Journal of Experimental Psychology:
Learning, Memory, and Cognition

© 2016 American Psychological Association
0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000223

Individual Differences in Incorrect Responding and the Ability to Discriminate the Source of the Products of Retrieval

Tanya R. Jonker
University of Waterloo

When memory is tested, researchers are often interested in the items that were correctly recalled or recognized, while ignoring or factoring out trials where one "recalls" or "recognizes" a nonstudied item. However, intrusions and false alarms are more than nuisance data and can provide key insights into the memory system. The present article reports 2 experiments demonstrating that people are remarkably consistent in the rate at which they incorrectly report memory for nonstudied items, even across a range of differing stimuli and test features. Experiment 1 found that individual differences in false alarms and intrusions were strongly related, even while controlling for the shared influence of memory ability on these incorrect response types. Furthermore, intrusion rate was found to be related to response bias as well, but this effect was suppressed by the shared influence of memory ability, demonstrating that an independent measure of memory ability is an important control in investigations into individual differences in response bias. Experiment 2 revealed that the relations between intrusions and false alarm rate and response bias were entirely explained by one's ability to discriminate episodic memories from semantic generation. This work links together previous work on individual differences in intrusions and in false alarms, and highlights the ability to identify the source of a memory as the key cognitive trait underlying incorrect response styles on various memory tests.

Keywords: individual differences, recognition, intrusions, source monitoring, retrieval discrimination

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

When thinking about memory, people tend to be interested in what one correctly remembers and what one forgets. However, relatively less attention is paid to cases where one "remembers" information that was not actually encoded. Yet, in memory research, participants regularly report remembering items that were never presented. In the case of research on recall, these extralist items—sometimes referred to as *false recall* or *intrusions*—are frequently ignored. In the case of research on recognition memory, false recognition of nonstudied items—referred to as *false alarms*—is most often accounted for statistically, with the goal of removing its influence from the data and having measures of true memory only (Abdi, 2007; Wixted, 2007). A popular method for accounting for false recognition comes from signal detection theory; this method accounts for response bias, which is a participant's tendency to identify items as studied (or "old") versus nonstudied (or "new"). Thus, false alarms are often treated as an artifact of one's tendency to respond liberally (more "old" responses) or conservatively (more "new" responses). A participant with a liberal response bias will tend to report more items as having been studied, which will result in many hits for studied items but also a relatively large number of false alarms to nonstudied items, whereas a conservative responder will report more items as being new, resulting in fewer hits and fewer false alarms.

Thus, two different types of memory tests tend to be associated with two very different treatments of incorrect responses: In recall research, incorrect responses tend to be ignored whereas, in recognition research, their influence tends to be removed from the data mathematically. Despite different methods, both approaches treat incorrect responses as nuisance data rather than responses that are interesting in and of themselves. The exception to this is a large literature on false memory, where "memory" for items that were never in fact presented is the focus of study. In this literature, tasks and stimuli are often designed to evoke specific types of intrusions (and sometimes false alarms). For example, in the Deese-Roediger-McDermott paradigm (DRM; Deese, 1959; Roediger & McDermott, 1995), participants are presented with a list of words that are related to one critical—but never presented—word, and interest is in false memory for the critical lure while other types of intrusions are often ignored (for a review, see Gallo, 2010). In the present article, the focus is not on incorrect recall produced by false memory paradigms designed to evoke intrusions, but instead on intrusions and false alarms produced in standard recall or recognition tests, with the hypothesis being that these incorrect responses are more than just noise in the data.

People are very consistent in the proportion of false alarms that they make on recognition tests, demonstrating reliability over time and across different stimuli (e.g., pictures vs. words; Kantner &

This research was supported by a Vanier Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council (NSERC) of Canada. I am very grateful to Dr. Colin MacLeod for feedback on an earlier version of the manuscript and to Dr. Erik Woody for his advisement on data analysis and modeling. I also extend thanks to Bonnie Armstrong, Iulia Banica, Sheldon Frey, Ailin Huang, Jennah Kerman, Anna Liargovas, Mariah Nanton, Melinda Recchia, and Trish Varao Sousa for their assistance with data collection.

Correspondence concerning this article should be addressed to Tanya R. Jonker, Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada. E-mail: trjonker@uwaterloo.ca

Lindsay, 2012, 2014). Similarly, people have been found to be consistent in the number of intrusions they make on various recall tests (Lövdén, 2003; Unsworth, 2009; Unsworth & Brewer, 2010). For example, Unsworth and Brewer (2010) used a latent variable analysis on intrusion rates from a number of different recall tasks, some of these being the DRM false memory paradigm, source-monitoring tasks in which participants must identify a context detail of an item (e.g., this item was presented in the upper left quadrant of the screen), and free recall of semantically related or unrelated words. Unsworth and Brewer found significant correlations among their measures of intrusions, and confirmatory factor analyses revealed that intrusion rates from the various tests fit best as one intrusion factor (see also Lövdén, 2003; Unsworth, 2009). In other words, participants who made many intrusions on one recall task also made many intrusions on other recall tasks, and these individual differences were consistent irrespective of the specific features of any individual recall task.

The work by Kantner and Lindsay (2012, 2014) and Unsworth and Brewer (2010) are examples of the recent trend to examine individual differences in incorrect responses and to relate them to various personality and cognitive traits (see also Aminoff et al., 2012). However, despite this recent trend, there has not yet been a thorough examination of the individual differences in incorrect response patterns *across* intrusions and false alarms. The most thorough examinations of the relation between intrusions and false alarms come from two different articles, both involving a false memory paradigm; these studies revealed a weak relation between false alarms on a recognition test and the tendency to report false memories using the DRM paradigm (Kantner & Lindsay, 2012; Zhu, Chen, Loftus, Lin, & Dong, 2013). However, these studies used only the DRM paradigm, which is a unique paradigm that involves the study of items that are strongly semantically related with particular emphasis on false memory for the related but nonstudied critical lure. It is not clear whether this result would extend beyond this paradigm to intrusions produced on various types of cued-recall tests. Therefore, the present work uses structural equation modeling to provide a thorough examination of whether intrusions and false alarms are in fact related to each other. Furthermore, this article also presents an examination of the cognitive factors that underlie any variance shared between intrusions and false alarms. In doing so, I demonstrate that incorrect responses (intrusions and false alarms) are not nuisance data, but instead (a) that they are highly consistent response patterns spanning different types of tests (e.g., recognition, cued recall, and free recall) and stimuli (e.g., pictures, words, and word pairs), (b) that they are not simply an artifact of memory ability, and (c) that the consistent response pattern is the result of one's ability to correctly identify the source of a memory.

Experiment 1

The purpose of Experiment 1 was to explore the relation between intrusions from cued-recall tests and response measures from recognition tests (i.e., response bias and false alarms) to determine whether there were consistent individual differences in incorrect response styles across various memory tests. In this experiment, participants completed a series of cued-recall and recognition tests. Intrusions, false alarms, and response bias were then examined using confirmatory factor analysis to determine whether response styles were consistent across different tests of the same type (e.g., across three different recognition tests), and structural equation modeling was used to explore the relation between different measures of guessing from the various memory tests. I hypothesized that there would be consistent individual differences across both intrusions and false alarms and that intrusions might be related to a more liberal response bias.

In this experiment, I also examined whether memory ability contributed to individual differences in incorrect responses. It is possible that consistent patterns in incorrect responding are simply a by-product of memory ability. This possibility arises from two findings. First, Unsworth (2009) has demonstrated that memory ability is a negative predictor of intrusions, establishing a relation on recall tests. Second, extensive research on the mirror effect in recognition demonstrates that strengthening memory (and thereby increasing hit rate) decreases the false alarm rate (e.g., Cary & Reder, 2003; Glanzer & Adams, 1990). In other words, false alarm rate moves as a function of memory strength. Thus, individual differences in intrusions or false alarm rates might exist only as a by-product of individual differences in memory ability, and if this were the case, they would not be very interesting. To address this possibility, memory ability was used as a predictor before assessing the relation between measures of incorrect responding. If false alarms and intrusions are simply a by-product of memory ability, then controlling for memory ability should yield no significant relation between the various measures of incorrect responding because all meaningful variance would be explained by memory ability. Alternatively, it is possible that incorrect response measures are affected by cognitive traits other than memory ability and, therefore, there should be a significant relation between measures of incorrect responding even after controlling for memory ability. This finding would lay the groundwork for examining cognitive traits that might underlie consistent incorrect response patterns across various memory tests.

Method

Participants. There were 156 students from the University of Waterloo who participated in exchange for partial course credit. All had normal or corrected-to-normal vision and reported fluency in written and spoken English.

Materials and procedure. Participants completed six blocks of tasks. In each block, participants completed a study phase, a distractor phase, and a test phase. The unique characteristics of each block are described below; however, there were some commonalities across all blocks. First, in all blocks, three stimuli were included at the beginning of the study phase to address the primacy effect; these items were not included in the test in any of the blocks and are not included in the descriptions of the materials below. Second, the distractor task was the same for each block: Participants were presented with a unique three-digit number and were asked to count backward by 7 for 30 s. This distractor task was used to prevent rehearsal between study and test.

Each block had a unique stimulus set and the test type was either cued recall or recognition. To create the word lists for some of the blocks, a large set of common nouns was selected from the MRC psycholinguistic database (Coltheart, 1981). All of these words had word frequency scores lower than 500.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Block 1. Forty pairs of unrelated words were constructed from the set of common nouns. The 40 cues were randomly selected, and the 40 targets were randomly selected with the constraint that they began with one of the seven following letters: a, b, c, h, m, p, or s. This constraint was included because participants would be tested using one-letter word stems (e.g., *thumb—m*). By restricting targets to a smaller set of first letters, the distinctiveness of the one-letter cue was reduced, and participants would have more opportunities to produce intrusions on the test. One set of cue-target pairs was constructed and was presented to all participants for study. Each pair was presented for 3,000 ms with a 250 ms interstimulus interval. For each test trial, participants were presented with the cue word and a one-letter word stem (e.g., *thumb—m*) and were given up to 10 s to complete the word stem with a studied word.

Block 2. Fifty-six words were selected from the set of common nouns. Each item was presented for 1,500 ms with a 250 ms interstimulus interval. Memory for the items was tested using a simple recognition task. In this test, individual items were presented and participants were to indicate whether each item was old (i.e., studied) or new. The test was comprised of the 56 studied words and 56 new words, which were intermixed.

Block 3. Category-exemplar pairs were selected from Battig and Montague's (1969) category norms. Eight exemplars from each of six categories were selected. Exemplars were selected with one important criterion: There must be another exemplar of higher frequency in that category that begins with the same letter. For example, *FRUIT—apricot* was selected because the exemplar *apple* is of higher frequency. This was done to reduce the likelihood that participants would guess the correct item; in the above example, if a participant could not remember the studied item and were to guess, he would most likely guess *apple*, which would be properly identified as an intrusion rather than mistakenly classified as a correctly recalled item. Participants first studied these category-exemplar word pairs individually for 2,000 ms each with a 250 ms interstimulus interval. After the distractor phase, participants were shown category cues along with one-letter word stems (e.g., *FRUIT—a*) and were given up to 10 s to complete the word stem with a studied exemplar. Testing was blocked, such that all exemplars for one category were tested before progressing to the next category.

Block 4. Fifty-six images of classic art pieces were presented for 1,500 ms each with a 250 ms interstimulus interval. During a recognition test, individual images were presented and participants were to indicate whether the image was old or new. The test was comprised of the 56 studied images and 56 new images, which were intermixed.

Block 5. The study session for Block 5 was identical to that of Block 3. The materials were also categorized, but they were from six new categories such that there was no overlap with Block 3. The items were selected from positions 6 to 40 from Battig and Montague's (1969) category norms. There was no additional requirement that there be a stronger exemplar that starts with the same letter (as in Block 3); this requirement was not necessary because the test phase involved category-cued free recall, which did not include one-letter word stems. Instead, participants were simply presented with a category label (e.g., *TRANSPORTATION*) and were to type in studied exemplars from that category as they came to mind. Thus, when selecting stimuli, the five strongest exemplars from each category were avoided because these would be the most likely candidates if participants were to guess. For this test, time was not restricted; participants were instructed to type "finish" when they could not remember any more exemplars from that category, after which the next category label appeared on the screen.

Block 6. Fifty-six words were selected from a larger set of common nouns. Each was presented for 1,500 ms with a 250 ms interstimulus interval. For the test, Block 6 used a recollection/familiarity recognition test, rather than a standard old/new recognition test. Before the test, participants were given extensive instruction on the difference between "re-experiencing" a memory and feeling that an item is "familiar." Emphasis was placed on remembering specific detail about the studied word (e.g., remembering what one thought when she saw the word); a memory without specific detail was to be labeled familiar, whereas a memory with specific contextual detail was to be labeled "re-experienced." Items for which the participant had no memory were to be labeled new. The test was comprised of the 56 studied words and 56 new words, which were intermixed.

For all memory tests, no instructions were given regarding guessing or incorrect responding. This was done intentionally to allow participants to complete the tests guided by their individual response styles. If participants asked whether they could guess, which happened on only a couple of occasions, the research assistant responded by telling the participants to respond in the way that they thought best.

Because this is an individual differences study, one random sequence of study and test stimuli was created for each block and this sequence was presented to all participants. This was done to reduce any possible noise produced by different sequences of items. All study and distractor words were carefully reviewed to ensure that words were not repeated across blocks.

Results

Data from one participant were removed from analyses because Block 6 was not completed correctly (no new responses were given), resulting in a final sample size of 155 participants.

The measure of response bias used in the present work is *c* (criterion). Kantner and Lindsay (2012) found that different measures of response bias (e.g., c_a, *B"*) had similar correlations across their various tasks (see their Table 2). Therefore, because of its wide-spread use, *c* was selected to quantify response bias. A positive *c* value indicates conservative responding, with a tendency to respond "new" more frequently than "old".

Descriptive statistics for and correlations between all measures are displayed in Tables 1 and 2, respectively. When computing *c*, cases where participants achieved perfect performance on old or new items were computed by adjusting the score by .01, such that 1.00 proportion hits was converted to .99 and .00 proportion false alarms was converted to .01. Rather than analyzing raw intrusions scores, proportion scores were used. This was done to account for the fact that participants with good memory will correctly complete more of the word stems, leaving fewer opportunities for intrusions or guesses. Therefore, the number of intrusions was assessed relative to the number of incorrect stems (i.e., for Block 1: # intrusions/[40 – # correctly recalled]; for Block 3: # intrusions/

Table 1
Experiment 1: Descriptive Statistics for Hits, Intrusions, False Alarms, and c From Each Block

Variable	Mean	*SD*	Skew	Kurtosis	Reliability[b]
B1 hit	9.81	7.33	.90	.25	.72
B3 hit	18.65	5.81	.23	.48	.54
B5 hit	16.69	6.02	.05	.43	.83
B2 hit	.66	.19	−.43	−.28	.85
B4 hit	.56	.17	−.13	−.12	.75
B6 hit	.64	.21	−.39	−.61	.85
-B6 recollection	.34	.23	.48	−.60	.88
B1 intrusions	.46	.35	.93	1.97	.87
B3 intrusions	.48	.23	.01	−.79	.76
B5 intrusions	4.25[a]	4.22	1.90	5.01	.76
B2 FA	.21	.19	1.71	3.02	.85
B4 FA	.24	.12	.34	−.41	.72
B6 FA	.26	.16	.80	.88	.81
B2 *c*	.26	.42	−.41	.25	.74
B4 *c*	.29	.38	.30	.77	.73
B6 *c*	.15	.49	−.19	.01	.87

[a] Incorrect responses for Block 5 were not proportionalized [b] Reliability reflects split-half reliability, obtained by randomly splitting the trials for each participant into two groups and producing a Pearson correlation on these halves across all participants.

[48 – # correctly recalled]); this score reflects participants' tendency to produce an intrusion only in cases where they had the opportunity to do so. Intrusions for Block 5 were not proportionalized because the test was free recall rather than cued recall, so participants could make as many responses (importantly, intrusions) as they wanted.

Confirmatory factor analysis. Confirmatory factor analysis was used to assess the hypothesized structure of the data. The Intrusions factor was formed by having the intrusion measures from Blocks 1, 3, and 5 load together on the same factor. The FA factor was formed by having the false alarm rates from Blocks 2, 4, and 6 load together on the same factor. The Response Bias factor was formed by having *c* scores from Blocks 2, 4, and 6 load together on the same factor. And, finally, the Memory factor was formed by having the proportions of items correctly recalled from Blocks 1, 3, and 5 and the proportion of recollection ("re-experienced") responses from Block 6 load on the same factor. The Memory factor did not include hit rates from Blocks 2, 4, and 6 because *c* is calculated from both hit and false alarm rates. If the Memory factor included hit rates from the recognition tasks, it would not be independent from the Response Bias factor and, therefore, could not be estimated in the same model. For this same reason, FA and Response Bias could not be estimated in the same model (as a demonstration of the lack of independence, see the strong correlations between response bias, false alarms, and hits for Blocks 2, 4, and 6 in Table 2). Therefore, two separate confirmatory factor analyses were run to address the fact that FA and Response Bias were not independent. The first included the Memory, Intrusions, and FA factors, and the second included the Memory, Intrusions, and Response Bias factors. In each analysis, the latent factors were allowed to correlate with each other. Furthermore, error variables from the same task were allowed to covary (e.g., the error for intrusions from Block 1 was allowed to covary with the error for memory from Block 1) to address shared method variance, and the error variables for intrusions and for hits from Blocks 1 and 3 were allowed to covary to account for the similarity in the testing method for Blocks 1 and 3 (cued + stem recall) relative to Block 5 (category-cued free recall).

Model fits were assessed using a combination of several fit indices: χ^2, root-mean-square error of approximation (RMSEA), and the comparative fit index (CFI). A nonsignificant χ^2 statistic is typically desirable; however, large samples often yield significant values because χ^2 is sensitive to small deviations between the observed and the reproduced covariance matrices. Therefore, fit can be assessed using a ratio between χ^2 and *df*. Acceptable fit is indicated by a χ^2 ratio of two or less, a RMSEA value of less than .08, and a CFI value of greater than .90.

Table 2
Experiment 1: Summary of Correlations for Hits, Intrusions, False Alarms, and c From Each Block

Variable	3H	5H	2H	4H	6H	6R	1I	3I	5I	2FA	4FA	6FA	2*c*	4*c*	6*c*
Block 1 hit	**.45****	**.40****	*.32***	*.24**	*.34***	*.33***	−.35**	−.28**	−.31**	−.25**	−.26**	−.18*	−.02	.00	−.13
Block 3 hit		**.55****	*.38***	*.34***	*.46***	*.40***	−.17*	−.16*	−.29**	−.10	−.09	.00	−.18*	−.19*	−.31**
Block 5 hit			*.30***	*.28***	*.42***	*.42***	−.28**	−.29**	−.36**	−.19*	−.15	−.18*	−.02	−.09	−.16
Block 2 hit				**.38****	**.56****	**.54****	−.20	−.15	−.06	.15	.07	.02	−.56**	−.28**	−.37**
Block 4 hit					**.60****	**.41****	.00	−.08	−.10	.17*	.23**	.18*	−.41**	−.84**	−.48**
Block 6 hit						**.79****	−.19*	−.16	−.11	.05	.06	.30**	−.42**	−.43**	−.82**
-Block 6 recollection							−.13	−.11	−.08	−.13	.03	.15	−.27**	−.30**	−.63**
Block 1 intrusion								**.75****	**.46****	*.20**	*.30***	*.20**	−.07	−.19*	.02
Block 3 intrusion									**.51****	*.19**	*.34***	*.26***	−.09	−.16	−.04
Block 5 intrusion										*.20**	*.26***	*.31***	−.16*	−.08	−.10
Block 2 FA											**.33****	**.40****	−.69**	−.31**	−.24**
Block 4 FA												**.40****	−.35*	−.74**	−.26**
Block 6 FA													−.39**	−.36**	−.75**
Block 2 *c*														**.47****	**.49****
Block 4 *c*															**.48****
Block 6 *c*															

Note. Correlations for the same type of measure are bolded (e.g., correlations between all measures of false alarms). Correlations for similar types of measures are italicized (e.g., correlations between false alarm and intrusion measures).
* Significant at $p < .05$. ** Significant at $p < .01$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
Experiment 1: Standardized Regression Weights and Interfactor Correlations for Memory, Intrusions, and False Alarms

Measure	Latent variable: Memory	Intrusions	FA
B1 hits	.55***		
B3 hits	.68***		
B5 hits	.79***		
B6 recollection	.52***		
B1 intrusions		.61***	
B3 intrusions		.68***	
B5 intrusions		.74***	
B2 FA			.58***
B4 FA			.59***
B6 FA			.65***
	Interfactor correlations		
Intrusions	−.53**		
FA	−.32*	.56***	

* Significant at $p < .05$. ** Significant at $p < .01$. *** Significant at $p < .001$.

This and all subsequent models were run using Amos 22 for SPSS (Chicago, IL). Results from each confirmatory factor analysis are reported in Tables 3 and 4.

Each measure loaded significantly on its factor and the fit of both models was good (model with FA: $\chi^2(26) = 34.53$, $p = .12$, $\chi^2/df = 1.33$, RMSEA = .05, CFI = .98; model with Response Bias: $\chi^2(26) = 48.64$, $p = .005$, $\chi^2/df = 1.87$, RMSEA = .08, CFI = .96), indicating that the measures loaded onto the factors as expected. As predicted, the interfactor correlations revealed that FA and Intrusions were related. Furthermore, Memory correlated negatively with both FA and Intrusions, replicating the finding that those with superior memory tend to make fewer false alarms and fewer intrusions (e.g., Unsworth, 2009).

An interesting note: Memory correlated negatively with Response Bias, indicating that those with superior memory tended to respond more liberally on recognition tests. Given that the Memory factor was formed using only recall and recollection responses, the relation between Memory and Response Bias could be interpreted in the following way: Individuals with a more liberal response bias produce more correct items on a recall test because they use a strategy that involves guessing, which sometimes results in a correct item. However, it seems unlikely that pure guessing in the absence of any true memory would frequently result in more correct cued-recall performance, particularly when the recall cue is an unrelated word (as in Block 1) or when the exemplar is not the strongest associate of the test cue (as in Blocks 3 and 5). A more likely interpretation is that those with a more liberal response bias require less memory strength to report an item, whereas those with a more conservative response bias required greater certainty before reporting an item and will, therefore, forgo reporting recalled items for which their memory is weak, even though the memory is correct. This explanation is consistent with the logic of signal detection theory: Response bias represents one's threshold of memory strength for endorsement, such that a conservative responder requires more memory strength before endorsing an item—studied or nonstudied—as old. Applying this to a cued-recall task, a conservative responder might internally generate a target, but fail to report it because it does not surpass her threshold of necessary memory strength for reporting. Thus, the conservative responder will report fewer correct items, not because she has poor memory, but because she reports only the strongest of items.

When considering the interfactor correlations in Table 4, it is surprising that Response Bias was not related to Intrusions. Specifically, Response Bias and Intrusions were hypothesized to be negatively related (i.e., conservative responders make fewer intrusions), but this relation was not found to be significant. However, the negative relation between Memory and Response Bias suggests that there might be suppression in the relation between Response Bias and Intrusions. That is, those with the highest Memory scores tended to be more liberal and to make fewer intrusions, but those who made fewer intrusions were hypothesized to be conservative responders. Thus, the relation between Response Bias and Intrusions could be influenced both by the shared relation with Memory and the underlying negative relation between Response Bias and Intrusions. In this case, the positivity from the shared relation with Memory might have suppressed the underlying negative relation. Therefore, in the subsequent section, structural equation modeling was used to control for the shared influence of Memory on these latent factors.

Structural equation modeling. Of primary interest was whether Intrusions were related to Response Bias. Confirmatory factor analysis yielded a nonsignificant interfactor correlation between Intrusions and Response Bias. However, the pattern of interfactor correlations between Response Bias and Memory and between Memory and Intrusions suggests that the interfactor correlation between Response Bias and Intrusions might be suppressed. To assess the possibility of suppression, Memory was included as a predictor of Intrusions and of Response Bias, while Intrusions and Response Bias were allowed to correlate. This model allowed for the assessment of the relation between Intrusions and Response Bias, after controlling for the shared influence of Memory. The model is presented in Figure 1A.

Once the shared influence of Memory was controlled, the relation between Intrusion and Response Bias was significant. In other words, this model demonstrated that the relation between Response Bias and

Table 4
Experiment 1: Standardized Regression Weights and Interfactor Correlations for Memory, Intrusions, and Response Bias

Measure	Latent variable: Memory	Intrusions	Response bias
B1 hits	.55***		
B3 hits	.73***		
B5 hits	.75***		
B6 recollection	.49***		
B1 intrusions		.57***	
B3 intrusions		.64***	
B5 intrusions		.79***	
B2 response bias			.68***
B4 response bias			.68***
B6 response bias			.60***
	Interfactor correlations		
Intrusions	−.52**		
Response bias	−.34**	−.17	

** Significant at $p < .01$. *** Significant at $p < .001$.

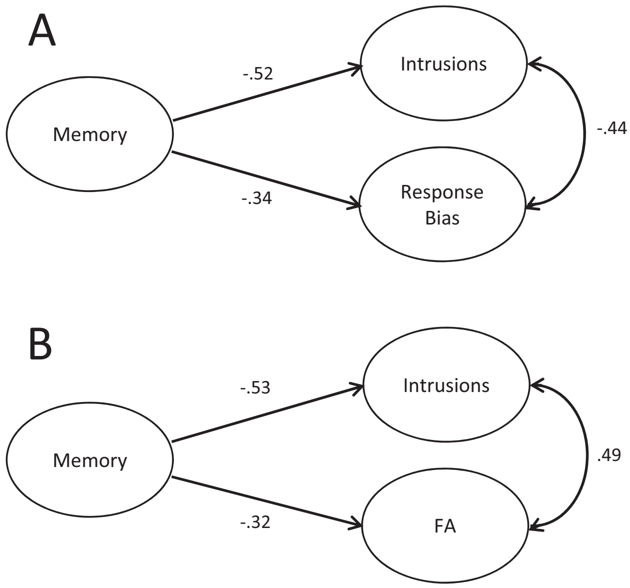

Figure 1. (A) Structural equation model assessing the relation between Intrusions and Response Bias after controlling for Memory. (B) Structural equation model assessing the relation between Intrusions and FA after controlling for Memory. The values displayed are the standardized estimates. All paths are significant at $p < .05$.

Intrusions was suppressed by the shared influence of Memory, and that controlling for this suppression revealed an underlying relation. The fit of the model was good, $\chi^2(26) = 48.64$, $p = .005$, $\chi^2/df = 1.87$, RMSEA = .08, CFI = .96.

A similar analysis was performed on the factors Memory, FA, and Intrusions. This second structural equation model was developed to determine whether the influence of Memory fully explains the relation between FA and Intrusions. It has been demonstrated that memory ability is a negative predictor of intrusions (Unsworth, 2009). Given that Memory was found to negatively correlate with FA, it is possible that FA and Intrusions were related because they were both affected by Memory, with stronger memory ability resulting in fewer intrusions and fewer false alarms. If this were the case, then the relation between measures of guessing from different memory tests would be nothing more than a by-product of memory ability. Alternatively, if a relation was found after controlling for Memory, then this would demonstrate that consistent individual differences in incorrect responding are not simply a by-produce of memory ability but instead reflect an underlying cognitive trait. This model is displayed in Figure 1B.

The model revealed that individual differences in incorrect response patterns were not simply a by-product of memory ability. Instead, the relation between FA and Intrusions remained significant even when controlling for the shared influence of Memory, suggesting that something other than memory ability underlies consistent incorrect responding on different memory tests. The fit of the model was good, $\chi^2(26) = 34.53$, $p = .012$, $\chi^2/df = 1.32$, RMSEA = .05, CFI = .98.

Discussion

This study examined the relation between different measures of guessing on cued-recall and recognition tests. Structural equation modeling revealed that FA was related to Intrusions on cued-recall tasks, and this relation was significant even after controlling for the shared influence of Memory. Similarly, Response Bias was found to be related to Intrusions, but the relation was suppressed by a shared influence from Memory and was only observed when controlling for the shared influence of Memory. The demonstration of this suppression effect is extremely important for the current status of individual differences research on incorrect responding. Particularly, observations of no relation between different measures (e.g., intrusions and response bias) could be because of polarized relations between memory and the measures of interest. Therefore, this experiment demonstrates the importance of including an independent measure of memory ability as a control in studies examining the relation between response bias and measures of incorrect responding.

The key finding from the present work was a robust relation between measures of incorrect responding, and one that is not explained by memory ability. This finding demonstrates that a cognitive trait other than memory ability drives incorrect responding on various memory tests. In Experiment 2, I explored a cognitive trait that might underlie incorrect response tendencies across various memory tests.

Experiment 2

In Experiment 1, the rate at which participants made incorrect responses was consistent, even across different test characteristics and stimuli. That is, incorrect responding on cued-recall tests was strongly related to false alarm rate on recognition tests, and this pattern remained even when controlling for the shared influence of memory ability. This finding suggests that there is a cognitive trait that is separable from pure memory ability that underlies incorrect responding, and that this cognitive trait affects performance on memory tests similarly even when test demands or stimulus features differ.

The purpose of Experiment 2 was to explore a cognitive trait hypothesized to underlie the variance shared between incorrect responding on cued-recall and on recognition tests. This trait involves correctly attributing the feeling of familiarity of an item (Whittlesea & Williams, 2000) to its proper source. For example, if upon seeing *FRUIT—a*, "*apple*" comes to mind, can one accurately determine whether the source of *apple* was a study list (i.e., episodic memory) versus existing knowledge of the strong association between *fruit* and *apple* (semantic memory)? The ability to do so was predicted to underlie the variance shared between intrusions and false alarms. That is, individuals who fail to discriminate between episodic memories and familiarity arising from other sources, such as semantic generation or similarity to a study stimulus, will on average produce more incorrect responses on cued-recall tests. A similar process was thought to occur during recognition: When individuals encounter a lure that feels familiar, if they are unable to easily identify the derivation of that familiarity, such as similarity or recent exposure in an extraexperimental context, then they might be more likely to incorrectly endorse that item. This type of incorrect responding on both recall and recognition tests could occur because participants mistakenly attribute some of the familiarity to episodic memory rather than to other sources, or because they are unsure of the origin of possible responses and decide to endorse any familiar memories. Thus, I

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

predicted that individual differences in the ability to discriminate episodic retrieval from other types of "memories" would underlie the relation between intrusions and false alarms. In the present article, I refer to this ability as retrieval discrimination.

To measure retrieval discrimination, a novel test was introduced: After completing the entire cued-recall task, participants were given a surprise response-confidence test during which they were again presented with each of the cues from the test (*FRUIT—a*) along with the response that they had provided (*apple*). They were asked to indicate for each response their confidence in whether that item had been presented during study or was a guess. This test allowed me to assess participants' accuracy at identifying whether their responses reflected true episodic memory versus other forms of memory (i.e., a "guess"). A participant with poor retrieval discrimination would have low accuracy on this response confidence task, frequently indicating that incorrect responses reflected studied items and that correct responses reflected guesses, whereas a participant with strong retrieval discrimination would have high accuracy, correctly indicating that correct responses reflected studied items and that incorrect responses reflected guesses. High versus low confidence did not factor into this measure—only accuracy of classification mattered.

I hypothesized that individuals with poor retrieval discrimination would produce more intrusions and false alarms because they would not be able to accurately attribute the feeling of familiarity during test to semantic activation rather than to true episodic memory for studied items. The result of this would be more intrusions on a cued-recall test and more false alarms on a recognition test. More important, in line with this logic, I predicted that retrieval discrimination would explain a significant portion of the *variance shared* between these two types of incorrect response patterns on memory tests, positioning retrieval discrimination as a key cognitive trait underlying incorrect responding across a variety of memory tests.

The surprise response-confidence task allowed for the measurement of another trait in addition to retrieval discrimination, which provided discriminant validity for the retrieval discrimination trait. This second trait was hypothesized to be a significant predictor of intrusions only, and was not predicted to explain variance shared between intrusions and false alarms. It reflects an individual's tendency to report a guess based on semantic memory, even while being confident that the response was not part of the study set (e.g., providing *apple* for the cue *FRUIT—a* even while being certain that *apple* was not a studied item; a precise description of how intentional guessing was quantified is described in the Method section below). Here, this trait is referred to as intentional guessing and might reflect a maximizing strategy; it is not a failure of the memory system because—despite reporting it—the individual is quite certain that the guess was not part of the study set.

Intentional guessing was a suspected predictor of cued-recall intrusions, but not of false alarms. If intentional guessing were to affect recognition responses, then the participant would likely respond with "old" to every item, resulting in very high (or perfect) hits as well as very high false alarms. Although participants differ in their tendency toward more liberal or more conservative responding, extreme liberal responding would be an odd strategy for a participant to adopt on a recognition test and would result in unusable data.[1] Therefore, intentional guessing was not hypothesized to be a predictor of false alarms. Thus, retrieval discrimination was predicted to explain significant variance shared between intrusions and false alarms, but intentional guessing was not. This expected pattern would provide a demonstration that not all traits explain shared variance; instead, it would support the key role of retrieval discrimination and provide discriminant validity for the construct.

Method

Participants. There were 103 students from the University of Waterloo who participated in exchange for partial course credit. All had normal or corrected-to-normal vision and reported fluency in written and spoken English.

Materials and procedure. Participants first completed a recognition block, which was followed by a cued-recall block. As in Experiment 1, the same sequence of stimuli was presented to all participants for all study and test sessions.

Recognition block. The recognition procedure was similar to that of Block 2 from Experiment 1. Forty words were selected from the set of common nouns. Each item was presented for 2,000 ms with a 250 ms interstimulus interval. Three additional items began the study phase to address the primacy effect, but memory for these three items was not tested. After the study session, participants completed the distractor task used in the previous experiments. The task lasted 2 min and during that time participants were presented with a new three-digit number every 30 s and were to count backward by 7 s until the number disappeared. Memory for the study items was tested using a simple recognition task. In this test, individual items were presented and participants were to indicate whether each item was old (i.e., studied) or new. The test was comprised of 40 studied words and 40 new words, which were intermixed. For this test and the cued-recall test, no instructions were given regarding guessing or incorrect responding.

Cued-recall block. The cued-recall procedure was similar to that of Block 3 from Experiment 1. Participants first studied 48 category-exemplar word pairs individually for 2,000 ms each with a 250 ms interstimulus interval. Three additional pairs were included at the beginning of the study phase to address the primacy effect; these were not included in the test. After the study, participants were presented with a three-digit number and were given 30 s to count backward by 7 s. During a test, participants were then shown category cues along with one-letter word stems (e.g., *FRUIT—a*) and were given up to 10 s to complete the word stem with a studied exemplar. Eighteen new cues, which could not be completed by any of the study items, were included in this test (e.g., *FRUIT—o*, because there were no studied fruit items that began with "o"). These were included so that intrusions could be measured independently of memory ability. In Experiment 1, intrusions were measured as a proportion of the number of cues that were not correctly completed, and the total number of cues not correctly completed differed across participants depending on how many study items they remembered. Even though a proportion

[1] It could be the case that intentional guessing correlates with response bias, such that those who respond more liberally on recognition tests would also demonstrate higher intentional guessing. If this were the case, then it would suggest that the same strategic approach underlies response bias setting as well as intentional guessing. This possibility was also assessed in Experiment 2, but was not supported by the model.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

score was used to address the problem of the effect of memory ability on the number of trials on which a participant can guess, there is still the problem of some dependency in the measures. The inclusion of new cues in Experiment 2 allowed for the determination whether the proportion score effectively controlled for the effect of memory ability: If the proportion score did effectively control for memory ability, then in Experiment 2, the intrusion rate for old cues (using the proportion score) should be highly correlated with intrusion rate for new cues (that they were; see Table 6).

One random order of test trials was created, and the test was not blocked by category (as it was in Experiment 1), but was fully randomized across categories. After the cued-recall test was completed, participants performed a surprise response-confidence task. In this task, participants were shown each cue again along with their response to that cue in quotation marks and they were to rate their confidence in the accuracy of the response that they had given. Participants were shown four response options: (a) "I am certain that this response is incorrect," (b) "I think that this response is incorrect," (c) "I think that this response is correct," and (d) "I am certain that this response is correct." For example, a participant would see the cue *FRUIT—a* along with her cued-recall response, which might have been *apple.* In this case, she might then respond with "2" to indicate that she thinks that the response she gave on the test was not part of the study list but that she is not certain (if she was certain that it was not part of the study set, then she should select "1"). Participants were told that they did not have to use the whole response scale; instead, they were encouraged to select the most accurate option for each cued-recall response. They did not make confidence responses for trials on which they had made no response. More important, the confidence test was administered after the cued-recall test to prevent it from having an effect on participants' response strategy during the cued-recall test. If the response-confidence judgments were made after each cued-recall trial, it might lead participants to deviate from their natural response styles on the cued-recall test. Therefore, participants were not aware that they would be completing the response-confidence task until the cued-recall test was complete.

The confidence test allowed for the computation of two unique variables. The first was a retrieval discrimination score, which reflected participants' ability to accurately distinguish true recall from nonstudied responses and was a proportion of the number of items correctly identified as their type (studied, guess) out of the total number of responses made on the cued-recall test. In other words, if the individual produced 10 intrusions and 10 correct responses to the studied cues on the cued-recall test, and then correctly identified 7 of the intrusions as nonstudied items (i.e., response options 1 and 2) and 8 of the studied responses as correct (i.e., response options 3 and 4), then that participant would have a retrieval discrimination score of .75 (15/20), indicating that he could correctly classify the identify of 75% of his responses from the cued-recall test.

The second variable was a score that reflected an individual's tendency to produce intrusions on the cued-recall test even while being very confident that that item did not occur during study, here referred to as intentional guessing. This was computed as the proportion of intrusions for which participants gave a "1" response on the confidence test relative to all cues that were not correctly completed (i.e., # intrusions with "1" response/[48 + 18 – # correctly recalled]). For example, if a participant had given the response "orange" to the cue *FRUIT—o*, which was not a studied item (there was in fact no studied "o" exemplar in the *FRUIT* category), and then during the confidence test she responded with "1" ("I am certain that this response is incorrect"), this would increase the intentional guessing score. If she had correctly recalled 16 items, and gave a "1" for only one of the intrusions, then she would have an intentional guessing score of .02 (1/[48 + 18 − 16]), indicating that she produced an intentional guess 2% of the time when she had not correctly recalled anything.

Results

One participant performed at chance on the recognition test and was, therefore, excluded from all subsequent analyses. The final dataset included 102 participants. However, one participant did not have a criterion measure because performance on the recognition memory test was perfect (100% hits, 0% false alarms). Analyses included this participant only when a model did not require data from the missing cell.

Two measures of intrusions could be obtained from the cued-recall test: one from the studied cues (# old intrusions/[48 – # correctly recalled]) and one from the new cues (# new intrusions/18). Descriptive statistics for and correlations between all measures are displayed in Tables 5 and 6, respectively.[2]

Structural equation modeling. To first assess whether the general pattern of results found in Experiment 1 were replicated using only one task of each type, two models were created in the likenesses of those from Figure 1. Because this experiment used only one cue-recall and recognition test, measures were used directly in the model in most cases (rather than latent factors): Memory was the number of items correctly recalled on the cued-recall test, FA was the proportion of false alarms from the recognition test, and Response Bias was the criterion value from the recognition test. An Intrusions factor was derived using the proportion of intrusions on new test cues and the proportion of intrusions on studied test cues.

As shown in Figure 2, Intrusions correlated with Response Bias (Panel A) and with FA (Panel B), successfully replicating Experiment 1. That is, Intrusions and Response Bias correlated negatively, indicating that more conservative responders have fewer intrusions on a cued-recall test, and Intrusions and FA correlated positively, indicating that those who endorsed more new items on a recognition test also made more intrusions on a cued-recall test. The predictive paths for Memory were—for the most part—not significant. This is likely because of reduced power from using only one measure for Memory (rather than four, as in Experiment 1).

To extend the findings from Experiment 1, two novel predictors were added to the models. The predictors Retrieval Discrimination and Intentional Guessing were included to determine whether these two traits might explain the variance shared between Intrusions and Response Bias and FA. These models are shown in

[2] Reliability for new intrusions was found to be quite weak. This is likely because of the fact that the split-half method reduced the samples to nine per half. Although reliability was low, this measure correlated strongly with study intrusions (r = .78), and, furthermore, latent variable analysis was used to estimate an Intrusions factor, which would reduce some of the influence of unreliability on the path estimates.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 5
Experiment 2: Descriptive Statistics for Hits, Intrusions, False Alarms, c, Retrieval Discrimination, and Intentional Guessing Variables

Variable	Mean	*SD*	Skew	Kurtosis	Reliability[a]
Hit	16.22	5.97	.09	.27	.61
New intrusions	.41	.18	−.19	−.48	.44
Study intrusions	.45	.20	−.13	−.34	.69
Total intrusions	.43	.18	−.29	−.28	.72
FA	.17	.13	.93	.70	.72
c	.18	.42	.86	1.41	.67
Retrieval discrimination	.73	.16	−.64	.19	.49
Intentional guessing	.08	.10	2.32	7.38	.75

[a] Reliability reflects split-half odd-even reliability, obtained by producing a Pearson correlation on these halves across all participants.

Figure 3. Both models fit the data well (Panel A model, $\chi^2(5) = 4.47$, $p = .35$, $\chi^2/df = 1.12$, RMSEA = .03, CFI = 1.00; Panel B model, $\chi^2(5) = 4.94$, $p = .29$, $\chi^2/df = 1.24$, RMSEA = .05, CFI = 1.00), and, importantly, with the inclusion of the additional predictors, the correlations between incorrect response measures were not significant. Therefore, these models demonstrate that together Retrieval Discrimination, Intentional Guessing, and Memory explained the majority of the variance shared between Intrusions and FA and Intrusions and Response Bias.

When Retrieval Discrimination and Intentional guessing were included in the models as predictors of Intrusions and Response Bias or Intrusions and FA, both were found to predict the outcome variables. However, the pattern of relations differed. Specifically, as seen in Figure 3, Retrieval Discrimination was found to be a significant predictor of Intrusions, Response Bias and FA, whereas Intentional Guessing, on the other hand, was a significant predictor only of Intrusions and had no relation with either of the recognition test measures. This outcome is predicted by signal detection theory, as response bias is thought to reflect the threshold at which signal outweighs noise and leads to an old response, even if the item is a new item. Therefore, signal detection theory predicts that false alarms occur for items that participants believe were studied items; they should not reflect intentional guesses that participants know are not correct, but instead a failure to correctly discriminate episodic memories from familiar but nonepisodic memories. Therefore, intentional guessing should not affect recognition tests, a prediction that was confirmed by the model.

The observation that Retrieval Discrimination was a predictor of Intrusions, FA, and Response Bias suggests that Retrieval Discrimination is likely the key trait underlying individual differences in incorrect responding across various memory tests. Intentional Guessing is an unlikely candidate for explaining variance shared because it was found to be a significant predictor only of Intrusions. To explicitly test whether Retrieval Discrimination explained the variance shared between incorrect response measures, I broke down the models of Figure 3 into separate models: One set of models had Retrieval Discrimination and Memory as predictors, and the other set had Intentional Guessing and Memory as predictors. Each of these models was then compared with a nearly identical model with only one change: Intrusions and Response Bias or Intrusions and FA were not allowed to correlate (see an example of a pair of models in Figure 4). By contrasting the fit of the model with correlated measures to the fit of the model with uncorrelated measures, I could determine whether the predictors explain the majority of the variance shared between Intrusions and Response Bias or Intrusions and FA. In other words, this analysis determined whether a correlation between the measures of incorrect responding significantly improved the fit of the model. If the two models do not differ statistically, then this would demonstrate that the predictors explained all significant variance shared between the measures of incorrect responding, because a correlation was not necessary for good fit.

To compare fits, the χ^2 difference between the two models was computed, and the difference was assessed for significance at $p < .05$ ($df = 1$, because only the correlation parameter changed). The models and their fits are detailed in Table 7.

When Retrieval Discrimination was used as a predictor, it did not matter whether Intrusions and FA or Intrusions and Response Bias were allowed to correlate; the model fits did not differ statistically. This outcome suggests that Retrieval Discrimination underlay individual differences in measures of incorrect responding across cued-recall and recognition tests, supported by the lack of significant correlation in both models (Model 1B: $r = .05$, $p = .61$; Model 2B: $r = -.18$, $p = .10$). That is, individuals' consistency in producing intrusions on a cued-recall test and false alarms or a liberal response bias on a recognition test was due to poor retrieval discrimination ability; after accounting for this trait, there was no significant shared variance left to explain.

Retrieval Discrimination was not the only trait related to individual differences on memory tests, but it was the only trait that explained the variance shared across recall and recognition tests. Intentional Guessing was not found to explain shared variance, which is evident in Figure 4. When contrasting the fit of a model in which Intrusions and FA or Intrusions and Response Bias were uncorrelated with a model in which they were allowed to correlate, fit differed significantly; thus, Intentional Guessing did not explain the variance shared between these measures of incorrect responding and correlation was necessary for good fit in this model (Model 3B: $r = .33$, $p = .003$; Model 4B: $r = -.30$, $p = .006$). However, Intentional Guessing was found to explain unique variance on cued-recall tests, beyond that of Retrieval Discrimination. That is, Intentional Guessing was found to be a discrete trait that influenced Intrusions. The Intentional Guessing measure is likely measuring a tendency toward response completion or maximization. This could be a strategy, but one that is different from simple conservative versus liberal responding and is limited to cued-recall tests only.

Table 6
Experiment 2: Summary of Correlations for Hits, Intrusions, False Alarms, c, Retrieval Discrimination, and Intentional Guessing Variables

Variable	NI	SI	FA	*c*	RD	IG
Cued-recall hit	.02	.13	−.31**	−.11	.56**	−.02
New intrusions		**.78****	.23*	−.25*	−.36**	.38**
Study intrusions			.22*	−.28**	−.24*	.42**
FA				−.61**	−.58**	.00
C					.17	−.04
Retrieval discrimination						.18
Intentional guessing						

Note. The correlation between two measures of intrusions is bolded.
* Significant at $p < .05$. ** Significant at $p < .01$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

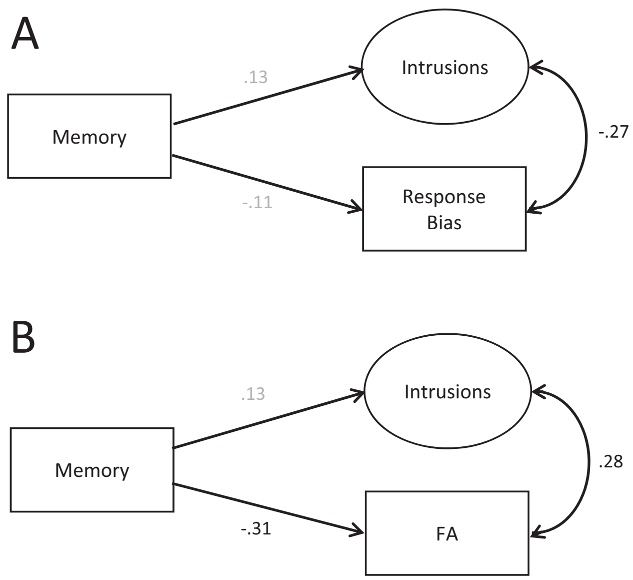

Figure 2. (A) Structural equation model assessing the relation between Intrusions and Response Bias while controlling for Memory. (B) Structural equation model assessing the relation between Intrusions and FA while controlling for Memory. The values displayed are the standardized estimates. Direct measures are displayed as squares, whereas derived (or latent) factors are displayed as ovals. Paths significant at $p < .05$ are represented with black text; nonsignificant paths are represented with grey text.

General Discussion

In two experiments, I examined individual differences in intrusions on cued-recall tests and explored how they related to measures from recognition tests. Experiment 1 revealed that intrusion rate was related to false alarm rate, and this relation was independent of the shared influence of memory ability. Thus, individual differences in incorrect response measures across cued-recall and recognition tests were the result of more than just memory strength. This highlights the importance of incorrect responses: They are not simply nuisance data or a by-product of memory ability, but instead reflect meaningful individual differences in response style.

Experiment 2 revealed the underpinnings of these meaningful individual differences: Specifically, I found that the ability to effectively discriminate true episodic memories from nonstudied responses explained the variance shared between intrusions and false alarms, as well as the variance shared between intrusions and response bias, lending support to the notion that consistent individual differences in incorrect responding on cued-recall and recognition tests are due primarily to one's ability to correctly identify the source of the products of retrieval. Indeed, in Experiment 2, retrieval discrimination was found to explain all significant variance shared between intrusions and false alarms, suggesting that it is the key trait underlying incorrect response patterns across different memory tests.

One limitation of this conclusion, however, is the short time-course of these experiments. All measures were collected within a single session. Therefore, it is possible that relations are because of consistency in the participants' states, rather than true underlying differences that are stable over time. However, research by Kantner and Lindsay (2012, 2014) has revealed strong consistency in individual differences in recognition performance. Considering that false alarm rate, response bias, and intrusions seem to be strongly related, the recognition work by Kantner and Lindsay suggests that similar measures in cued-recall might be equally stable.

When it came to retrieval discrimination and response bias, it was interesting to discover a negative relation. This effect was interesting because one could expect a parabolic relation, with poorer retrieval discrimination as one moves further away from response bias = 0. That is, those with strong retrieval discrimination might be expected to set the most optimized response bias for detecting signal from noise ($c = 0$); those with weak retrieval discrimination, on the other hand, might be expected to set response biases that were either too liberal or too conservative. However, this was not in fact the case in Experiment 2: A quadratic relation did not significantly improve a regression model where Retrieval Discrimination predicted Response Bias, $\Delta R^2 = .02$, $F(1, 98) = 2.16$, $p = .15$, demonstrating that the linear relation provides a superior fit for the data. Thus, the data suggest that those who are poor at monitoring the source of their memories tend toward conservatism, perhaps recognizing that they cannot easily distinguish true memory from semantic activation and thereby proceeding cautiously in reporting.

It is worth noting that the retrieval discrimination variable measured in the present work is conceptually similar to the well-known cognitive trait of *source monitoring*. Source monitoring is the ability to correctly determine the episodic context in which an item was learned (Johnson, Hashtroudi, & Lindsay, 1993; Mitchell & Johnson, 2009). For example, words could be presented in male

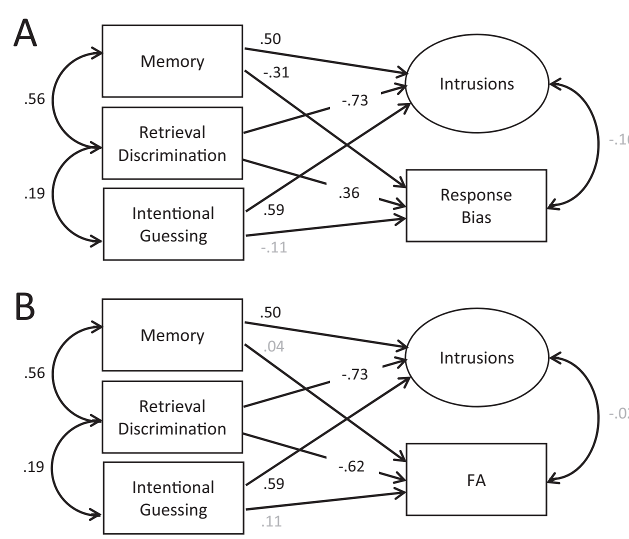

Figure 3. (A) Structural equation model assessing the relation between Intrusions and Response Bias with Memory, Retrieval Discrimination, and Intentional Guessing as predictors. (B) Structural equation model assessing the relation between Intrusions and FA with Memory, Retrieval Discrimination, and Intentional Guessing as predictors. The values displayed are the standardized estimates. Direct measures are displayed as squares, whereas latent factors are displayed as ovals. Retrieval Discrimination and Memory were allowed to correlate to address the effect that Retrieval Discrimination had on recall, and Retrieval Discrimination and Intentional Guessing were allowed to correlate to address variance shared between them. Paths significant at $p < .05$ are represented with black text; nonsignificant paths are represented with grey text.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

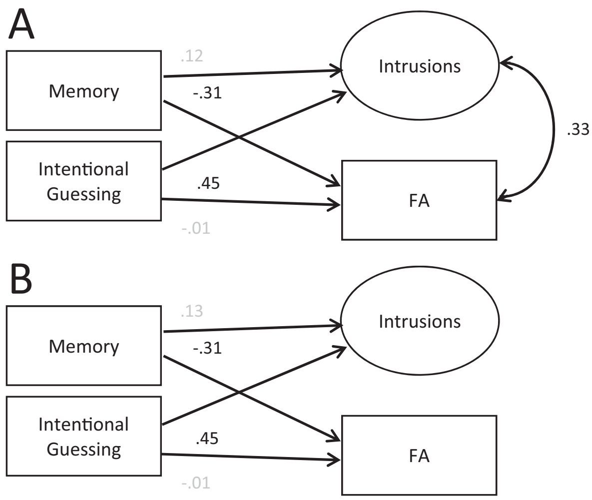

Figure 4. (A) Structural equation model assessing the correlation between Intrusions and False Alarms with Memory and Intentional Guessing as predictors (Model 3B in Table 7). (B) Structural equation model with no correlation between Intrusions and False Alarms and with Memory and Intentional Guessing as predictors (Model 3A in Table 7). The values displayed are the standardized estimates. Direct measures are displayed as squares, whereas derived factors are displayed as ovals. Paths significant at $p < .05$ are represented with black text; nonsignificant paths are represented with grey text.

and female voices, and source monitoring could be measured as the probability of correct identification of whether a word had been spoken by a male or female. Source monitoring can also involve correctly attributing a memory to a particular environmental context in which an item was learned, the format in which an item was originally presented, the list in which the item occurred, or other types of sources. Important to the present work is that source monitoring has been implicated as a key factor in intrusions in recall and false memory (Koriat & Goldsmith, 1996; Qin, Ogle, & Goodman, 2008; Unsworth & Brewer, 2010; Winograd, Peluso, & Glover, 1998). For example, Unsworth and Brewer (2010) found that poor source monitoring was a strong predictor of increased intrusions on a variety of recall tests, such as delayed free recall, the DRM false memory paradigm, and paired associates recall. Koriat and Goldsmith (1996) also demonstrated a link between monitoring ability and intrusions, suggesting that "people cannot readily improve the overall quality of information that they *retrieve*, but they can improve the quality of what they *report*" (p. 507, authors' italics); in other words, monitoring processes cannot increase correct recall, but they can affect decisions about reporting based on the quality of retrieved information. This view of monitoring (see also postretrieval monitoring; e.g., Hayama & Rugg, 2009) shares conceptual overlap with the retrieval discrimination measure in the present study: In both cases, the proposed traits emphasize one's ability to identify the source of memory activation, whether that be one of several environmental conditions or features (source monitoring) or between episodic versus semantic retrieval (retrieval discrimination). Thus, it is possible that retrieval discrimination and source monitoring both rely on the same underlying cognitive mechanism. Future research on this topic would be beneficial in understanding the possible link between source monitoring and what has been referred to here as retrieval discrimination.

A second key finding yielded by the present experiments was the suppression effect observed when examining the relation between intrusions and response bias in Experiment 1. On their own, the response bias and intrusions factors were not found to negatively correlate, but a more complex model demonstrated that this was because of memory ability being negatively related both to intrusions and to response bias. Thus, individuals with the strongest memory ability tended to be less conservative in their response patterns and also to make fewer intrusions, but those who made fewer intrusions were hypothesized to be conservative responders. When the relation with memory ability was controlled using a structural equation model, the suppressed relation between intrusions and response bias was revealed. This demonstrates that memory ability is an important control when researchers assess the overlap between intrusion rate and response bias.

It is worth noting that the present work does not rule out the possible influence of other cognitive and personality traits on incorrect response rates. For example, Aminoff and colleagues (2012) found that a cognitive style of verbalizing and some personality factors (e.g., negative affect, fun-seeking behavior) were related to a participant's tendency to shift response bias as task demands changed, and Winograd and colleagues (1998) found that intrusion and false alarm rates correlated with reports of dissociative experiences. However, these studies examined correlations between cognitive and personality traits with a particular response pattern on a particular type of test. The present work more broadly demonstrates that the cognitive trait of retrieval discrimination underlies the majority of the variance shared across various tests and measures of incorrect responding. Other cognitive and personality traits might account for the small amount of remaining shared variance (although there was no significant shared variance left to explain in Experiment 2; see Figure 3),

Table 7
Experiment 2: Comparison of the χ^2 Results From Models With and Without Correlated Measures of Incorrect Responding

Model	Predictors	Model dependent variables	χ^2 (*df*)	χ^2 difference	*p*
1A	Memory, Retrieval Discrimination	Uncorrelated: Intrusions, FA	4.61 (3)		
1B	Memory, Retrieval Discrimination	Correlated: Intrusions, FA	4.37 (2)	.24	.62
2A	Memory, Retrieval Discrimination	Uncorrelated: Intrusions, Response Bias	6.74 (3)		
2B	Memory, Retrieval Discrimination	Correlated: Intrusions, Response Bias	4.02 (2)	2.72	.10
3A	Memory, Intentional Guessing	Uncorrelated: Intrusions, FA	12.61 (4)		
3B	Memory, Intentional Guessing	Correlated: Intrusions, FA	2.34 (3)	10.27	.001
4A	Memory, Intentional Guessing	Uncorrelated: Intrusions, Response Bias	10.63 (4)		
4B	Memory, Intentional Guessing	Correlated: Intrusions, Response Bias	2.31 (3)	8.32	.004

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

and there are likely other cognitive and personality traits that affect recognition or cued-recall tests independently, as was found with the intentional guessing measure in Experiment 2. However, when it comes to understanding the cognitive trait that affects consistent performance across a variety of memory tests, the present work demonstrates that the ability to discriminate the source of a retrieved memory plays a significant role in producing incorrect responses across a variety of memory tasks.

Conclusion

People respond to various memory tests and stimuli with remarkable consistency, suggesting that certain cognitive and perhaps personality traits guide performance on these tasks. In fact, this does seem to be the case: One's ability to determine the source of a memory affects not only correct responses on memory tests, but also the rate at which one makes incorrect responses across different test types. This demonstrates the ubiquitous impact of monitoring on memory. Indeed, activating a memory is not enough; to accurately remember, one must also be able to discriminate true episodic memory from other types of activation. When able to do so, one improves not only the ability to properly produce a true memory, but also the ability to edit out inappropriate or false memories.

References

Abdi, H. (2007). Signal Detection Theory (SDT). In N. J. Salkind & K. Rasmussen (Eds.), *Encyclopedia of measurement and statistics* (pp. 886–889). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412952644.n405

Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Tipper, C., Johnson, A., . . . Miller, M. B. (2012). Individual differences in shifting decision criterion: A recognition memory study. *Memory & Cognition, 40,* 1016–1030. http://dx.doi.org/10.3758/s13421-012-0204-6

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80,* 1–46. http://dx.doi.org/10.1037/h0027577

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language, 49,* 231–248. http://dx.doi.org/10.1016/S0749-596X(03)00061-5

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

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22. http://dx.doi.org/10.1037/h0046671

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38,* 833–848. http://dx.doi.org/10.3758/MC.38.7.833

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16. http://dx.doi.org/10.1037/0278-7393.16.1.5

Hayama, H. R., & Rugg, M. D. (2009). Right dorsolateral prefrontal cortex is engaged during post-retrieval processing of both episodic and semantic information. *Neuropsychologia, 47,* 2409–2416. http://dx.doi.org/10.1016/j.neuropsychologia.2009.04.010

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28. http://dx.doi.org/10.1037/0033-2909.114.1.3

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition, 40,* 1163–1177. http://dx.doi.org/10.3758/s13421-012-0226-0

Kantner, J., & Lindsay, D. S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review, 21,* 1272–1280. http://dx.doi.org/10.3758/s13423-014-0608-3

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

Lövdén, M. (2003). The episodic memory and inhibition accounts of age-related increases in false memories: A consistency check. *Journal of Memory and Language, 49,* 268–283. http://dx.doi.org/10.1016/S0749-596X(03)00069-X

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135,* 638–677. http://dx.doi.org/10.1037/a0015849

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. http://dx.doi.org/10.1037/a0014309

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. http://dx.doi.org/10.1037/0278-7393.21.4.803

Unsworth, N. (2009). Variation in working memory capacity, fluid intelligence, and episodic recall: A latent variable examination of differences in the dynamics of free recall. *Memory & Cognition, 37,* 837–849. http://dx.doi.org/10.3758/MC.37.6.837

Unsworth, N., & Brewer, G. (2010). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language, 62,* 19–34. http://dx.doi.org/10.1016/j.jml.2009.08.002

Whittlesea, B. W., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26,* 547–565.

Winograd, E., Peluso, J. P., & Glover, T. (1998). Individual Differences in Susceptibility to Memory Illusions. *Applied Cognitive Psychology, 12,* S5–S27. http://dx.doi.org/10.1002/(SICI)1099-0720(199812)12:7<S5::AID-ACP553>3.0.CO;2-D

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152–176. http://dx.doi.org/10.1037/0033-295X.114.1.152

Zhu, B., Chen, C., Loftus, E. F., Lin, C., & Dong, Q. (2013). The relationship between DRM and misinformation false memories. *Memory & Cognition, 41,* 832–838. http://dx.doi.org/10.3758/s13421-013-0300-2

Received July 7, 2015
Revision received September 28, 2015
Accepted October 19, 2015 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.