for each participant.

General knowledge task. Instructions stated that participants were to respond to each of a series of questions with a range of years within which they were “reasonably certain the event in question occurred, such that you would be comfortable giving this information to a friend if asked.” Responses were non-speeded. A 1-s ITI separated each trial. A single practice trial preceded the 50 test trials.

Results and Discussion

The general knowledge task data from five participants were removed from the analyses reported below. Three were new to North America, one had lived through an inordinate number of the events in question, and one gave several ranges beginning much earlier than 200 years ago despite instructions to the contrary. These participants’ DRM and recognition data were analyzed.

Recognition test means are displayed in Table 2. Performance on the two tests was very similar: sensitivity and bias were both statistically equivalent (both $t_s < 1$). Mean bias across all participants was again approximately neutral.

Table 3

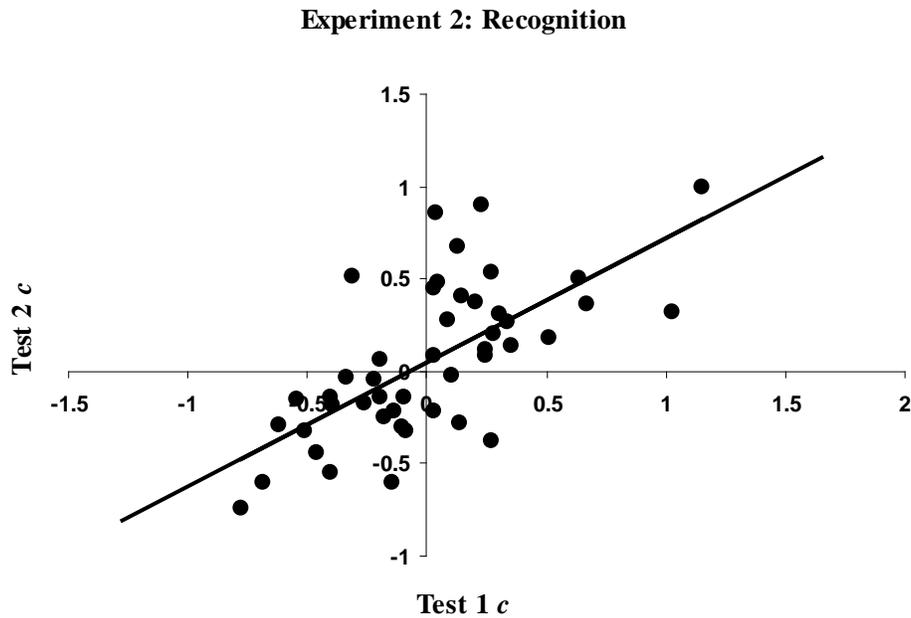
Recognition means in Experiment 2.

	H		FA		c		A_z	
	M	SD	M	SD	M	SD	M	SD
Test 1	.71	.16	.28	.14	.01	.41	.80	.10
Test 2	.70	.14	.28	.17	.06	.41	.79	.10

As in Experiment 1, variability in bias was substantial across subjects, with single test values ranging from 1.15 ($H = .40$, $FA = .02$) to -0.78 ($H = .92$, $FA = .56$). Test 1 c is plotted against Test 2 c in Figure 3. The correlation of bias across the two tests was again highly significant, $r(42) = .73$, $p < .001$, and directionally greater than in Experiment 1, when the two tests were only 10 minutes apart. Variability in the coefficients calculated from the supplementary bias measures was greater than in Experiment 1 (see Table 2), but each correlation was again indicative of a strong positive relationship, with three of the five correlations directionally higher than their counterparts in Experiment 1.

Figure 3

Correlation of recognition bias at Test 1 and Test 2 in Experiment 2.



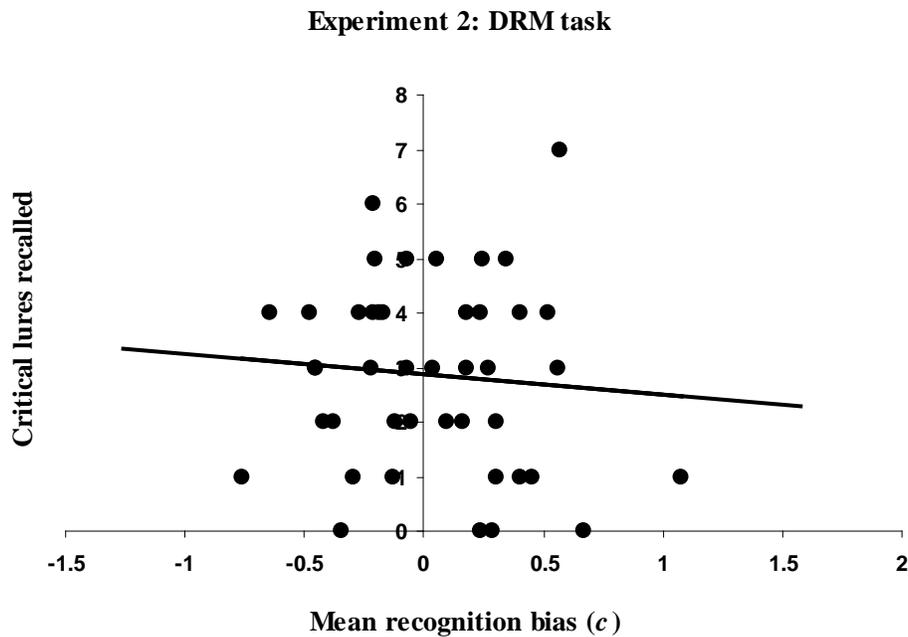
On the DRM test, participants falsely recalled an average of 2.86 critical lures out of 8 possible ($SD = 1.73$, range = 0 to 7). The correlation of the number of critical lures recalled and the mean of Test 1 c and Test 2 c for each participant (correcting for mean A_z) was negative but did not approach significance, $r(40) = -.12$, $p = .45$ (see Figure 4, top

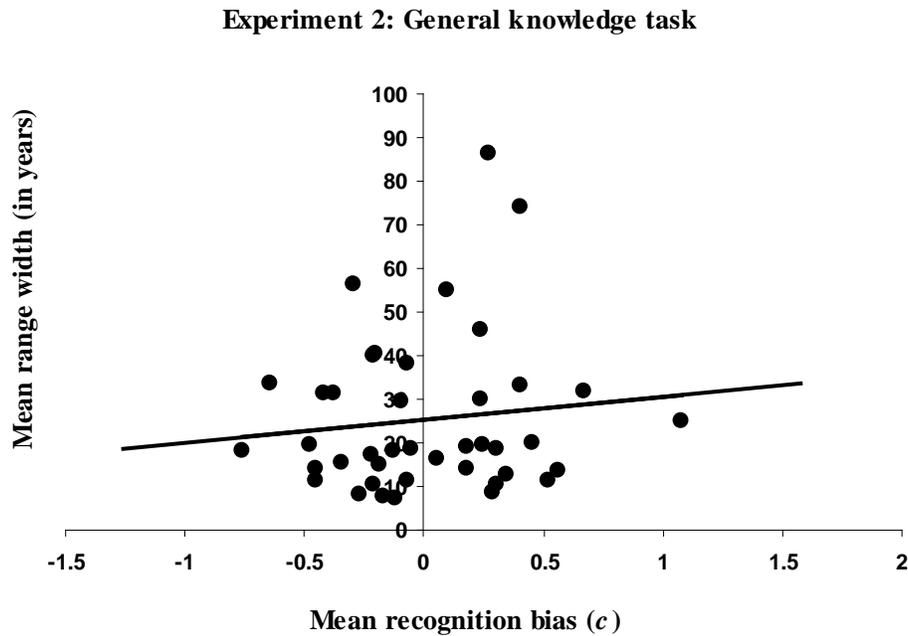
panel). The supplementary bias measures also yielded negative but very weak correlations (strongest $r = -.17$, lowest $p = .28$).

Across participants, the mean range width in the general knowledge task was 25.4 years ($SD = 17.9$, range = 7.3 to 86.2). The split-half reliability of the range width measure was .82. Mean range width was not significantly correlated with c , $r(38) = .12$, $p = .47$; see Figure 4, bottom panel), nor the supplementary measures (strongest $r = -.19$ [using FA], lowest $p = .24$).

Figure 4

Correlation of recognition bias and performance on non-recognition tasks in Experiment 2: frequency of DRM false recall (top panel) and mean range width in the general knowledge task (bottom panel).





While bias was highly correlated across a one-week interval, evidence of a relationship between bias and DRM and general knowledge task performance was not obtained. Unfortunately, the null relationships are difficult to interpret. Despite the tendency for participants to overestimate their own knowledge levels in the general knowledge task, prior knowledge may have driven variability in range sizes to a far greater degree than response bias. Even more problematic was the timing of Experiment 2, which coincided with lectures on the DRM paradigm in several of the psychology courses taken by our participants. Interviews conducted during the debriefing revealed that more than half of the participants came to the experiment with fresh insight that they should avoid recall of critical lures. We addressed these potential impediments to identifying a relationship between recognition bias and the DRM and grain size tasks in Experiment 4.

Experiment 3

In Experiments 1 and 2, the correlated bias measures were derived from two tests of word recognition, leaving open the possibility that bias is consistent for words (or, more generally, that it is consistent within the same stimulus domain), but differs unpredictably when the to-be-recognized stimuli change. To address this possibility, Experiment 3 included conditions in which two recognition study/test cycles varied in the class of materials used.

The stimulus domains chosen for the experiment were words and digital images of masterwork paintings. These materials are well suited to an examination of bias consistency across stimuli in two respects. First, words and paintings share few features beyond their visual presentation modality and contrast sharply along several dimensions: Paintings are richly detailed, complex in subject matter, and thematically (and sometimes emotionally) evocative, whereas the common word stimuli used in the present experiments possess none of these attributes. The use of such qualitatively distinct stimulus sets provides a strong test of the within-individual consistency of bias across materials.

A second advantage of words and paintings in providing a rigorous test of bias consistency is their tendency to elicit very different magnitudes of bias on recognition tests: Whereas common words tend to produce roughly neutral responding on average, paintings are associated with dramatic conservatism (Lindsay & Kantner, 2011). Note that bias consistency does not require that the obtained measure of bias is the same or even similar for a given participant across tests if the two tests use different stimuli. Rather, it requires that bias on Test 1 *predicts* bias on Test 2, such that a participant with

a more liberal than average word recognition bias should show a more liberal than average painting recognition bias. A finding that bias remains correlated across words and paintings despite sharp cross-stimulus differences would provide substantial evidence of trait-like stability across materials.

Method

Participants. A total of 143 undergraduates participated, each randomly assigned to one of four conditions: the Word-Word (WW) condition (words in the first study/test cycle, words in the second study/test cycle), the Painting-Painting (PP) condition, the Word-Painting (WP) condition, and the Painting-Word (PW) condition. The WW, PP, WP, and PW conditions included 40, 37, 35, and 31 participants, respectively.

Materials. Word stimuli were identical to those used in Experiments 1 and 2. Several hundred images of masterwork paintings were obtained from a computer-based memory training game called Art Dealer by permission of its creator, Jeffrey P. Toth. This set contains large, full color, high definition images of works by well-known artists from the 17th to early 20th centuries (e.g., Rembrandt, Matisse, Modigliani). Two hundred and four of these images, representing a wide array of artists, styles, and subject matter, were selected for use in Experiment 3. Very famous works (e.g., Van Gogh's self-portraits) were avoided. Paintings and words were assigned to study and test phase(s) by the same method as words in Experiments 1 and 2.

Procedure. The procedure was essentially identical to that of Experiment 1 with the exceptions of the materials manipulation and a slight modification to the filler task.

Results and Discussion

Recognition means are displayed in Table 3. Test 1 bias is plotted against Test 2 bias for each condition in Figure 5. Individuals' bias values were again marked by considerable variability. Ranges of c were 1.27 ($H = .31$, $FA = .02$) to -0.96 ($H = .88$, $FA = .77$) in the WW condition, 1.10 ($H = .33$, $FA = .04$) to -0.68 ($H = 1$ [uncorrected], $FA = .17$) in the PP condition, 0.90 ($H = .60$, $FA = .02$) to -1.21 ($H = .96$, $FA = .75$) in the PW condition, and 0.75 ($H = .46$, $FA = .08$) to -0.65 ($H = .92$, $FA = .46$) in the WP condition.

Table 4

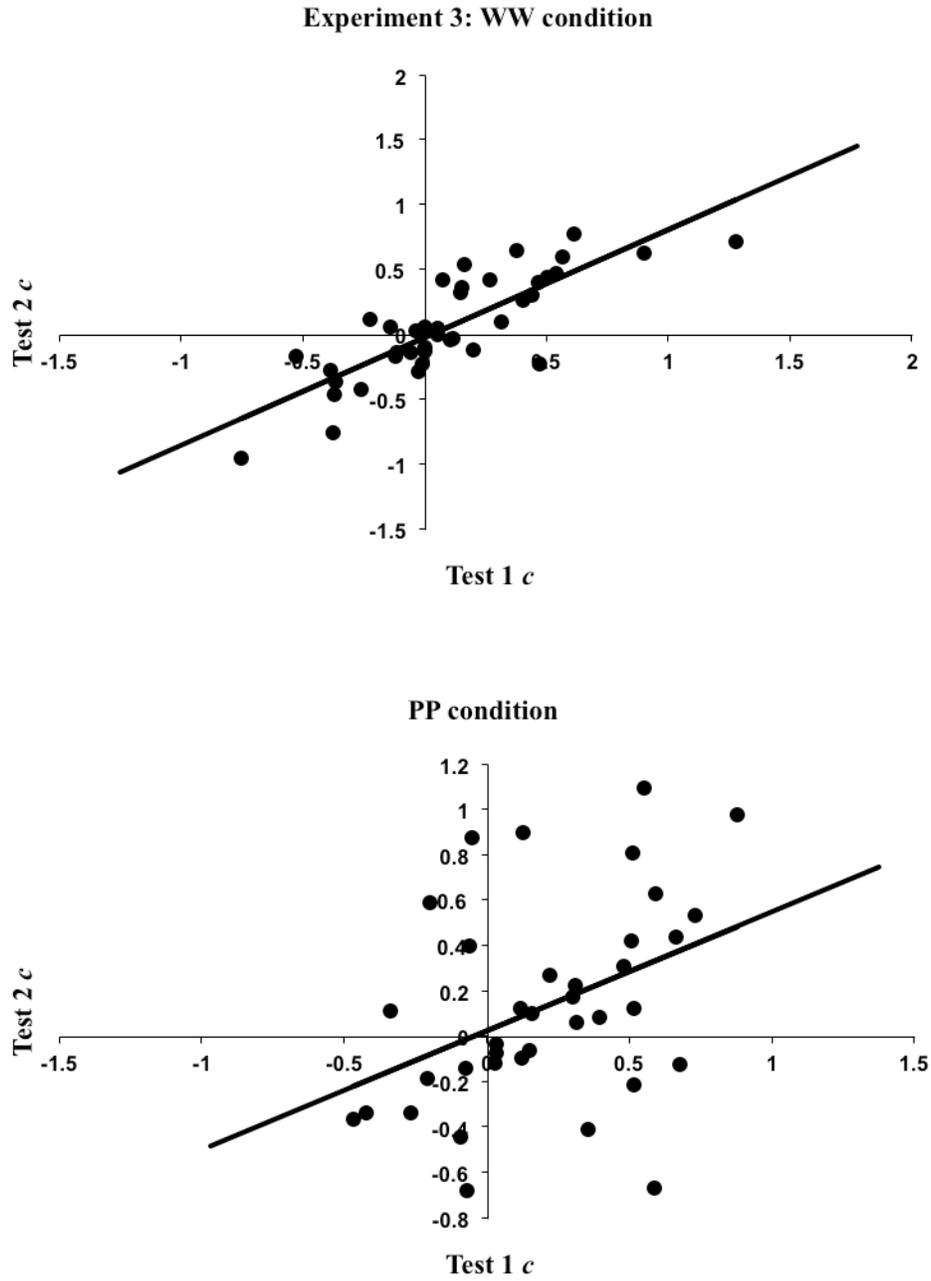
Recognition means in Experiment 3.

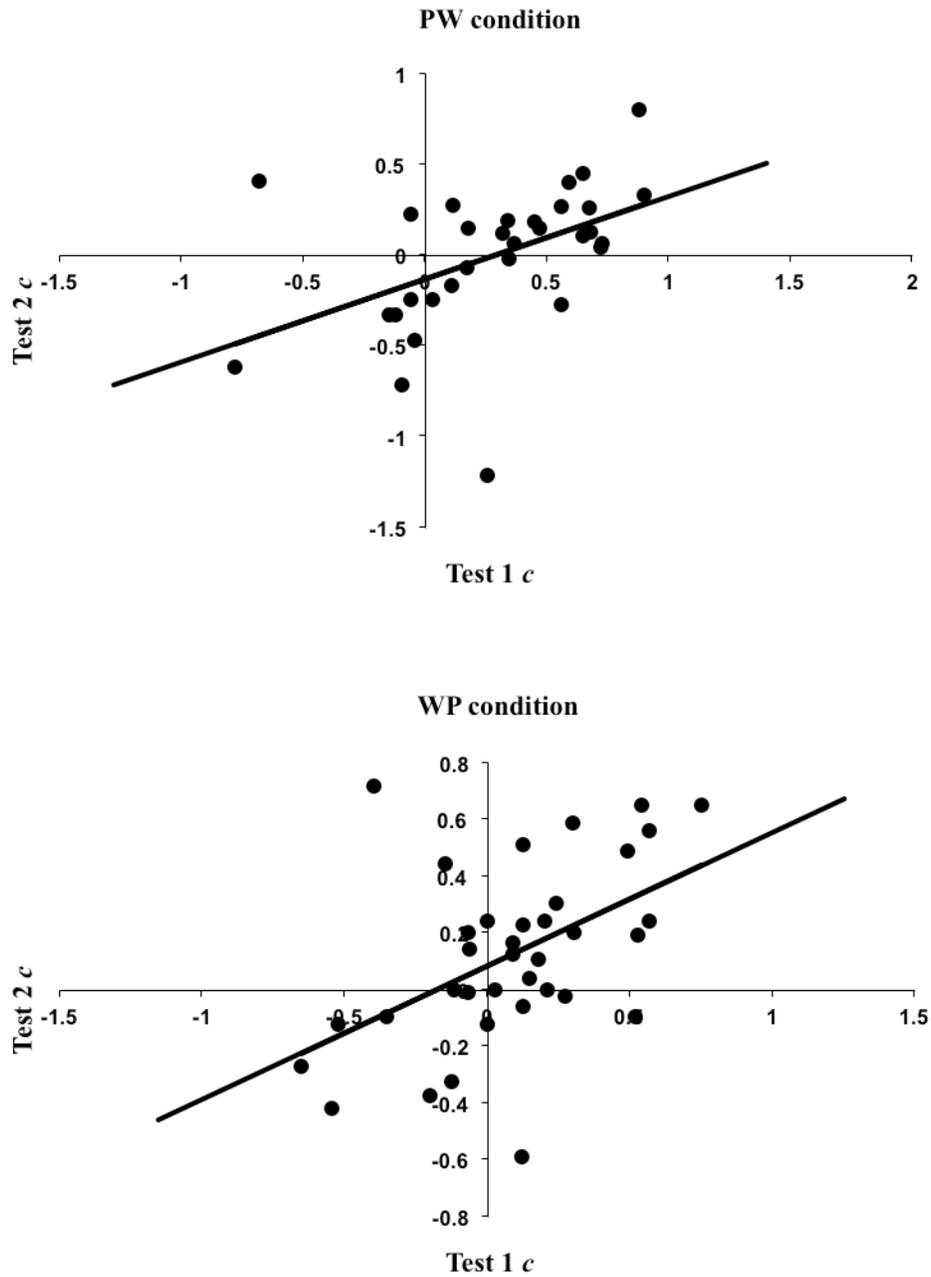
	H		FA		<i>c</i>		A_z	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
WW Condition								
Test 1 (Words)	.68	.15	.26	.15	.11	.39	.78	.12
Test 2 (Words)	.71	.13	.26	.17	.07	.40	.79	.09
PP Condition								
Test 1 (Paintings)	.69	.15	.19	.12	.21	.35	.82	.08
Test 2 (Paintings)	.77	.14	.16	.12	.14	.45	.87	.10
PW Condition								
Test 1 (Paintings)	.68	.13	.17	.12	.28	.41	.82	.08
Test 2 (Words)	.75	.14	.26	.17	.00	.40	.82	.09
WP Condition								
Test 1 (Words)	.70	.12	.24	.15	.09	.34	.80	.09
Test 2 (Paintings)	.76	.13	.16	.11	.13	.32	.86	.08

Note. WW = word-word, PP = painting-painting, PW = painting-word, WP = word-painting. Type of stimulus used in each test is in parentheses.

Figure 5

Correlation of recognition bias at Test 1 and Test 2 in Experiment 3.





WW condition. Neither A_z nor c differed significantly across tests, both $ps > .28$. c was significantly correlated across the tests, $r(36) = .81, p < .001$, replicating the findings of Experiments 1 and 2. The strength of the relationship was consistent across all bias measures (see Table 2).

PP condition. Sensitivity rose significantly from Test 1 to Test 2, $t(36) = -3.45, p$

$< .01$, while bias was unchanged ($t < 1$). c was positively correlated across tests, but the relationship did not survive the partialing out of A_z , $r(33) = .31$, $p = .07$. The supplementary measures yielded similar results: In all but one case (FA), the magnitude of the correlation was reduced below significance when controlling for sensitivity. The average p -value for the bias correlation across all five measures was .056.

PW condition. There was no significant difference in sensitivity on the painting and word tests ($t < 1$). As expected, painting bias was much more conservative than word bias, $t(30) = 3.837$, $p < .001$. The bias correlation across tests was significant, $r(27) = .64$, $p < .001$. The supplementary measures yielded concordant results.

WP condition. Group differences in sensitivity and bias followed the opposite pattern of the PW condition: sensitivity differed significantly, $t(34) = 3.246$, $p < .01$, while bias was statistically equivalent ($t < 1$). The correlation of c was again significant, $r(31) = .45$, $p < .01$. The spread of correlation coefficients across bias measures was unusually wide (range = .31 to .75), partially due to differences across measures in the presence or absence of score removals via the 3 SD cutoff. Correlations were significant in four of five cases, however, and were substantially positive in all cases.

Thus, Test 1 bias remained strongly predictive of Test 2 bias when different materials were used in the two tests. Stability was not equivalent across all four conditions, however. Fisher's tests on bias correlations averaged across the five measures confirmed that the magnitude of the WW correlation ($M = .76$) was significantly greater than that of the PP condition ($M = .33$), $z = 2.59$, $p < .01$. Correlations in the PW ($M = .60$) and WP ($M = .52$) conditions did not differ significantly from those of any other condition. The directionally lower stability in the PW and WP conditions than the WW

condition is not surprising and suggests that consistency in stimuli contributes to consistency in bias across tests. The relatively weak bias relationship observed in the PP condition, however, was an unexpected result, given that materials did not differ in this condition. We are currently investigating the possibility, suggested by the increase in sensitivity across tests, that some participants in the PP condition used the experience gained in the first study/test cycle to alter their approach to the second, producing a change in response bias across the tests and thereby reducing the correlation. For the present purposes, the important point is that even when the two tests used different materials individuals who were conservative (liberal) on one test tended to be conservative (liberal) on the other.

Experiment 4

While Experiments 1-3 provided support for the characterization of recognition bias as a trait, evidence in the form of generalization beyond recognition memory has been absent. In Experiment 2, bias was uncorrelated with performance on a DRM free recall test and a general knowledge task tapping strategic adjustments of grain size, two tasks hypothesized to involve the same evidence criterion at work in producing recognition bias. Experiment 4 returned to the DRM and grain size paradigms under conditions expected to increase the likelihood of detecting a relationship with recognition bias if one exists.

The DRM task in Experiment 4 was unchanged from Experiment 2, but the timing of the experiment within the course of the academic term was believed to better suit the DRM paradigm. Experiment 2 took place in the middle of an academic term, and, as noted above, overlapped with lectures on the DRM paradigm in some psychology

courses, giving many participants recent pre-experimental insight regarding the nature of the test lists. In experimental settings, warnings about the critical lure decrease DRM false alarm rates (see Starns, Lane, Alonzo, & Roussel [2007] for a review). Variability in this foreknowledge across Experiment 2 participants may have driven differences in false recall of critical lures, undermining the detection of other mediators (e.g., inherent response bias). Experiment 4 was conducted in the first half of the fall semester, at which time very few of the introductory psychology students who constitute the majority of research participants had been familiarized with the DRM paradigm. As in Experiment 2, liberal recognizers were predicted to recall a higher proportion of critical lures than conservative recognizers.

The general knowledge task was revised for Experiment 4 on the suspicion that the null result in Experiment 2 arose from the use of range size as the dependent measure. Given that participants were free to choose whatever range sizes they felt were appropriate, this measure might have been driven by variability in knowledge of the subject matter (either actual or assumed) to the extent that any potential relationship with response bias was obscured. Moreover, the incentive for using particular interval sizes in Experiment 2 (accurate yet informative answers) was merely implied; any response bias associated with the general knowledge task might be brought out more effectively with more explicit consequence for responses. Therefore, a new version of the task was created in which each question was accompanied by two response options, one of which was correct (e.g., “What year did CBC make its first television broadcast? a. 1953 b. 1963”). Participants were informed that they would gain 10 cents for every correct answer and lose 10 cents for every incorrect answer. They were also given the right to

“pass” on any question to which they did not feel confident giving an answer (called *report option* by Koriat & Goldsmith [1994]), in which case they neither gained nor lost money on that trial. In this scenario, any question to which the subject does not have prior knowledge of the answer is a small gamble in which giving a response incurs risk that can be avoided with the exercise of report option.

The dependent measure of interest is the proportion of trials on which liberal versus conservative recognizers use the pass option. Conservative recognizers, assumed to require more memory evidence than liberal recognizers before committing to an “old” judgment, were hypothesized to require more confidence in their knowledge of the answer to a given question before committing to the gamble. Thus, conservative recognizers should exercise report option significantly more often. This task is particularly appealing, given the goals of this experiment, in that risk-taking behaviors have been associated with extraversion (Patterson & Newman, 1993), and extraversion, in turn, has been associated with a liberal recognition bias (Gillespie & Eysenck, 1980).

Method

Participants. There were 50 participants in Experiment 4.

Materials. The same words used in Experiments 1-3 served as recognition task materials. DRM task materials were identical to those of Experiment 2. The general knowledge task included 50 questions, each with two response alternatives.

Approximately half of the questions were retained from Experiment 2; in order to increase variety within the task, the remainder of the set comprised questions requiring numerical responses other than names of years. These questions were drawn from the original pool of 208 (see Experiment 2 Method).

Two response alternatives were prepared for each question. One alternative was always the correct answer. The second option was chosen by the first author and was designed to pose as a plausible alternative that would generate uncertainty without making the task overly difficult. Generally, the incorrect alternative was a value of moderate numerical distance from the correct answer.

Procedure. Experiment 4 consisted of three stages: a recognition study/test cycle, the DRM task, and the general knowledge task with report option. Participants were tested in groups of 1 to 3, a measure taken to increase the efficiency of data collection given the unusual length of the experiment. In sessions of more than one participant, a second experimenter was present and aided in the transition between phases. The order of the recognition, DRM, and general knowledge tasks was counterbalanced across groups; within groups, each participant completed the tasks in the same order.

The recognition task followed a procedure identical to that of Experiments 1-3 (note, however, that only one study/test cycle was included in Experiment 4). The procedure for the DRM task was identical to that of Experiment 2, with two exceptions required by the group testing format. First, participants were asked to read list words silently. Second, a black and white flashing screen (rather than an auditory tone) alerted participants to the end of each two-minute recall period.

The general knowledge task was similar to the one used in Experiment 2, with differences reflecting the change to a two-alternative forced choice (2AFC) response format with report option. Task instructions were analogous to those in Experiment 2, with an additional component informing participants that they would gain ten cents for each correct response, lose ten cents for each incorrect response, and gain or lose nothing

by choosing to “pass” on answering a given question. Instructions emphasized that a negative balance at the end of the task would not result in any loss of money.

Questions were again presented near the top of the screen with two boxes positioned underneath; these boxes contained response options A and B. Near the bottom of the screen appeared the words, “Press spacebar to pass.” Participants chose an answer by entering ‘a’ or ‘b’ or passed by hitting the spacebar. Passing initiated the next trial. Selection of one of the response alternatives prompted the appearance of a confidence scale ranging from 50% to 100% near the top of the screen. Participants indicated their confidence in the selected answer via key press, initiating the next trial.

Results and Discussion

The data of two participants near chance in recognition accuracy were removed from subsequent analyses. The general knowledge test data of one additional participant were deleted due to a failure to follow task instructions. Therefore, the following analyses included 48 participants in the DRM and recognition tasks and 47 participants in the general knowledge task.

Table 5

Recognition means in Experiment 4.

	H		FA		<i>c</i>		A_z	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Test 1	.66	.15	.28	.15	.11	.41	.76	.08

Table 6

Correlations of bias and DRM false recall as a function of bias measure in Experiment 4.

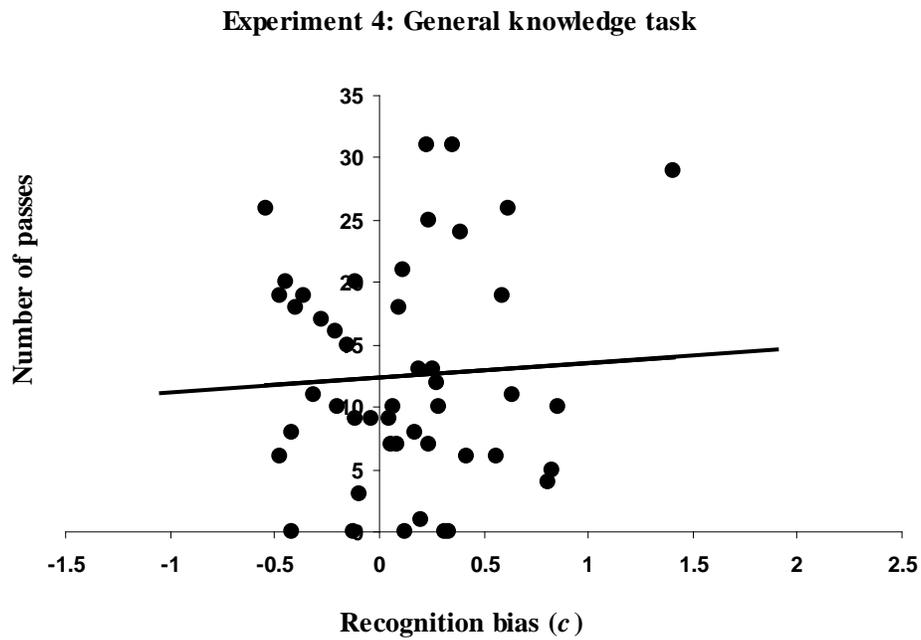
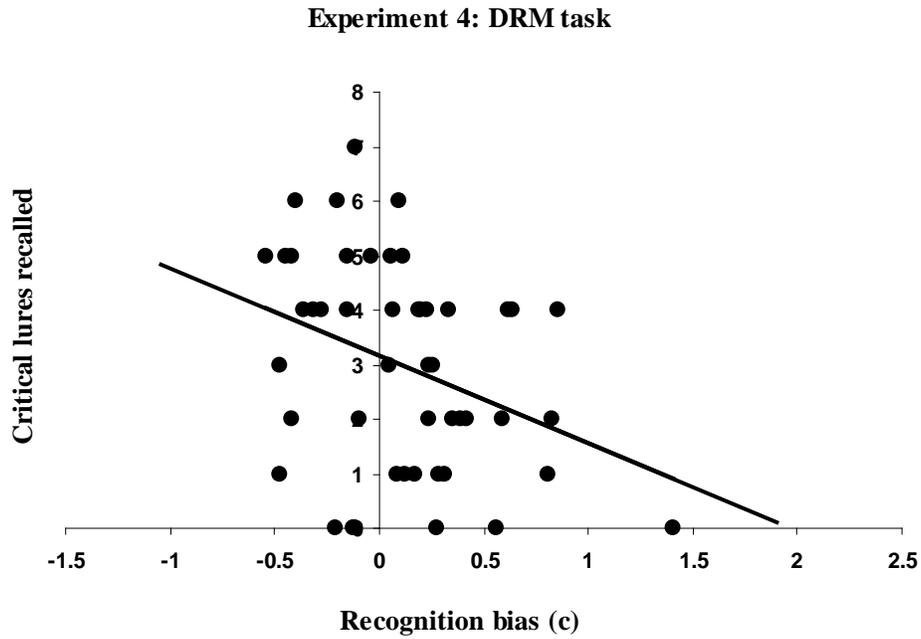
	c	c_a	$\ln(\beta)$	B''	FA
DRM false recall	-.32*	-.39**	-.23	-.33*	.43**

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

Group recognition measures are displayed in Table 4. Values of c ranged from 1.40 ($H = .31$, $FA = 0$ [uncorrected]) to -0.53 ($H = .81$, $FA = .58$). As in Experiments 1-3, all bias scores greater than 3 SD from the mean for a given parameter were removed prior to correlational analysis (0.83% of all scores) and all bias correlations controlled for A_z . In the DRM task, participants falsely recalled an average of 2.98 critical lures out of 8 possible ($SD = 1.92$, range = 0 to 7). Unlike in Experiment 2, values of c were significantly correlated with frequency of false recall, $r(44) = -0.32$, $p < .05$, with a negative relationship indicating that increasing liberality of recognition bias was associated with increasing frequency of false recall (see Figure 6). Results using the supplementary bias measures were consistent with those using c (see Table 5); only $\ln(\beta)$ yielded a non-significant relationship. The correlation of A_z and DRM false recall (controlling for c) was nearly reliable, $r(44) = -0.29$, $p = .051$. These analyses suggest that both bias and sensitivity are related to recall of critical lures in the DRM paradigm.

Figure 6

Correlation of recognition bias and frequency of DRM false recall (top panel) and correlation of bias and number of passes in the general knowledge task (bottom panel), Experiment 4.



Participants chose to pass on an average of 12.53 ($SD = 8.69$) out of 50 general knowledge questions (25.1%). Individual participants' use of the pass option ranged from 0 to 31 times. When questions were answered with one of the two response alternatives, mean accuracy was 68.9% ($SD = 10.5\%$) and mean confidence was 58.2% ($SD = 12.5\%$). Split-half reliability of the pass measure was high (.81). No relationship was detected between c and frequency of passing ($r = 0.06$), accuracy ($r < .001$), or confidence ($r = -.08$). The strongest relationship yielded by the supplementary bias estimates was $r = -.14$, $p = .35$. Recognition sensitivity was also unrelated to these measures (strongest $r = -.06$).

One further analysis concerned individuals' frequencies of passing versus giving responses at a 50% confidence level. Since both types of responses signify an expectation of chance-level ability in answering a question, it was expected that this comparison would discriminate liberal and conservative responders: the former should be more likely to risk an incorrect response while the latter should be more likely to pass. However, preference for the pass option (the number of passes minus the number of 50% confidence responses) was uncorrelated with c ($r = 0.09$) and with the supplementary measures (strongest $r = .08$)

Experiment 4 provided the first indication of a relationship between recognition response bias and performance on a non-recognition task. Individuals using a more lax criterion for calling items old in a recognition test also used a more lax standard for recalling related but non-presented list items in a DRM procedure. Though replication of this relationship is warranted, its presence in Experiment 4 is consistent with the suspicion that the lack of relationship in Experiment 2 was due to the noise added by widespread foreknowledge of the task (although we note that, contrary to that

speculation, the absolute rate of DRM intrusions was not higher in Experiment 4 than in Experiment 2; it may simply be that the bias-DRM relationship is weak and that there was insufficient power to detect it in Experiment 2).

No relationship was observed between bias and performance in the general knowledge task, consistent with the results of Experiment 2. It might be the case that conservatism in a recognition task and conservatism in a general knowledge test have independent cognitive substrates, and that trait-like stability in recognition bias is not relevant to the class of decisions exemplified in the general knowledge task.

Alternatively, the general knowledge task might not have elicited systematic biases relevant to a required evidence criterion. We consider this possibility in the General Discussion.

General Discussion

The construct of response bias provides a basis for understanding how recognition decisions are reached under conditions of uncertainty and how they are affected by factors independent of memory. The present experiments examined whether measured response bias might also provide a basis for understanding individual recognizers. The perspective taken was a departure from that of most previous research on bias: Instead of asking what factors influence bias from without (e.g., task and stimulus manipulations), the present work asked to what extent bias is founded within an individual as a stable cognitive trait.

The present experiments provided evidence of substantial within-individual consistency. Experiment 1 established that one's response bias does not vary freely from one recognition test to the next; rather, bias on an initial test is highly predictive of bias

on a subsequent test. Experiment 2 extended this finding by demonstrating a similar correlation of bias on two tests one week apart. While this comparison spans separate experiments and groups of participants, it is nonetheless worth emphasizing that the differences between the ten-minute and one-week intervals transcend duration. With a 10-minute interval, participants remain within the context of the experiment between tests, changing only the task with which they are engaged; when one week separates the tests, by contrast, participants return to the laboratory for Test 2 having accumulated a week of life experiences since Test 1. The fact that these two intervals were associated with similar correlations of bias is strongly suggestive of trait-like stability.

In Experiment 3, significant bias correlations obtained even with substantial differences in the stimuli across two tests (the PW and WP conditions). The presence of large, significant correlations across stimulus classes sharing few common features suggests that participant predisposition accounts for a good deal of the variance in bias. The similarity of the average correlations observed in the PW and WP conditions is sensible given the identical content of the two conditions. It is informative, however, in light of the divergent trends distinguishing the two conditions. The PW condition showed a sizable shift in bias from paintings to words but no change in sensitivity; the WP condition showed no shift in bias across stimuli (contrary to expectation) but significantly greater sensitivity to paintings than words. Bias stability, then, is apparently not reliant on a match in general discrimination or response bias across tests. The former finding is consistent with the signal detection theory assumption that discrimination and bias are independent properties, while the latter confirms the hypothesis that bias need not be similar across tests to have a predictive relationship across tests.

Is the bias observed in recognition tests a special case of a general proclivity to act on the basis of a certain level of evidence? Experiments 2 and 4 explored this question by allowing recognition bias to be correlated with performance on non-recognition tasks. These experiments produced mixed results. Bias did not predict DRM false recall in Experiment 2, but the interpretation of this result was clouded by the fact that many participants had very recently learned of the DRM paradigm in classroom lectures. Experiment 4 was timed to resolve this concern and did show a significant tendency for participants with a more liberal recognition bias to recall more critical lures in the DRM task. These findings are noteworthy in two respects: first, they represent a correspondence of recognition bias with performance on a task outside of recognition memory (i.e., free recall), suggesting a common cognitive substrate. Second, they suggest that liberality in response bias is related to general false memory proneness, which has itself been discussed as a possible trait (e.g., Geraerts et al., 2009).

Recognition bias was not, however, associated with degree of conservatism in answering general knowledge questions. Whether conservatism was operationalized as the use of wide ranges in numerical estimates (Experiment 2) or the frequent use of the “pass” option to avoid the risk of an incorrect response (Experiments 4), there was no apparent relationship with recognition bias. As noted earlier, these null findings may result from a true lack of relationship between recognition bias and risk tolerance in estimation. Based on the current findings, it might be the case that bias correlates with performance on memory tasks that are episodic in nature (e.g., DRM), but not with other forms of decision making, such as estimation. However, it might be the case that the general knowledge task did not motivate patterns of responding driven by a liberal or

conservative bias. The payoffs and losses associated with individual responses were small (10 cents), and participants were aware that they could not lose money on balance. A design fostering loss aversion (Kahneman & Tversky, 1984) or including trial-by-trial feedback might increase investment in responses and bring out a decisional bias related to recognition bias.

We view the current evidence for within-individual bias stability as broadly consistent with past research (cited above) indicating that certain groups of participants defined by inherent characteristics (e.g., Alzheimer's patients, panic disorder patients) can be characterized as liberally biased recognizers. We believe the present results are also consistent with the Blair et al. (2002) finding of substantial within-individual consistency in false recognition of critical lures from the same DRM lists tested two weeks apart, although Blair et al. argued against response bias consistency as an account of their results (p. 595). Our conclusions seem at odds, however, with those of Salthouse and Siedlecki (2007), who did not find predictive relationships of false DRM recognition across tests differing in materials or reliable correlations between false recognition and established individual difference variables. The Salthouse and Siedlecki (2007) results suggest that recognition of critical lures is governed largely by task specific factors, while our results point to an element of response bias that is inherent to an individual and is consistent across time and task variations. These findings are not irreconcilable, however: while response bias and the tendency to commit DRM errors are likely related to one another (Miller, Guerin, & Wolford, 2011), the relationship is imperfect. In the current Experiment 4, for example, bias on a recognition test predicted the number of critical lures recalled by participants in a DRM free recall task, but the modest strength of the

relationship ($r = -0.32$) suggests independent properties as well. It might be the case that individual differences in recognition response bias are more stable than individual differences in DRM false recognition. Future research should determine whether bias is related to the individual difference measures tested by Salthouse and Siedlecki (2007).

The possibility that response bias is a trait carries at least three general implications. First, it suggests that bias proclivities at the level of the individual represent a source of variance in recognition memory experiments, one that can be identified in models of recognition. Models of individual subject data can be constrained if an individual's predisposition to bias is identified beforehand. If not, trait bias suggests a psychological interpretation of model parameters that index response bias.

Second, the concept of trait response bias suggests that an individual's bias on a recognition test is a product of both inherent tendencies and any experimental manipulations that affect bias. The relative influence of predisposition and a given experimental manipulation in determining a subject's overall level of bias is an open question, but there are clearly scenarios in which the two factors could come into conflict. For example, a growing literature has demonstrated that participants tend not to make appropriate criterion shifts in response to changing classes of items (e.g., strongly encoded versus weakly encoded) when the shifts must occur within a single test list (Stretch & Wixted, 1998; Verde & Rotello, 2007). Generally, within-list criterion shifts are not observed unless the two item classes are strikingly different (Bruno, Higham, & Perfect, 2009; Singer, 2009; Singer & Wixted, 2006) or corrective feedback is administered (Rhodes & Jacoby, 2007; Verde & Rotello, 2007). Participants' resistance to within-list criterion shifting might be a partial result of inherent bias tendencies that

anchor shifting behavior. Similarly, the remarkable unwillingness of participants to exercise appropriate levels of bias even when they are aware that test lists are composed only of targets or lures (Cox & Dobbins, 2011) might be explained in part by such an anchoring effect. Even when manipulations such as feedback, instructional motivation, and payoffs are successful in moving response bias, such shifts are usually sub-optimal (Aminoff et al., in press; Healy & Kubovy, 1978), another potential influence of trait bias.

The third and most significant implication of the results presented here is that individuals can be placed along a continuum of recognition bias from more conservative to more liberal, and, critically, that this placement characterizes the *individual*, not just his or her performance on a given test. An important question is whether the trait-like qualities of response bias are limited to recognition memory measures or extend to multiple classes of decisions. Preliminary evidence for the generalization of bias beyond recognition memory was obtained with the DRM free recall task used in Experiment 4, but not with the general knowledge task. Further work with such tasks and with personality measures relevant to criterial levels of decision evidence (e.g., impulsivity, maximizing tendencies [Simon, 1956]) will help to determine the extent to which recognition bias is a special case of a more general, intra-individually stable decision making heuristic.

The idea that people are inherently disposed to some tendency or other is commonplace in the domain of personality psychology; indeed, the very notion of a personality trait implies that persons possess enduring attributes that guide their thinking and behavior across different situations. The notion of a memory trait is less explored.

The current experiments suggest that response bias observed in a recognition test is tethered to a predisposition internal to the recognizer, and that a complete theory of recognition memory performance should account for this factor.

References

- Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting--with and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1224-1245.
- Aminoff, E. M., Freeman, S., Clewett, D., Tipper, C., Frithsen, A., Johnson, A., Grafton, S. T., & Miller, M. B. (in press). Individual differences in shifting decision criterion: A recognition memory study. *Memory & Cognition*.
- Benjamin, A., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language*, *51*, 159-172.
- Beth, E. H., Budson, A. E., Waring, J. D., & Ally, B. A. (2009). Response bias for picture recognition in patients with alzheimer disease. *Cognitive and Behavioral Neurology*, *22*, 229-235.
- Blair, I. V., Lenton, A. P., & Hastie, R. (2002). The reliability of the DRM paradigm as a measure of individual differences in false memories. *Psychonomic Bulletin & Review*, *9*, 590-596.
- Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *The Quarterly Journal of Experimental Psychology*, *29*, 461-473.
- Bruno, D., Higham, P. A., Perfect, T. J. (2009). Global subjective memorability and the strength-based mirror effect in recognition memory. *Memory & Cognition*, *37*, 807-818.

- Budson, A. E., Todman, R. W., Chong, H., Adams, E. H., Kensinger, E. A., Krangel, T. S., et al. (2006). False recognition of emotional word lists in aging and alzheimer disease. *Cognitive and Behavioral Neurology*, *19*, 71-78.
- Carlin, M. T., Toglia, M. P., Wakeford, Y., Jakway, A., Sullivan, K., & Hasel, L. (2008). Veridical and false pictorial memory in individuals with and without mental retardation. *American Journal on Mental Retardation*, *113*, 201-213.
- Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *33A*, 497-505.
- Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *The Journal of Abnormal and Social Psychology*, *60*, 151-174.
- Cox, J. C., & Dobbins, I. G. (2011). The striking similarities between standard, distractor-free, and target-free recognition. *Memory & Cognition*, *39*, 925-940.
- Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in Cognitive Science*, *4*, 135-150.
- Dougal, S., & Rotello, C. M. (2007). 'Remembering' emotional words is based on response bias, not recollection. *Psychonomic Bulletin & Review*, *14*, 423-429.
- Egan, J. P. (1958). Recognition memory and the operating characteristic. *USAF Operational Applications Laboratory Technical Note*, 58-51.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833-848.
- Geraerts, E., Lindsay, D. S., Merckelbach, H., Jelicic, M., Raymaekers, L., Arnold, M. M., et al. (2009). Cognitive mechanisms underlying recovered-memory experiences of childhood sexual abuse. *Psychological Science*, *20*, 92-98.

- Gillespie, C. R., & Eysenck, M. W. (1980). Effects of introversion–extraversion on continuous recognition memory. *Bulletin of the Psychonomic Society, 15*, 233-235.
- Gold, C. A., Marchant, N. L., Koutstaal, W., Schacter, D. L., & Budson, A. E. (2007). Conceptual fluency at test shifts recognition response bias in Alzheimer's disease: Implications for increased false recognition. *Neuropsychologia, 45*, 2791-2801.
- Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General, 131*, 73-95.
- Grier, J. B. (1971). Non-parametric indices for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*, 424-429.
- Harvey, L. O., Jr. (2005). Parameter estimation of signal detection models: RscorePlus (Version 5.6.0). [Computer software and manual]. Retrieved from <http://psych.colorado.edu/~lharvey/html/software.html>
- Healy, A. F., & Kubovy, M. (1978). The effects of payoffs and prior probabilities on indices of performance and cutoff location in recognition memory. *Memory & Cognition, 6*, 544-553.
- Hockley, W. E. (2011). Criterion changes: How flexible are recognition decision processes? In P. Higham & J. Leboe (Eds.), *Constructions of remembering and metacognition: Essays in honor of Bruce Whittlesea*. Houndmills, UK: Palgrave Macmillan.
- Huh, T. J., Kramer, J. H., Gazzaley, A., & Delis, D. C. (2006). Response bias and aging on a recognition memory task. *Journal of the International Neuropsychological Society, 12*, 1-7.

- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341-350.
- Kantner, J., & Lindsay, D. S. (2010). Can corrective feedback improve recognition memory? *Memory & Cognition*, *38*, 389-406.
- Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, *123*, 297-315.
- Kramer, J. H., Rosen, H. J., Du, A., Schuff, N., Hollnagel, C., Weiner, M. W., et al. (2005). Dissociations in hippocampal and frontal contributions to episodic memory performance. *Neuropsychology*, *19*, 799-805.
- Lindsay, D. S., & Kantner, J. (2011). A search for influences of feedback on recognition of music, poetry, and art. In P. Higham & J. Leboe (Eds.), *Constructions of remembering and metacognition: Essays in honor of Bruce Whittlesea*. Houndmills, UK: Palgrave Macmillan.
- Macmillan, N. A. (1993). Signal detection theory as data analysis method and psychological decision model. In G. Keren, C. Lewis, G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues*. (pp. 21-57). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.
- Macmillan, N. A., & Creelman, C. D. (1996). Triangles in ROC space: History and theory of “nonparametric” measures of sensitivity and response bias. *Psychonomic Bulletin & Review*, *3*, 164-170.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed.)*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

- 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*, 185-199.
- Miller, M. B., Guerin, S. A., & Wolford, G. L. (2011). The strategic nature of false recognition in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1228-1235.
- Moritz, S., Woodward, T. S., Jelinek, L., & Klinge, R. (2008). Memory and metamemory in schizophrenia: A liberal acceptance account of psychosis. *Psychological Medicine*, *38*, 825-832.
- Patterson, C. M., & Newman, J. P. (1993). Reflectivity and learning from aversive events: Toward a psychological mechanism for the syndromes of disinhibition. *Psychological Review*, *100*, 716-736.
- Qin, J., Ogle, C. M., & Goodman, G. S. (2008). Adults' memories of childhood: True and false reports. *Journal of Experimental Psychology: Applied*, *14*, 373-391.
- Ratcliff, R., Sheu, C., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*, 518-535.
- 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*, 305-320.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803-814.

- Rotello, C. M., & Macmillan, N. A. (2008). Response bias in recognition memory. In B. H. Ross (Ed.), *Skill and strategy in memory use*. (pp. 61-94). San Diego, CA US: Elsevier Academic Press.
- Salthouse, T. A., & Siedlecki, K. L. (2007). An individual differences analysis of false recognition. *American Journal of Psychology*, *120*, 429-458.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129-138.
- Singer, M. (2009). Strength-based criterion shifts in recognition memory. *Memory & Cognition*, *37*, 976-984.
- Singer, M., & Wixted, J. T. (2006). Effect of delay on recognition decisions: Evidence for a criterion shift. *Memory & Cognition*, *34*, 125-137.
- Stadler, M. A., Roediger, H. L. III, & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, *27*, 494-500.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137-149.
- Starns, J. J., Lane, S. M., Alonzo, J. D., & Roussel, C. C. (2007). Metamnemonic control over the discriminability of memory evidence: A signal detection analysis of warning effects in the associative list paradigm. *Journal of Memory and Language*, *56*, 592-607.
- Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1379-1396.

- Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 582-600.
- Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit and false alarm rate: The accuracy, precision, and robustness of d' , A_z , and A' . *Perception and Psychophysics*, 68, 643-654.
- Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition*, 35, 254-262.
- Windmann, S., & Krüger, T. (1998). Subconscious detection of threat as reflected by an enhanced response bias. *Consciousness and Cognition*, 7, 603-633.
- Woodard, J. L., Axelrod, B. N., Mordecai, K. L., & Shannon, K. D. (2004). Value of signal detection theory indexes for wechsler memory scale-III recognition measures. *Journal of Clinical and Experimental Neuropsychology*, 26, 577-586.

Author Note

Justin Kantner and D. Stephen Lindsay, Department of Psychology, University of Victoria, Victoria, British Columbia, Canada.

This research was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to Steve Lindsay. We thank Mayumi Okamoto, Caitlin Malli, Graeme Austin, and Emily Cameron for their assistance with data collection.

Justin Kantner is now at the University of California, Santa Barbara.

Correspondence concerning this article should be addressed to Justin Kantner, Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, California, 93106. Email: kantner@psych.ucsb.edu. Phone: (805) 845-1906.