Published in final edited form as:
Q J Exp Psychol (Hove). 2011 August ; 64(8): 1515–1542. doi:10.1080/17470218.2011.560272.

False recall in the Deese–Roediger–McDermott paradigm: The roles of gist and associative strength

David R. Cann[1], **Ken McRae**[2], and **Albert N. Katz**[2]

[1]Department of Psychology, Mount Royal University, Calgary, AB, Canada

[2]Department of Psychology, The University of Western Ontario, London, ON, Canada

Abstract

Theories of false memories, particularly in the Deese–Roediger–McDermott (DRM) paradigm, focus on word association strength and gist. Backward associative strength (BAS) is a strong predictor of false recall in this paradigm. However, other than being defined as a measure of association between studied list words and falsely recalled nonpresented critical words, there is little understanding of this variable. In Experiment 1, we used a knowledge-type taxonomy to classify the semantic relations in DRM stimuli. These knowledge types predicted false-recall probability, as well as BAS itself, with the most important being situation features, synonyms, and taxonomic relations. In three subsequent experiments, we demonstrated that lists composed solely of situation features can elicit a gist and produce false memories, particularly when monitoring processes are made more difficult. Our results identify the semantic factors that underlie BAS and suggest how considering semantic relations leads to a better understanding of gist formation.

Keywords

False memories; Gist; Situation knowledge; Fuzzy-trace theory; Activation/monitoring theory

Roediger and McDermott's (1995) technique, originally used by Deese (1959b), has resulted in robust rates of false recall and recognition across numerous studies. Commonly known as the Deese–Roediger–McDermott (DRM) paradigm, participants are presented with word lists composed of the strongest associates of a critical non-presented word, as determined by word association norms. On subsequent recall and recognition tests, participants tend to intrude the nonpresented critical word as having been studied previously.

Much of the research using this paradigm (see Roediger, McDermott, & Robinson, 1998, for a review) has focused on manipulating variables to reduce false-memory rates, and, for these purposes, data generally are collapsed across multiple DRM lists. However, a theoretically important and somewhat neglected issue concerns the underlying characteristics of the word lists that produce different false-memory rates (but see Brainerd, Yang, Reyna, Howe, & Mills, 2008). Crucially, not all lists are equally effective in eliciting false memories despite being constructed in the same manner. Moreover, the two main theoretical explanations have opposing strengths and weaknesses, and neither by themselves provides a comprehensive explanation for the observed differences in list effectiveness.

Correspondence should be addressed to David R. Cann, Department of Psychology, Mount Royal University, Calgary, Alberta, Canada, T3E 6K6. dcann@mtroyal.ca.

One approach, based on the strength of normative association from list words to the nonpresented critical word (backward associative strength, or BAS), provides a clear operational definition of a variable known to be effective in producing false memories (Roediger, Watson, McDermott, & Gallo, 2001). However, it is theoretically underdetermined. The other major approach, based on fuzzy-trace theory (Brainerd & Reyna, 2002), provides a strong theoretical perspective based on gist formation, but it has been argued that it is less informative regarding the effectiveness of gist in producing false memories (Roediger, Watson, et al., 2001). One main goal of the current study is to broach this divide by examining the underlying characteristics of the DRM lists using a feature-based coding scheme developed by Wu and Barsalou (2009). The experiments reported herein deconstruct BAS into semantic relations. These relations (particularly situation features) were employed as an index of gist formation, and the relative contributions of BAS and situation features were examined in DRM experiments.

Not all DRM lists are created equal

The 55 commonly used DRM lists were constructed in a similar manner and are presented in Roediger, Watson, et al. (2001). Each list is composed of the top 15 associates of the nonpresented critical word taken from word association norms (Nelson, McEvoy, & Schreiber, 1999; Russell & Jenkins, 1954). Words are presented to participants in descending order of their associative strength to the critical word. Consider the lists for *doctor* (*nurse, sick, lawyer, medicine, health, hospital, dentist, physician, ill, patient, office, stethoscope, surgeon, clinic, cure*) and *king* (*queen, England, crown, prince, George, dictator, palace, throne, chess, rule, subjects, monarch, royal, leader, reign*). Both critical words are nouns referring to types of people, and the lists are virtually identical in terms of word association (BAS of .245 and .230, and forward associative strength, FAS, the probability of the critical word eliciting a list word in a word association task, of .053 and .059, respectively). Despite these similarities, the mean false-recall rate is .60 for the *doctor* list, whereas it is .10 for the *king* list.

Roediger, Watson, et al. (2001) examined underlying characteristics of the commonly used DRM stimuli. In multiple regression analyses, the dependent variable was the probability with which a list produced a false memory (the probability of falsely recalling the critical word as having been studied previously), and the predictors were both word list variables and nonpresented critical-word variables. The word list variables included FAS, BAS, interitem associative strength (a measure of connection strength among list items), and veridical recall—the probability of correctly recalling list items. The critical-word variables were word length, word frequency (Kuçera & Francis, 1967), and rated concreteness.

BAS and veridical recall were significant predictors and together accounted for 68% of the variance in false recall. Thus, Roediger, Watson, et al. provided an empirical demonstration of why there are list differences in the elicitation of false memories; the probability of false recall increases with higher BAS and lower veridical memory for list items.

Theories of false memory in the DRM paradigm

Roediger and colleagues (Gallo & Roediger, 2002; Roediger, Balota, & Watson, 2001; Roediger & McDermott, 1995, 2000; Roediger, Watson, et al., 2001) have argued for an activation/monitoring framework to account for false memories in the DRM paradigm. This is a hybrid of ideas, including implicit associative responses (Underwood, 1965), spreading activation (Collins & Loftus, 1975), and source monitoring (Johnson, Hashtroudi, & Lindsay, 1993). Two processes—activation and monitoring—can occur during encoding and retrieval. A central assumption is that information not explicitly presented during encoding can be inferentially activated and processed. In the DRM paradigm, encoding of word lists

produces activation that spreads in a lexical–semantic system due to backward association and can result in the creation of an implicit associative response (an associate to the encoded words, see Underwood, 1965). If it is not rejected as being internally generated (a failure of source monitoring), this generated word will be "remembered" as being presented. False memory is a function of the likelihood that an implicit associative response is activated during either encoding or retrieval, irrespective of being activated consciously (McDermott, 1997) or unconsciously (Seamon, Luo, & Gallo, 1998). The stronger the set of connections from list words to the nonpresented critical word (as indexed by BAS), the more likely the critical word will be falsely remembered. Note that this theory is mute on what underlies BAS itself. Other than being an index of the probability of a list word eliciting the critical word in a word association task, no proposals have been offered regarding the types of semantic relations that underlie those probabilities—that is, the relations that drive participants' responses in a word association task. In fact, Nelson, McEvoy, and Dennis (2000), who are strong proponents of the use of word association as a research tool, acknowledge that little is known about the representations and processes that underlie word association.

The second major framework used to account for false memory in the DRM paradigm is fuzzy-trace theory (Brainerd & Reyna, 2002; Payne, Elie, Blackwell, & Neuschatz, 1996). This theory predates that of Roediger, Watson, et al. (2001) in accounting for false memories and suggestibility effects (for a review see Reyna & Brainerd, 1995). The primary assumption of fuzzy-trace theory is that a verbatim trace and a gist trace result from an encoding experience, and the corresponding processes operate in parallel. A verbatim trace represents the surface form of the presented list items, whereas the gist trace represents the semantic content, including meaning, relations, and patterns (Brainerd & Reyna, 2002). False recall is attributed to gist extraction that occurs during encoding, whereas veridical recall is due to verbatim traces. That is, the fuzzy-trace approach implicates gist extraction or episodic interpretation of the semantic content of the list words as the primary mechanism underlying false memories. This is not to imply that gist extraction will necessarily result in a false memory. In fact, fuzzy-trace theory predicts that false memories can be suppressed through the recollection rejection process (Brainerd, Wright, Reyna, & Mojardin, 2001) that is based on the verbatim trace (for a review, see Brainerd & Reyna, 2005). Roediger, Watson, et al. (2001) argue however, that the notion of "gist extraction" is operationally underspecified, and for the fuzzy-trace account to be complete, the gist representation needs to be specified. Evidence has been reported that supports gist specification in the context of studies examining transitive inference (Brainerd & Kingma, 1984; Reyna & Brainerd, 1990), inclusion illusions (Brainerd & Reyna, 1990, 1995), and false memory for discourse (Brainerd & Reyna, 1993; Reyna & Kiernan, 1994, 1995); however. Brainerd and Reyna (1998) have argued that the gist trace results from various meanings cued by the surface form of the presented list items, and these meanings are represented by the nonpresented critical list word, or list theme. Gallo and Roediger (2002) suggest that gist theory adequately accounts for lists that produce high levels of false memory, but it fails to explain why other lists produce low levels given that all lists are constructed in a similar manner and appear to converge semantically on their respective nonpresented critical item. The present research tests the notion that certain types of semantic relations might be more successful in eliciting a gist that contains the concept corresponding to the critical word.

Semantic relations and knowledge types

One way to address the potential relations underlying BAS is to consider them as feature or knowledge types. Various researchers (Anisfeld & Knapp, 1968; Fillenbaum, 1969; Grossman & Eagle, 1970) have used this approach to understand errors in recognition tasks. In general, these studies demonstrate that synonyms of a presented word are more likely to

be falsely recognized than are control words. They concluded that words are stored as feature complexes, and because synonyms share numerous features, they are more likely to be recognized as having been previously studied. Recently, Brainerd et al. (2008) conducted factor analyses on DRM list associative and lexical variables, as well as numerous semantic variables: seven from Toglia and Battig's (1978) norms, and three regarding emotionality from Bradley and Lang (1999), as well the Wu and Barsalou (2009) knowledge types used below (i.e., Brainerd et al. used our values from Experiment 1 below). They found numerous indicators of strong meaningfulness (stronger than average words) in the DRM lists, in both list words and critical nonpresented words. They also found that false recall, BAS, and meaningfulness loaded on the same factor and thus concluded that they vary together.

In this article, we focus on semantic relations between list items and critical nonpresented words. Recently, Cree and McRae (2003; see also McRae & Cree, 2002) used a knowledge-type taxonomy developed by Wu and Barsalou (2009) to account for behavioural data of category-specific semantic deficit patients. Cree and McRae classified features from a large set of feature production norms for 541 object concepts, taken from McRae, Cree, Seidenberg, and McNorgan (2005). Although those concepts are all concrete nouns, this is not the case in the DRM lists. For example, determining the features of *dog* is fairly straightforward in comparison to determining the features of *wish.* Nonetheless, Wu and Barsalou's taxonomy is a useful tool to understand DRM lists because it affords the opportunity to classify not only features such as physical components of an object, but also knowledge types such as situation features (e.g., for the concept *buy*, <in a store> would be coded as a location feature), and other relations as well. If DRM list words are considered as features (broadly defined) of the critical word, and these relations are classified using the knowledge-type taxonomy, insight into the role of semantic relations in BAS and false recall may be possible.

In Experiment 1, we used Wu and Barsalou's (2009) knowledge-type taxonomy to examine whether BAS varies as a function of different types of semantic relations and whether semantic relations themselves contribute to false memories. Demonstrating that knowledge types are determinants of BAS may provide explanatory power for the notion of "gist". To foreshadow, we found consistent effects of situation features, synonyms, and taxonomic relations, followed by less consistent effects of entity features. Based on these results, new word lists were developed that consist entirely of situation features of a nonpresented event or location, with the situation features having quite low BAS with the critical nonpresented word. These situational feature lists were used to test predictions of activation/monitoring theory and fuzzy-trace theory. The results provide varying degrees of support for both.

EXPERIMENT 1

Method

Materials and procedure—We used the 55 word lists from Roediger, Watson, et al. (2001). Each contains the strongest 15 associates of the nonpresented critical word, although Roediger and McDermott (1995) state that a few words were substituted in some lists because they seemed more likely to elicit the critical word. In addition, probability of false recall and BAS were taken from Roediger, Watson, et al. They obtained the majority of BAS values from Nelson et al.'s (1999) word association norms, whereas for words not included in those norms, they collected BAS values using Nelson et al.'s procedure.

Wu and Barsalou's (2009) knowledge-type taxonomy, with a few minor alterations, was used to classify the relationship between the critical word and each list word. For example, for the critical word *doctor*, *physician* was coded as a synonym and *hospital* as a location, which falls under the major knowledge-type *situation features*. The taxonomy consisted of

32 knowledge types, organized within six major classes (see Appendix A). Three minor alterations to Wu and Barsalou's taxonomy were made to accommodate the DRM lists. As in Cree and McRae (2003), the knowledge type *made of* was added to the major class of *entity features.* Also, *synonyms* was removed from *taxonomic relations* and was treated as a separate major knowledge type due to its assumed importance in eliciting false memories. Furthermore, *antonyms* was added as a major knowledge type because although no antonyms are produced when people list features for object concepts, they are produced in the word association task, particularly when the stimulus is an adjective or verb. This resulted in six major knowledge types that formed the bases of our analyses: situation features, synonyms, taxonomic relations, antonyms, entity features, and introspective features.

Two independent raters coded the semantic relations between each of the 825 DRM list words and their respective nonpresented critical words. One rater was the first author, and another was a Research Assistant who was unaware of the goals of the present study. Disagreements were discussed with the result that 813 of the DRM list words were coded for an agreement rate of 98.5%. Only items in which agreement was reached were used in any analysis reported below (9 items coded as miscellaneous were also excluded). The initial step was to use WordNet (2001; Version 1.7.1) to determine synonyms and antonyms of the critical word. Following this, the remaining list words were coded at the specific level in the taxonomy. The individual feature types were then summed for each of the six major knowledge types within which they were nested. The number of words per DRM list in each of these categories potentially ranged from 0 to 15 and served as predictor variables.

Results and discussion

Descriptive statistics for knowledge types—We begin by providing descriptive statistics regarding the prevalence of the knowledge-type relations in the DRM lists (see Appendix B). The prevalence of relations depends to some extent on the nature of the critical word. For example, synonyms and antonyms are most likely to occur for adjectives such as *beautiful* (*lovely*), or verbs such as *sleep* (*snooze*), although they do occur for some concrete nouns such as *trash* (*garbage*, *refuse*). Twenty-seven of the 55 lists contained at least 1 synonym, with 11 lists containing 2 or more. Overall, there were fewer antonyms (24) than synonyms (56), with only 16 lists containing at least 1 antonym. There were 200 list words referring to taxonomic knowledge, with about half being category coordinates (referring to somewhat similar concepts). Forty-three lists contained at least one taxonomically related word. There were 146 entity features across 33 lists, with the vast majority occurring for critical words denoting a concrete object or entity. Ninety-one list words spread across 33 lists referred to introspections. Finally, of greatest relevance to Experiments 2, 3, and 4, the most frequent semantic relation was aspects of situations in which the concept denoted by a critical word typically takes part (287). Fifty-two lists included at least one aspect of a situation.

We used regression analyses and analyses of variance (ANOVAs) to show that the probability of false recall is higher for lists containing a greater number of aspects of situations to which the critical concept applies, synonyms of the critical nonpresented word, taxonomic relations, and, to a lesser extent, a greater number of physical features of the entity or object denoted by the critical word.

Predicting BAS with knowledge types—To show that BAS itself reflects differences in the six knowledge types, they were regressed onto BAS. However, multicollinearity issues were indicated by tolerance levels of the knowledge types being close to zero (.011, .011, .012, .017, .045, .201) and the condition index exceeding the suggested value of 30

(100.76; Tabachnick & Fidell, 2001). Therefore, we removed introspective features from these regression analyses. This resolved the multicollinearity problem, and other analyses indicated that introspective features played little role in accounting for either BAS or false recall (primarily because there were few of them). Without introspective features, both the tolerance levels (.465, .483, .498, .596, .798) and condition index (12) were well within acceptable levels (Tabachnick & Fidell, 2001). The number of study words per list corresponding to each of the remaining five knowledge types accounted for 25% of the variance in BAS in a simultaneous regression, $F(5, 49) = 3.20$, $p < .01$, $R = .50$ (see Table 1 for the results). Partial correlations showed that the number of situation features, taxonomic relations, and antonyms significantly predicted BAS. Synonyms were a marginal predictor ($p = .053$). Thus, lists that contain a greater number of situation features, taxonomic relations, antonyms, and synonyms tend to have higher mean BAS in word association.

Predicting false recall with BAS and knowledge types—To further understand the relation between BAS and its constituent components, a simultaneous multiple regression analysis was conducted in which BAS and the five knowledge types were used to predict the probability of false recall. These variables accounted for 59% of the variance in false recall, $F(6, 48) = 11.31$, $p < .001$, $R = .77$ (see Table 2). BAS was the sole significant predictor of false recall, further suggesting that the knowledge types are components of BAS because they do not contribute unique variance in predicting false recall.

To further support the argument that knowledge types, at least in part, are constituent components of BAS, we predicted the probability of false recall using aggregated regression scores, which were the standardized residuals from the first analysis in which knowledge types were regressed onto BAS. This aggregate represents only the variance that the knowledge types accounted for in BAS, and it significantly predicted false recall, $F(1, 53) = 39.28$, $p < .001$, $R = .65$, accounting for 43% of the variance. However, this percentage of variance is based on the possibility of accounting for 100% of the variance in false recall, and as outlined in Roediger, Watson, et al. (2001), the maximum amount of explainable variance is 81% (based on the reported split-half correlation of +.90 for the DRM lists). After adjustment, 53% of the explainable variance in false recall was accounted for by the knowledge types ($.426 \div .81 \times 100 = 52.6\%$). This further supports the assumption that knowledge types are at least one underlying component of BAS and that they aid in understanding and explaining why false memories occur in the DRM paradigm.

The key measures of BAS and knowledge types differ on critical dimensions. As noted earlier, BAS provides little insight into the types of semantic relations that link list words to their corresponding critical word. However, it provides an operational measure of relational strength—namely, the number of participants in a word association task who produced the critical word given a list word as a stimulus. In contrast, the knowledge types denote semantic relations, but for each DRM list in the previous analyses, we simply counted, for example, the number of situation features. Thus, each situational feature was weighted identically in the previous analyses. For example, if two DRM lists each contain 4 situational features, the value of the situational feature measure in the previous analyses was 4 for each list. However, those analyses fail to take into account the possibility that the situational features on each list may have different BAS values. That is, one DRM list with 4 situational features, where each situational feature has a value of .20, would have a total BAS value of .80 for the 4 situational features on that list. In contrast, another DRM list with 4 situational features, in which each situational feature has a BAS value of .10, would result in a total BAS value of .40 for the 4 situational features on that list. Critical analyses for identifying a specific intrusion, therefore, may be ones in which these two measures are combined, with Wu and Barsalou (2009) knowledge types providing semantic relations and BAS values providing an estimate of the relational strength.

To conduct this weighted knowledge-type analysis, we summed the BAS values across all of the list words of each knowledge type within a list (e.g., summing the BAS values for the three synonyms of *anger* in that list). Because no multi-collinearity issues existed, introspective features were included (tolerance levels = .743, .768, .806, .860, .930, .946, and condition index = 3.4). In a simultaneous regression, these weighted knowledge-type scores predicted 52% of the variance in probability of false recall, $F(6, 48) = 8.82$, $p < .001$, $R = .72$ (see Table 3). Taking 81% of the variance as the maximum possible value, the weighted knowledge types predicted 64% of the explainable variance in false recall. Weighted measures of situational knowledge, synonyms, taxonomic relations, and entity features were significant predictors.

Finally, we conducted one-way analyses of variance to examine further the influence of weighted knowledge types on false-recall rates. We divided the 55 DRM lists into three groups based on the probability of false recall, resulting in 18 low-intrusion lists with a false-recall probability of .18 or less, 19 medium-level intrusion lists (.19 to .37), and 18 high-intrusion lists (greater than .38). Summed BAS for each knowledge type was the dependent variable, and level of false recall (low vs. medium vs. high) was the independent variable. Thus, the analyses tested whether the high-, medium-, and low-intrusion lists differed in terms of each weighted knowledge type. Table 4 presents the means and standard deviations for each analysis. There were significant differences across the levels of false recall for three weighted knowledge types: situation features, $F(2, 54) = 7.53$, $MSE = 0.49$, $p < .001$; synonyms, $F(2, 54) = 5.76$, $MSE = 0.23$, $p < .01$; and taxonomic relations, $F(2, 54) = 5.26$, $MSE = 0.56$, $p < .01$. Finally, entity features approached significance, $F(2, 54) = 2.68$, $MSE = 0.15$, $p < .08$, whereas antonyms, $F(2, 54) = 2.19$, $MSE = 0.11$, $p > .1$, and introspective features, $F(2, 54) = 1.37$, $MSE = 0.005$, $p > .2$, were nonsignificant.

Tukey's post hoc tests showed that situation-based knowledge had larger weighted scores for high-intrusion lists ($M = 1.04$) than for low lists ($M = .14$), but high did not differ significantly from medium lists. Medium- and low-intrusion lists did not differ significantly. Synonym scores were larger for high-intrusion lists ($M = .58$) than for low ($M = .06$) and medium lists ($M = .18$; ps $< .05$). Medium and low lists did not differ significantly. For taxonomic relations, only high- ($M = .86$) and low-intrusion ($M = .08$) lists differed significantly.

In summary, the primary result is that it is possible to describe BAS, the variable most often implicated as a predictor of false memories in the DRM paradigm, in terms of semantic relations represented by knowledge types. Aggregating across all of the analyses, situation features, synonyms, and taxonomic relations were consistently predictors of BAS and false recall. This finding nicely converges with results reported in Brainerd et al. (2008; see their Table 6) in which they examined a stratified random sample of word pairs from Nelson et al. (1999) using the Wu and Barsalou (2009) variables (and other semantic variables as well). Brainerd et al. found that Wu and Barsalou knowledge types accounted for variance in word association strength, demonstrating that the current findings generalize beyond DRM lists. Thus, these results suggest a response to the criticism that fuzzy-trace theory cannot specify when gist is extracted from a list of words. We propose that false-recall variability observed in the DRM lists is influenced by the number of situation features and other knowledge types, these semantic relations evoke a gist representation, and that the extraction of gist contributes to false recall of nonpresented critical words. Elements of a situation promote gist extraction because in combination they provide cues to a scenario that is described by a critical nonpresented word, or a scenario in which the nonpresented concept participates. Synonyms are somewhat different in that they promote gist because the concept overlaps with the nonpresented critical word. Finally, the influence of taxonomic relations, many of which are category coordinates, is slightly less clear. It could be the case that they promote

gist because of featural overlap with the critical non-presented concept (akin to synonyms). On the other hand, category coordinates often are found in the same situations, in which case they may exert more of a situational influence.

EXPERIMENT 2

The purpose of Experiment 2 was to test whether DRM-like lists composed of entirely situation features would be successful in producing a gist. That is, as a preliminary to studying false-recall rates, as was done in Experiments 3 and 4, we began by asking participants to explicitly produce a word corresponding to the gist of a word list (for a similar procedure see Brainerd & Reyna, 1998; Brainerd et al., 2001). The idea that gist or a schema-like representation influences memory performance is not new. For instance, Bartlett (1932) introduced the idea of schemas when arguing for the reconstructive nature of memory in accounting for the recall of complex stories. Situation features are one knowledge type that could lead to Brainerd and Reyna's (2002) notion of gist extraction. Specifically, Reyna and Brainerd (1995) define gist as memory for the substance of an experience, and in the current context this would be represented by the word or theme that summarizes a particular word list. Gist is an emergent property that arises from a verbatim experience, with the latter being more specific in content and detail. We hypothesize that if all list words were schema related (i.e., formed a set of coherently related situation features), then participants would be highly likely to intrude other aspects of the situation, or the label for that situation. Consider the following list: *barn*, *tractor*, *hay*, *fence*, *pig*, *mud*, *field*, *house*, *stables*, *cow*, *pitchfork*, and *rooster*. This study list, which has not been used in DRM studies, should elicit a gist or schema for a farm, thus potentially resulting in intrusions of the word *farm* (or of other things that are found at farms, such as *horse*).

In Experiment 2, 12 DRM-like lists were constructed based on situation features (rather than BAS). The goal was to test whether these lists evoke a gist or semantic theme, irrespective of associative strength. We compared these situation lists to 6 DRM lists, divided into 3 low-to-mid and 3 high false-recall lists. According to fuzzy-trace theory (Brainerd & Reyna, 2002), false memories in the DRM paradigm result from the episodic interpretation of a gist representation, and, therefore, higher rates of false recall in the DRM paradigm result from lists that better produce a coherent gist. As such, participants may respond more often with the gist (specifically, the nonpresented critical word) for the higher than for the lower recall lists. Moreover, because the situation lists were constructed to create a strong gist representation, these lists may produce a higher proportion of gist responses than even the high-recall DRM list.

Method

Participants—Forty-three undergraduate students from an introductory cognitive psychology course at King's University College (affiliated with the University of Western Ontario) participated as part of a classroom demonstration on memory.

Materials—Twelve situation lists, such as the farm list presented above, were constructed (see Appendix C). Each contained 12 situation features corresponding to people, animals, or objects typically found at common events or locations. The situation lists were based on production norms collected by Hare, Jones, Thomson, Kelly, and McRae (2009). Hare et al. presented participants with an event or location noun (e.g., farm) and asked them to list either types of things, or types of people and animals, that typically are found at those locations or events. Participants provided up to five responses. Responses were scored on the basis of the number of participants that produced them, as well as on their rank. A weighted score was calculated for each response by summing the number of participants who produced it first times five, second times four, and so on. For each list, the top 12

responses were used, except when avoiding having identical items on multiple lists, or when a top 12 response consisted of multiple words (e.g., *fire engine* for the event *accident*).

In addition, six DRM lists were used. Three (*window, smell, sleep)* were selected from Roediger, Watson, et al.'s (2001) norms for their high rates of false recall, and three (*king, fruit, lion)* for their relatively low rates. To equate the number of DRM-list items with the 12-item situation lists, the final three items from each of the 15-item DRM lists were removed (those with the lowest FAS values). To test whether high false-recall lists more successfully converge on a single concept regardless of gist-inducing features, the two DRM groups were roughly balanced overall in terms of gist-inducing features identified in Experiment 1, with the high false-recall group containing a greater number of synonyms on average (2.0 vs. 0), and situation features (4.3 vs. 3.7), but fewer taxonomic relations (0.7 vs. 5.3, mainly due to the eight exemplars on the *fruit* list).

Word association norms for the situation lists: To minimize effects of BAS, we ensured that the situation lists were low on this variable. Where possible, BAS values were taken from Nelson et al.'s (1999) norms. Of the 144 situation list items, 62 had BAS values in Nelson et al.'s norms. For the remaining 82 items, norms were collected using their procedure. Over 180 undergraduate students (who did not take part in other experiments reported herein) were given one of two norming sheets, each containing half of the items. Thirty-three filler items were added to each norming sheet and at least 5 items intervened between any 2 items from a single situation list. Participants were instructed to write on the blank line provided next to each item the first word it brought to mind. Over 90 observations were obtained for each item, and many had slightly over 100 observations. Mean BAS for the situation lists were low, ranging from .02 to .11 (*accident* = .02, *airport* = .03, *breakfast* = .03, *concert* = .03, *theatre* = .04, *bathroom* = .05, *arrest* = .05, *farm* = .06, *gym* = .06, *funeral* = .09, *casino* = .09, and *war* = .11). By comparison, DRM lists that are effective at producing false recall typically have BAS values above .2 with less effective DRM lists typically being below .1 (Roediger, Watson, et al., 2001). As such, BAS for our situation lists are more in line with less effective DRM lists. The 12 situation lists had a mean BAS of .06, whereas the 3 low–mid false-recall DRM lists used in this study averaged .21, and the 3 high false-recall DRM lists averaged .33.

Procedure—Participants were instructed that they would hear a series of word lists and that they would be asked to write down one word that they thought best described the lists of words they had just heard. They were provided with a sheet of paper that had 18 lines labelled "list 1", "list 2", and so on. Each list was read aloud by the first author at an approximate rate of 2 seconds per word. Participants then were given a few seconds to respond. This procedure was repeated for all 18 lists. The lists were presented in the following semirandom order: *accident, window, breakfast, funeral, concert, smell, farm, gym, sleep, theatre, king, bathroom, fruit, war, casino, airport, lion*, and *arrest*. The task took approximately 10 minutes.

Results and discussion

The purpose was to determine whether the situation lists produce a gist representation, which was defined as participants producing the nonpresented critical word. In the present study, this response is not a false memory per se, although high rates of consistent gist-based responding indicates that the situation lists are able to elicit gist. This would suggest that these lists might be able to create false memories in a DRM paradigm.

Over the 12 situation lists, 64% (*SD* = 23%) of participants produced the critical event or location noun (see Table 5). The ability of the situation lists to elicit the critical word ranged

from 5% for *arrest* to 86% for *casino*. Note that the second worst list was *airport* at 51%. Although the *arrest* list is therefore an anomaly, if a more liberal criterion is applied by including all gist-consistent responses, then it becomes one of the best situation lists. For example, although only 2 participants produced the word *arrest*, there were 16 other gist-consistent responses (e.g., *criminal*, *illegal*, *law*, *cops*, *legal*, *guilty*, *trial*). As such, even this list can be considered fairly successful at eliciting gist.

A repeated measures ANOVA was conducted to compare the proportion of critical-word responses for the situation, high false-recall DRM, and low–mid DRM lists. This analysis used the strict response criterion (i.e., using only the actual event or location names). List type influenced the proportion of critical-word responses, $F(2, 72) = 25.16$, $p < .001$, $MSE = .05$ (Greenhouse–Geisser degrees of freedom). Planned comparisons showed that the mean proportion of critical-word responses was greater for the situation lists ($M = .64$, $SD = .36$) than for the high-recall ($M = .44$, $SD = .36$), $F(1, 42) = 25.81$, $p < .001$, and low–mid-recall ($M = .33$, $SD = .24$) DRM lists, $F(1, 42) = 61.15$, $p < .001$. Thus, consistent with fuzzy-trace theory, the stronger the assumed gist of the lists, the higher the probability that participants produced the critical nonpresented word. In addition, high-recall DRM lists were more effective than the low–mid-recall lists, $F(1, 42) = 4.41$, $p < .05$. Thus, as in the Experiment 1 weighted knowledge-type analyses, and consistent with activation/monitoring theory (Roediger, Watson, et al., 2001), relational strength as measured by BAS played a role as well.

It should be noted that there was substantial variability in the low–mid-recall lists that highlights another aspect of gist. That is, it is not simply the number of list items of various relation types that matters but their coherence and the degree to which they direct people toward a specific concept matters as well. For example, 65% of participants responded with "fruit" given the *fruit* list. This occurred presumably because this list contains eight types of fruit, making the gist quite obvious. Note that although gist is strong for the *fruit* list, false-recall probability is only .2, presumably because source monitoring allows many participants to distinguish the superordinate category name from the basic-level fruits on the list. On the other hand, only 5% of participants responded with *king* given the *king* list. This list was mentioned in the introduction as one with a surprisingly low false-recall level despite strong BAS. In Experiment 1, participants did get the gist (e.g., 15 produced *royalty*), but responses varied (*majesty*, *noble*, *regal*, etc.), and the gist did not converge directly onto *king*.

In summary, two notable findings emerged. First, situation lists evoked a clear gist representation. That is, they successfully converged on a single concept that represents the overall theme or gist. This result was not due to BAS because the situation lists had low BAS. Second, the high false-recall DRM lists produced more consistent gist responses than did the low–mid recall lists. This finding supports arguments proposed by Gallo (2006), Gallo and Roediger (2002), and Roediger, Watson, et al. (2001), that DRM lists produce robust rates of false recall because they converge on a single critical word. These authors argue that this is due to BAS.

Second, the results are consistent with Brainerd and Reyna's (2002) suggestion that lists with stronger gist should produce higher rates of false memories. Although Experiment 2 did not assess false memory per se, the results support our contention that situation lists provide one basis for a valid operational definition of gist extraction and that, consistent with the results of Experiment 1, high false-recall lists may be those with semantic relations that encourage gist extraction. The highest gist response rates were for situation lists, followed by high false-recall DRM lists, and then low–mid DRM lists, even though the situation lists had the lowest BAS. We have not shown however, that the situational lists

actually produce false memories in a DRM paradigm. This was the aim of the next two experiments.

EXPERIMENT 3

The purpose was to test whether word lists composed entirely of situational information produce gist-based false memories. There is some uncertainty, however, about the levels at which false memories would be expected. False-memory production depends on two processes: the generation of plausible alternative list members, and the failure of monitoring processes (source monitoring as in activation/monitoring theory, Roediger, Watson, et al., 2001, or verbatim-based recollection rejection, as in fuzzy-trace theory, Brainerd, Reyna, Wright, & Mojardin, 2003). With the situation lists, one possibility is that they might lead participants to internally generate a strong, coherent gist-related item that is likely to be attributed to an external source, thus resulting in high false-memory rates. Another possibility is that because the gist is either an event or location, whereas list items denote types of people, animals, or objects, the gist concepts are obviously distinct and thus might easily be rejected during monitoring.

False memories are measured in DRM studies solely in terms of the critical nonpresented word (Roediger & McDermott, 1995). Using measures such as BAS and FAS as the central variables leads to a perspective in which a researcher measures intrusions for a single predetermined concept, the critical word that is the basis for the word association norms. However, based on fuzzy-trace theory (Brainerd & Reyna, 2002) and the idea that gist extraction results from a schema-like representation formed during episodic interpretation of an experienced event (Alba & Hasher, 1983), situation lists may also produce other intrusions. That is, a knowledge-type or gist explanation of false recall leads to a difference in how false memories might be measured. A gist account in which a DRM list consists of, for example, primarily aspects of a *farm*, leads a researcher to measure intrusions for a set of concepts that include the label for that situation, but also other aspects of that situation as well (e.g., other farm animals such as *horse*, which is not on the *farm* list).

Method

Participants—Sixty-one undergraduate students from an introductory psychology course at King's University College participated as part of a classroom demonstration on memory.

Materials and procedure—The 12 situation lists from Experiment 2 were used. Participants were instructed that they would hear a series of word lists and that after hearing each, they would be asked to recall as many of the words as they could remember by writing them on a sheet provided. As is typical in DRM studies, the participants were instructed not to guess, but to include only words that they were sure they heard when the list was read to them (Roediger & McDermott, 1995). The first author read lists aloud at an approximate rate of 2 seconds per word. After each list, participants were given 2 minutes to recall words. This procedure was repeated until all 12 situation lists had been presented for study and recall. The task took approximately 30 minutes.

Results and discussion

Three primary measures are of interest. First, veridical memory was measured as the proportion of list items accurately recalled. Second, DRM-style false memories were measured as the proportion of intrusions of the critical nonpresented word (e.g., *farm*). Third, we measured schema- or gist-consistent false memories, such as *horse* for the *farm* list. The data are presented in Table 6.

Veridical memory performance—Mean veridical recall rate was .67 (*SD* = .10). The average across all 55 DRM lists in Roediger, Watson, et al. (2001) is a comparable .62 (*SD* = .06). As such, veridical memory for situation list items is similar to that in studies using BAS-based DRM lists.

DRM-like false memories—Participants produced low rates of nonpresented critical-word intrusions. The mean proportion was .05 (*SD* = .08), ranging from .02 for *concert* to . 11 for *farm*. These are similar to some of the least successful DRM lists and is probably due to successful monitoring processes, as tested in Experiment 4.

Gist-consistent false memories—Gist-consistent intrusions are reported in terms of their mean number per participant for each list rather than as a proportion because there are potentially greater than one per list. Not a single random response was produced; all bore an obvious relation to the general theme of the lists. For example, for the *war* list, gist-consistent intrusions included *military*, *bullet*, and *Air Force*, and for the *arrest* list, they included *prison, plaintiff, judge*, and *criminal.* Summing over the 12 situation lists, participants produced an average of 2.05 (*SD* = 2.35) schema-consistent intrusions (or false memories), ranging from 0 to 11. Situation lists varied considerably, ranging from very few intrusions for *breakfast* (*M* = 0.03, *SD* = 0.18) to a moderate number for *arrest* (*M* = 0.33, *SD* = 0.70). In general, and consistent with earlier schema research such as Brewer and Treyens (1981), the situation lists lead to gist-consistent false memories.

Relations among the three memory measures—Correlations were examined to determine the relations among veridical situation feature list memory and the two measures of false memory. There was a negative correlation between veridical memory and DRM-like intrusions indicating that the greater the number of list items that participants correctly recalled, the less likely they were to falsely recall the nonpresented critical word, $r(59) = -.30$, $p < .02$. This finding is consistent with studies that have used DRM lists (Gallo, 2006). The relation between veridical memory and gist-consistent intrusions was also negative and significant, $r(59) = -.29$, $p < .02$. Participants were less likely to make a gist-consistent intrusion as their veridical memory increased. The two false-memory measures correlated positively, indicating that as participants made more false-memory intrusions of the event or location names, they also tended to make more schema-consistent intrusions, $r(59) = .61$, $p < .001$.

In summary, the situation lists elicit false memories at a low rate, particularly when the event or location word was the target. The data are thus somewhat inconsistent. On the one hand, they are weakly supportive of the idea that false memories can result from forming a schema-like representation (Alba & Hasher, 1983). On the other hand, although Experiment 2 shows that the situation lists strongly evoke gist, Roediger and colleagues (Gallo & Roediger, 2002; Roediger, Watson, et al., 2001) could claim that the low rates of nonpresented critical-word intrusions are best explained by their activation/monitoring theory. According to activation/monitoring theory, DRM false memories are much less likely to occur when BAS is low, which it was for the situation lists.

However, an alternative explanation that is tested in Experiment 4 relates to the source-monitoring process proposed to operate in activation/monitoring theory (Roediger, Watson, et al., 2001) and the recollection rejection process in fuzzy-trace theory (Brainerd et al., 2001). In fuzzy-trace theory, editing out gist-consistent information from a memory report results from recollection rejection (Brainerd et al., 2003). This process relies on the verbatim trace of the studied list items to suppress false memories. Specifically, recollection rejection can occur when verbatim traces are strong and thus are able to edit out gist reconstructions regardless of the strength of the gist trace. As such, it may be the case that the low rates of

critical-word intrusions in Experiment 3 result from verbatim traces overriding gist traces, resulting in successful monitoring. The veridical recall rates are consistent with this idea.

The activation/monitoring account proposes that regardless of whether the nonpresented critical word is activated during encoding or retrieval, false memories are produced when people fail to successfully monitor the source of their memory (Roediger, Watson, et al., 2001). Source misattributions in memory can result from confusion of two external sources (e.g., confusing who told an amusing story, Dave or Albert), or of an internal and external source, as is argued to be the case in the DRM paradigm (Johnson et al., 1993). As such, the low false-recall rates in Experiment 3 might reflect correct source monitoring decisions. This explanation is reasonable given the materials. The nonpresented critical item for each of the situation lists was either a location (*farm*) or an event (*breakfast*), whereas the items included types of objects and people that are commonly found at those locations and events. Because the event and location concepts are quite distinct from the concepts presented for study, it may be easy for participants to distinguish between them and therefore reject locations and event names as having been presented.

EXPERIMENT 4

The primary purpose was to examine the monitoring issues raised above. To do so, a two-minute distractor task, during which participants performed arithmetic calculations, was inserted between the study and free-recall phases. The purpose was to prevent overt rehearsal of the studied list words, which may produce greater monitoring difficulty during free recall, or verbatim suppression in terms of fuzzy-trace theory (Brainerd et al., 2003). If the low false-recall rates in Experiment 3 were due to participants' successful monitoring processes, the delay should lead to higher rates.

In contrast to Experiment 3, we used four types of lists. Three situation lists comprised one list type, and the other three list types were each taken from DRM word lists (Roediger, Watson, et al., 2001). The three DRM list types were chosen on the basis of their BAS and the distribution of their knowledge types from Experiment 1. This allowed direct comparison among the four list types to examine false memories that result from word lists differing in terms of BAS and gist strength. We also included a recognition test at the end of the study. This should also reduce the effectiveness of monitoring because it tests memory for all lists simultaneously. Note that we are not claiming that source monitoring or recollection rejection plays no role in a recognition test. Rather, participants' ability to source monitor (or recollection reject) is a matter of degree. Monitoring should be more difficult in general when a participant is asked whether a word appeared on any of the 12 previously presented lists (i.e., whether it was 1 of 144 presented words, as was the case in Experiment 4), versus recall in which a participant is dealing with a single list at a time. In addition, because there is a longer time period between study and test in the recognition than in the recall task, it should be more difficult to correctly monitor the source of an item in the recognition task. This should be the case particularly for the situation lists because, in the recognition task, all critical nonpresented words were not events or locations, and all list words were not types of people, animals, and objects.

Further, the recognition task allows for the assessment of participants' phenomenological experience during recollection by using the remember/know procedure developed by Tulving (1985). The recognition test requires participants to judge each item as old (previously studied item) or new (nonstudied item). For each item judged as old, participants indicate if they remember or know that the item was present on the study list. As Roediger and McDermott (1995) note, a remember judgement is defined as the mental reliving of an experience by the participant based on information such as the physical characteristics

associated with the items' presentation, what the participant was doing when they experienced the item, and what the item made them think of, and perhaps by recalling the neighbours of the item on the word list. In contrast, a know judgement reflects that the participant is confident that the item was on the list but they are unable to mentally reexperience its actual occurrence. The primary purpose for including the remember/know procedure was to examine the pattern of judgements for items that were previously studied and for the nonpresented critical items.

Method

Participants—Thirty-seven undergraduate students from an introductory psychology course at King's University College participated in groups of 5 or fewer.

Materials—There were 12 lists, each composed of 12 items. Three were situation lists from Experiment 2: *breakfast*, *casino*, and *farm*. The other 9 lists differed on the probability of false recall and BAS from Roediger, Watson, et al.'s (2001) norms, and/or the number of features that we argued in Experiment 1 distinguished among lists that were likely to lead to gist extraction (i.e., situation features, synonyms, and taxonomic relations). Thus, there were three strong lists, three medium-gist lists, and three low-gist DRM lists. The three strong lists (*doctor*, *smoke*, *sleep*) had a mean false-recall probability of .58, mean BAS of .28, and an average of 10.7 items identified as gist-inducing knowledge types. Thus, these were "strong" on all dimensions. The mid-gist lists (*beautiful*, *bitter*, *butterfly*) had a mean false-recall probability of .02, mean BAS of .03, and an average of 9.3 gist-inducing items. The low-gist lists (*long*, *trouble*, *whistle*) had a mean false-recall probability of .08, mean BAS of .02, and an average of 6.00 gist-inducing items. The three situation lists had a mean BAS of .06, and their mean false-recall rate from Experiment 2 was .08. The lists were arranged so that no two lists of the same type appeared in consecutive order during presentation. The lists were recorded on audio tape in a male voice and were presented at a rate of 2 seconds per word.

Two critical dimensions are important for testing the claims of fuzzy-trace (Brainerd & Reyna, 2002) and activation/monitoring theory (Roediger, Watson, et al., 2001). The first is mean BAS because of its importance in activation/monitoring theory. A one-way ANOVA on BAS showed differences among list types, $F(3, 140) = 32.96$, $MSE = .020$, $p < .001$. Tukey's HSD tests indicated that the strong lists had higher BAS than the other three types, which did not differ. Therefore, if associative strength is the sole critical variable for producing false memories, then the strong lists should be most successful in doing so in both recall and recognition, with the other three being equal.

The second critical dimension is the number of gist-inducing items on each list. A one-way ANOVA on the total number of situation features, synonyms, plus taxonomic relations per list showed a significant effect of list type, $F(3, 8) = 6.82$, $MSE = 2.92$, $p < .02$. Tukey's HSD tests indicated that the number of these types of items did not differ between situation and strong lists (although the lists did differ on BAS). Situation lists contained a greater number of gist-inducing items than did either the mid- or low-gist lists. The strong and mid-gist lists did not differ significantly (although they did differ on BAS). The mid- and low-gist lists differed significantly on gist-inducing items.

The recognition test consisted of 36 studied items taken from Serial Positions 1, 8, and 10 on each of the 12 word lists. A further 36 items that had not been previously presented to participants were also included. These were the 12 nonpresented critical items along with 24 fillers randomly selected from other DRM word lists not used in the current study. These 72 items were randomly ordered. Participants were asked to circle "Old" or "New" next to each

item. Further, if they indicated that an item on the recognition test was "Old", they were asked to circle an "R" to indicate a *remember* judgement, or a "K" to indicate *know*.

Procedure—Participants were instructed as follows: "In a moment you will hear a list of words read one word at a time. After each list is presented you will be asked to complete 2 minutes of arithmetic problems and then you will have 1 minute to recall as many of the list words as possible." After the experimenter read these instructions, the first list was played. The participants then were instructed to work on arithmetic problems for 2 minutes. They then were given one minute to recall as many of the words that they could remember on a sheet provided. This procedure was repeated until all 12 lists had been presented and recalled.

Subsequently, the recognition test was administered. Participants were instructed that if they remembered that the word was presented earlier on the tape player, then they should circle "Old". If they did not have a memory for a particular word being presented earlier, they should circle "New". For any word given an old judgement, the participants were also asked to make a remember/know judgement by circling "R" or "K", respectively. To make a remember judgement, participants were instructed that they should have a specific memory of that word being on the lists presented earlier. This could include something distinct about the speaker's voice when he said the word, or perhaps they might remember a word or words that came before or after the specific word on the list. They were instructed to make a know judgement to an old word if they did not have a specific memory for the word being on one of the lists presented earlier, but that they were fairly certain that the word had been presented earlier. Participants were then asked whether they understood the old/knew and remember/know distinction before they were instructed to begin the recognition test. The experiment took approximately one hour.

Results and discussion

The results are presented in five sections. Veridical recall is reported first, then false recall. Veridical and false recognition then are presented. The final section presents the remember/know judgements and it is subdivided into these judgement types based on veridical and false recognition. Veridical memory results were followed up with Tukey's HSD post hoc tests because no specific predictions were made concerning those data. False-memory results were followed up with pairwise planned comparisons because specific tests were planned for those data. Veridical and false-recall and recognition rates are presented in Table 7. Remember/know rates are presented in Table 8.

Veridical recall—Veridical recall was measured as the proportion of correctly recalled studied items. Overall veridical recall collapsed across the four list types was quite good, .61. A repeated measures ANOVA showed an influence of list type, $F(3, 102) = 27.12$, $MSE = .006$, $p < .001$. Tukey's HSD revealed significantly higher veridical recall for the situation lists than for the strong ($p < .05$), mid-gist ($p < .05$), and low-gist lists ($p < .01$). Strong lists differed from low- ($p < .01$) but not mid-gist lists. The mid-gist lists also had a higher veridical recall rate than did the low-gist lists ($p < .01$). In general, a greater number of gist-inducing knowledge-type items on a word list resulted in better veridical recall, with the best recall being for situation lists.

False recall—False recall was measured as the proportion of participants who recalled the nonpresented critical list word for each list. A repeated measures ANOVA revealed a significant influence of list type, $F(3, 92) = 36.58$, $MSE = .057$, $p < .001$. Planned comparisons indicated significantly higher false-recall rates for the strong lists than for the situation, $F(1, 108) = 52.46$, $p < .0001$, mid-gist, $F(1, 108) = 87.69$, $p < .0001$, and low-gist

lists, $F(1, 108) = 71.92$, $p < .0001$. False recall was higher for the situation lists than for the mid-, $F(1, 108) = 4.50$, $p < .04$, but not the low-gist lists, $F(1, 108) = 1.54$, $p > .2$. Mid-and low-gist lists did not differ, $F < 1$. These findings are consistent with activation/monitoring theory (Roediger, Watson, et al., 2001) because the word lists with the highest BAS (the strong lists) elicited higher rates of false recall. In addition, they are somewhat consistent with fuzzy-trace theory (Brainerd & Reyna, 2002) because false-recall rates were higher for the situation lists than for the mid-gist lists that were matched on BAS.

Veridical recognition—Veridical recognition (or hit rates) was measured as the proportion of studied list words that were judged as old. A repeated measures ANOVA indicated that veridical recognition differed by list type, $F(3, 102) = 16.66$, $MSE = .012$, $p < .001$. Tukey's HSD revealed that it was higher for situation lists than for strong ($p < .01$), mid-gist ($p < .01$), and low-gist lists ($p < .01$). Veridical recognition was higher for strong than for low-gist lists ($p < .01$), but not mid-gist lists. The mid- and low-gist lists did not differ significantly. These analyses again show that a greater number of gist-inducing knowledge-type items on a word list results in better veridical recall.

False recognition—False recognition (or false-alarm rates), measured as the proportion of nonpresented critical words that were judged as old, differed by list type, $F(3, 100) = 24.11$, $MSE = .079$, $p < .001$. False recognition was higher for situation lists than for mid-gist, $F(1, 108) = 20.78$, $p < .0001$, and low-gist lists, $F(1, 108) = 30.85$, $p < .0001$. Situation and strong lists did not differ significantly, $F(1, 108) = 2.45$, $p > .1$. False recognition was higher for strong lists than for mid-, $F(1, 108) = 37.50$, $p < .0001$, and low-gist lists, $F(1, 108) = 50.73$, $p < .0001$. Finally, mid- and low-gist lists did not differ significantly, $F(1, 108) = 1.00$, $p > .3$. Overall, participants were more likely to falsely recognize a nonpresented critical word from lists that contained a greater number of gist-inducing semantic relations (i.e., situation and strong lists). This is the first case in which the situation lists produced a substantial false-memory effect, with a 65% false-recognition rate. Further, the situation and strong lists, which were equated on the number of gist-inducing items but differed on BAS, produced similar rates of false recognition (although there was a nonsignificant 10% advantage for the strong lists). These findings are consistent with explanations that attribute false memory to gist extraction.

Remember/know judgements

Correct recognition: The proportion of *remember* responses did not differ significantly across list type, $F(3, 99) = 1.27$, $MSE = .019$, $p > .2$. However, *know* responses did, $F(3, 85) = 5.21$, $MSE = .022$, $p < .002$. Tukey's HSD (honestly significant difference) indicated that a significantly greater proportion of *know* responses were produced for situation list items than for mid- ($p < .05$) and low-gist items ($p < .05$). There were no other reliable differences.

False recognition: The proportion of *remember* responses differed across list type, $F(3, 93) = 4.48$, $MSE = .087$, $p < .01$. Situation and strong lists did not differ, $F(1, 108) = 1.03$, $p > .3$. There was a greater proportion of *remember* responses for situation lists than for mid-, $F(1, 108) = 4.10$, $p < .05$, and low-gist lists, $F(1, 108) = 4.10$, $p < .05$. Similarly, a greater proportion of *remember* responses occurred for strong lists than for mid-, $F(1, 108) = 9.23$, $p < .01$, and low-gist lists, $F(1, 108) = 9.23$, $p < .01$. The proportion of *remember* responses was identical for mid- and low-gist lists, $F < 1$. Overall, *remember* responses were more likely to be produced for falsely recognized nonpresented critical items for lists that contain a greater number of gist-inducing items, a result that is consistent with the phantom recollections reported by Brainerd et al. (2001) and argued to result from strong gist-based memory traces.

The *know* responses to nonpresented critical words also differed by list type, $F(3, 99) = 8.45$, $MSE = .073$, $p < .001$. Situation and strong lists did not differ, $F < 1$. The proportion of *know* responses was greater for situation lists than for the mid-, $F(1, 108) = 6.39$, $p < .01$, and low-gist lists, $F(1, 108) = 13.05$, $p < .001$. The proportion of *know* responses was greater for strong than for mid-, $F(1, 108) = 10.83$, $p < .001$, and low-gist lists, $F(1, 108) = 18.88$, $p < .0001$. Mid- and low-gist lists did not differ, $F(1, 108) = 1.08$, $p > .05$. As with the *remember* data, *know* responses are more likely to be produced for falsely recognized nonpresented critical items for lists containing a greater number of gist-inducing items.

Experiment 4 adds further behavioural evidence concerning false-recall variability found in DRM lists and the related theoretical implications. The results support, in part, both activation/monitoring and fuzzy-trace theory. Activation/monitoring theory can generally account for the false-recall results, particularly the fact that high-BAS strong lists produced the highest false-recall rate. On the other hand, situation lists produced a higher false-recall rate than did mid-gist lists. The false-recognition results are more problematic. In activation/monitoring theory, BAS is the critical variable underlying false recall and recognition, but the situation lists produced a false-recognition rate that was similar to the strong lists despite much lower mean BAS. Furthermore, although the situation, mid-gist, and low-gist lists had similar mean BAS, situation lists produced a far higher rate of false recognition. Finally, false-remember rates were similar for situation and strong lists and were much higher for both than for mid- or low-gist lists.

Fuzzy-trace theory predicts that false memories result from processing the gist of an experienced event. In Experiment 4, gist was manipulated in two ways: using lists composed entirely of situation features, and by manipulating the number of gist-inducing items in the three groups of DRM lists. The false-recognition results fall nicely in line with the fuzzy-trace prediction that the stronger the gist of a list of words, the more likely false memories will be produced (although mid- and low-gist lists did not differ). The false-recall data are more potentially problematic, although successful recollection rejection may have decreased false recall of locations and events with the situation lists. That is, the two-minute distractor task between presentation and recall apparently did not sufficiently disrupt source monitoring or reduce the trace of the verbatim trace. As a result, it may have been reasonably easy for participants to discount the nonpresented critical locations and events. In contrast, the recognition task was completed at the end of the study, after all 12 word lists had been studied and recalled. The longer delay plus the intervening lists appear to have decreased the verbatim trace sufficiently.

The observed pattern of *remember* responses made by participants when falsely recognizing a nonpresented critical list word also supports a monitoring explanation. A greater proportion of *remember* responses were made to nonpresented critical items on situation and strong lists than on mid- and low-gist lists. This indicates that participants had more phenomenological experiences associated with nonpresented critical words from word lists designed to have stronger gist.

One potential issue concerns whether false-recognition rates were inflated due to previous recall. Some studies have reported higher rates of false recognition for lists that were previously recalled than for lists that were not. For example, Roediger and McDermott (1995; Experiment 2) found a 9% increase in false recognition of the critical nonpresented words when participants had previously recalled a list versus when they did arithmetic problems rather than recall.

However, although it intuitively seems that previous recall may increase false-recognition rates, in general, there is no or little evidence that it does. For example, Brainerd et al.

(2008) used a set of semantic variables to predict false-recall and recognition rates. They were concerned about the potential of previous list recall to influence recognition. Roediger, Watson, et al.'s (2001) data for 36 of the 55 lists were taken from Stadler, Roediger, and McDermott (1999). In Stadler et al., recall preceded recognition, as in our Experiment 4. Brainerd, Watson, et al. collected false-recognition rates for those 36 lists in an experiment in which the recall phase was excluded. The mean false-recognition rate for the 36 lists was actually slightly lower when recall was included. That is, without a recall phase, Brainerd et al. found a false-recognition rate of .69, and with a recall phase, Stadler et al. found a false-recognition rate of .66. Brainerd et al. also found that the two sets of false-alarm rates were extremely highly correlated at $r = .91$.

Further evidence for the lack of an influence of prior recall on false recognition comes from Gallo (2006). Gallo looked at 14 studies that compared recognition rates when there was or was not previous recall. Overall, false-recognition rates were only 2% higher when recall preceded recognition than when it did not. This difference was significant in Gallo's analyses, but it is only 2%. Therefore, including a previous recall task does not compromise Experiment 4.

Another point regarding the false-recognition rates for the situation lists is that the design of Experiment 4 enables direct comparison among DRM and situation lists. That is, the possibility that recall might inflate false-recognition rates would potentially be problematic if Experiment 4 had not included DRM lists as a direct comparison. Also, if falsely recalling a critical nonpresented word increases the likelihood of falsely recognizing that item, the false-recognition rate for the strong DRM lists should have been much more strongly influenced by the recall task. The false-recall rate was .52 for the strong DRM lists, and .15 for the situation lists.

A further aspect of the Experiment 4 data is that the difference between false-recognition and false-recall rates was much higher for the situation lists than for the DRM lists (situation: .65 – .15 = .50; strong DRM: .75 – .52 = .23; mid-gist DRM: .36 – .05 = .31; low-gist DRM: .30 – .09 = .21). These differences are consistent with our claims about the role of source monitoring with respect to the situation as compared to the DRM lists. When the situation lists are recalled following the presentation of each list, it is relatively easy for a participant to distinguish the name of the event or location from the types of people, animals, or objects that typically are involved in that event or typically are found at that location. However, source monitoring is more difficult in the recognition task, particularly with respect to the situation lists, because participants are now dealing with 12 lists, 3 of which are situation lists, and 9 of which are standard DRM lists (for which the critical nonpresented word is not qualitatively different from the list words, and the list words are a mixture of types of concepts). Therefore, not only has time passed, but that time has been filled with lists of different types. This produced a context in which, during the recognition task, participants had a more difficult time deciding whether the name of the event or location had actually occurred on the list, and thus false recognition of those critical nonpresented words was quite high, and there was no difference in remember judgements for situation versus strong DRM lists. In summary, because the recognition task makes it more difficult for participants to distinguish the event or location name from the list words, false recognition for the situation lists is significantly higher than for the DRM lists that were matched for BAS, and rates were similar to the rates for the strong DRM lists that contained numerous gist-inducing items and had a much higher BAS.

Finally, an experimenter oversight actually provides additional insight into gist formation. Although the low-gist list differed from the other types in mean number of gist-inducing items, *whistle* actually contained eight situation features. However, inspection of these items

illustrates that they do not converge coherently on a gist because they come from multiple types of situation. The *whistle* situation items were *stop*, *dog*, *train*, *song*, *boy*, *blow*, *tune*, and *lips*. Compare those items to those of the *doctor* strong list that contained the eight situation items *sick*, *medicine*, *hospital*, *ill*, *patient*, *office*, *stethoscope*, and *cure* (as well as the synonyms *physician* and *surgeon*). It is obvious that the *doctor* list is much more semantically coherent than is the *whistle* list. Of course, relational strength varied as well, with BAS for the *doctor* situation items being much higher than that for the *whistle* items. Apparently, producing high rates of false memories without high BAS demands list coherence, as in the situation lists that were coherent due to the method by which they were constructed. Supporting this idea is Toglia, Neuschatz, and Goodwin's (1999) finding that thematically grouping items on a list produces higher rates of false recall.

GENERAL DISCUSSION

The research conducted herein was designed primarily to address a central theoretical issue that relates to the two theories commonly used to explain false memory: activation/monitoring and fuzzy-trace. Although both can explain much of the same data, it has been argued that they have strengths and weaknesses. Associative strength provides an operational index of item activation, but is theoretically mute on what underlies BAS. On the other hand, fuzzy-trace theory has a well-developed theoretical rationale but could be more clearly elaborated regarding identifying when gist is aroused. The basic aim of the experiments reported here was to bridge these differences by identifying semantic relations that underlie BAS and by considering whether these relations can provide insight into gist extraction.

In Experiment 1, we used Wu and Barsalou's (2009) taxonomy to classify the semantic relations between DRM list words and their respective non-presented critical words. We found that BAS could be decomposed into various knowledge types and that situation features, synonyms, antonyms, and taxonomic relations predicted BAS. Further, these knowledge types, particularly situation features, synonyms, and taxonomic relations, predict the probability of false recall. As such, these findings provide a theoretical basis for considering associative strength in terms other than the mere probability of one word eliciting another word. Finally, because the lists commonly used in the DRM paradigm vary in both BAS and in the features we take as an index of gist extraction, consideration of gist factors can explain variability in false recall not predicted to date by associative explanations. These conclusions are similar to those of Brainerd et al. (2008), who emphasized that DRM lists have a great deal of strong semantic content, and that meaningfulness correlates with false recall and recognition.

In Experiment 2, a new type of word list was developed to understand the role of situation information in false memories. The situational feature lists strongly converged on the nonpresented event or location. Moreover, DRM lists that are likely to elicit a false memory are also likely to produce a gist response, supporting further the contention that published DRM lists confound BAS and gist characteristics. In Experiment 3, the situation lists produced false-recall rates that were more consistent with associative than gist extraction explanations.

Experiment 4 tested whether the low false-recall rate in Experiment 3 was due to successful monitoring processes engendered by the situation lists in which the intended false memory (an event or location) differed substantially from the list items. In false recall, the situation lists fell between DRM lists with higher and similar BAS. In false recognition however, the situation lists were as good as the strong DRM lists and much better than the mid- or low-gist lists. Remember responses also followed this pattern.

Theoretical implications

The results implicate both associative and gist extraction in false-memory production. Importantly, the findings give gist extraction a more central role than has been the case and thus support fuzzy-trace theory. Even the findings that false recall is better explained by associative strength is consistent with fuzzy-trace theory, which has argued that strong verbatim memory traces reduce false recall of nonpresented critical items through recollection rejection (Brainerd & Reyna, 2002). It is clear that the situation lists enabled participants to form strong verbatim traces because these lists produced high rates of veridical recall. As such, the lack of finding high rates of false recall with the situation lists can be attributed in part to the strength of the verbatim memory trace.

Stronger evidence for fuzzy-trace theory is provided by the Experiment 4 false-recognition results, in which the low-BAS situation lists were similar to the strong high-BAS DRM lists. Other results have discounted associative strength as the sole explanation in recent years. For instance, Meade, Watson, Balota, and Roediger (2007) reported semantic priming experiments in which the DRM lists were used in a lexical decision task. They found that activation processes due to encoding DRM word lists do not last much longer than 1 second (see also Tse & Neely, 2005, for a similar conclusion). They concluded that this activation process does not persist long enough to account for the false-recognition effect in DRM studies. One prediction from the present experiments is that our strongly gist-inducing situation lists should produce false memories that persist over time (see Toglia et al., 1999).

To account for the high false-recognition rates observed in typical DRM studies, Meade et al. (2007) argued that during encoding, related information is brought to mind as a result of spreading activation. However, because this activation persists for only a short time, unless it is maintained in some manner, it is unlikely that any activation received by the nonpresented critical word would be present when a participant engages in a recognition task (see also Zeelenberg & Pecher, 2002). They further argued that items on the recognition task can serve as cues to reactivate an episodic associative network based on a previous episodic experience, which would be the encoding phase. Therefore, false recognition is due to the reactivation of a previously encoded episodic experience. Importantly, Meade et al. maintain the importance of BAS during activation. That is, false recognition is more likely to result from high-BAS DRM word lists. Note that we agree with this conclusion to the extent that BAS indexes relation strength and thus the probability with which list words elicit a coherent gist.

Although the theoretical refinement outlined above is consistent with studies using DRM word lists, it fails to account for why the low-BAS situation lists elicited false-recognition rates that were similar to strong high-BAS DRM lists. Furthermore, reactivation of an episodic associative network fails to account for the false-recognition differences between the situation lists and the mid- and low-gist DRM lists that had similar mean BAS. Clearly, the fact that the situation lists elicited greater false recognition is problematic for theories that emphasize associative strength as the primary theoretical construct.

Another issue that deserves consideration relates to the associations among list items (i.e., interitem associative strength). Interitem associative strength may explain why the situation lists produced very few critical-word intrusions on the recall tasks, but did produce high rates of false recognition. McEvoy, Nelson, and Komatsu (1999; see also Deese, 1959a) have reported that word lists with stronger interitem associative strength result in reduced levels of false recall. McEvoy et al. explained this as follows. On a recall task, the list items activate each other as a result of their strong interitem associations, but in doing so they prevent, or interfere with, the activation of the nonpresented critical word. McEvoy et al. also reported that these strong interitem associations have the opposite effect on recognition

because the strength of these associations leads to more false alarms to the nonpresented critical item. That is, strong interitem association leads to a stronger thematic representation, or gist, of the list items, and this influences recognition more strongly than recall. Although interitem associative strength was not controlled in the current experiments, the interitem connectivity among list items on the situation lists is fairly strong, ranging from a low of 1.17 associative connections among list items on the *airport* list to a high of 3.08 associative connections among list items on the *funeral* list. Interitem connectivity is the number of items on a list that are associated to any degree. Future research with situation lists could be conducted in which this variable is manipulated between lists, while holding gist strength and backward associative strength constant, in order to see whether higher false-recall rates might be obtained. However, in general, a set of entities, things, and/or actions that typically co-occur in a particular type of situation should be associated, at least in the general sense of association.

The present experiments provide insight into at least some conditions that lead to gist extraction. Fuzzy-trace theory proposes that an encoding experience results in a verbatim trace and a gist trace. The verbatim trace represents the surface form of the actual episodic experience, and the gist trace represents the semantic content that results from encoding the surface form of the experience (Brainerd & Reyna, 2002). However, some have argued that the theory is unclear on how the semantic content of an episodic experience is constructed in the gist trace. One possibility is based on Barsalou's (1999) perceptual symbol system framework. We speculate that gist extraction might proceed in the following manner, at least in the case of situation lists. The presentation of list words produces the encoding of a verbatim trace, which through conscious or nonconscious awareness activates simulations for each list item on the basis of situational knowledge and other semantic relations. That is, conceptual processing is situation based, and situational simulations are intrinsic to it. Further, because gist processing occurs in parallel with verbatim processing, gist extraction results from integrating the individual simulations that are active at the time of encoding into a background situation, and this background situation represents the gist of the encoded episodic experience.

It also should be noted that although we have emphasized situational factors (as well as featural overlap in the form of synonyms, and taxonomic relations), false memories in experiments using the DRM procedure can occur in the absence of semantic information. For example, Zeelenberg, Boot, and Pecher (2005) found that high rates of false recognition result even when using nonwords that overlap orthographically and phonologically. Participants falsely recognized nonwords such as *ploost* after studying lists that contained items such as *froost*, *floost*, and *stoost*. Therefore, although semantic relations play a key role in false memories, other relations can matter as well.

Finally, we are not arguing that associative strength is unimportant in false-memory production because clearly in the "real world" some concepts are more likely to co-occur with other concepts in certain situations (e.g., doctor and nurse). Co-occurrence probabilities influence processing at many levels, and, as such, they influence, for example, the probability of constructing a situated simulation. Our point is that associative strength is not necessary nor by itself sufficient to explain false-memory effects in the DRM paradigm, nor in the real world. Rather, what is more important are certain types of semantic relations and the relations among them that are extracted from experiences with the world and then later used to reinstate that experience through simulation. From this perspective, research using the DRM paradigm is applicable to understanding real-world false memories based on the gist overlap that produces these illusory memories.

Acknowledgments

This work was supported by a SSHRC (Social Sciences and Humanities Research Council) Graduate Research Fellowship to David R. Cann while a graduate student at The University of Western Ontario, by Natural Sciences and Engineering Council Grant 06P007040 to Albert N. Katz, and by Natural Sciences and Engineering Council Grant OGP0155704 and National Institutes of Health Grant HD053136 to Ken McRae. We thank Charles Brainerd, Diane Pecher, and an anonymous reviewer for their helpful comments on an earlier version of this article.

References

Alba JW, Hasher L. Is memory schematic? Psychological Bulletin. 1983; 93:203–231.

Anisfeld M, Knapp M. Association, synonymity, and directionality in false recognition. Journal of Experimental Psychology. 1968; 77:171–179. [PubMed: 5655107]

Barsalou LW. Perceptual symbol systems. Behavioral and Brain Sciences. 1999; 22:577–609. [PubMed: 11301525]

Bartlett, FC. Remembering: A study in experimental and social psychology. Cambridge, MA: Cambridge University Press; 1932.

Bradley, MM.; Lang, PJ. Affective norms for English words (ANEW): Instruction manual and affective ratings (Tech. Rep. C-1). Gainesville, FL: University of Florida, Center for Research in Psychophysiology; 1999.

Brainerd CJ, Kingma J. Do children have to remember to reason? A fuzzy-trace theory of transitivity development. Developmental Review. 1984; 4:311–377.

Brainerd CJ, Reyna VF. Gist is the grist: Fuzzy-trace theory and the new intuitionism. Developmental Review. 1990; 10:3–47.

Brainerd CJ, Reyna VF. Memory independence and memory interference in cognitive development. Psychological Review. 1993; 100:42–67. [PubMed: 8426881]

Brainerd CJ, Reyna VF. Autosuggestibility in memory development. Cognitive Psychology. 1995; 28:65–101. [PubMed: 7895469]

Brainerd CJ, Reyna VF. When things that were never experienced are easier to “remember” than things that were. Psychological Science. 1998; 9:484–489.

Brainerd CJ, Reyna VF. Fuzzy-trace theory and false memory. Current Directions in Psychological Science. 2002; 11:164–169.

Brainerd, CJ.; Reyna, VF. The science of false memory. New York, NY: Oxford University Press; 2005.

Brainerd CJ, Reyna VF, Wright R, Mojardin AH. Recollection rejection: False-memory editing in children and adults. Psychological Review. 2003; 110:762–784. [PubMed: 14599242]

Brainerd CJ, Wright R, Reyna VF, Mojardin AH. Conjoint recognition and phantom recollection. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2001; 27:307–327.

Brainerd CJ, Yang Y, Reyna VF, Howe ML, Mills BA. Semantic processing in “associative” false memory. Psychonomic Bulletin & Review. 2008; 15:1035–1053. [PubMed: 19001566]

Brewer WF, Treyens JC. Role of schemata in memory for places. Cognitive Psychology. 1981; 13:207–230.

Collins AM, Loftus EF. A spreading-activation theory of semantic memory. Psychological Review. 1975; 82:407–428.

Cree GS, McRae K. Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). Journal of Experimental Psychology: General. 2003; 132:163–201. [PubMed: 12825636]

Deese J. Influence of interitem associative strength upon immediate free recall. Psychological Reports. 1959a; 5:235–241.

Deese J. On the prediction of occurrence of particular verbal intrusions in immediate recall. Journal of Experimental Psychology. 1959b; 58:17–22. [PubMed: 13664879]

Fillenbaum S. Words as feature complexes: False recognition of antonyms and synonyms. Journal of Experimental Psychology. 1969; 82:400–402.

Gallo, DA. Associative illusions of memory. New York, NY: Psychology Press; 2006.

Gallo DA, Roediger HL. Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. Journal of Memory and Language. 2002; 47:469–497.

Grossman L, Eagle M. Synonymity, antonymity, and association in false recognition responses. Journal of Experimental Psychology. 1970; 83:244–248.

Hare M, Jones M, Thomson C, Kelly S, McRae K. Activating event knowledge. Cognition. 2009; 111:151–167. [PubMed: 19298961]

Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. Psychological Bulletin. 1993; 114:3–28. [PubMed: 8346328]

Kucera, H.; Francis, WN. A computational analysis of present-day American English. Providence, RI: Brown University Press; 1967.

McDermott KB. Priming on perceptual implicit memory tests can be achieved through presentation of associates. Psychonomic Bulletin & Review. 1997; 4:582–586.

McEvoy CL, Nelson DL, Komatsu T. What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25(5):1177–1194.

McRae, K.; Cree, GS. Factors underlying category-specific semantic deficits. In: Forde, EME.; Humphreys, GW., editors. Category specificity in brain and mind. Hove, UK: Psychology Press; 2002. p. 211-249.

McRae K, Cree GS, Seidenberg MS, McNorgan C. Semantic feature production norms for a large set of living and nonliving things. Behavior Research Methods. 2005; 37:547–559. [PubMed: 16629288]

Meade ML, Watson JM, Balota DA, Roediger HL. The roles of spreading activation and retrieval mode in producing false recognition in the DRM paradigm. Journal of Memory and Language. 2007; 56:305–320.

Nelson DL, McEvoy CL, Dennis S. What is free association and what does it measure? Memory & Cognition. 2000; 28:887–899.

Nelson, DL.; McEvoy, CL.; Schreiber, TA. Unpublished manuscript. University of South Florida; Tampa: 1999. The University of South Florida word association, rhyme, and word fragment norms.

Payne DG, Elie CJ, Blackwell JM, Neuschatz JS. Memory illusions: Recalling, recognizing, and recollecting events that never occurred. Journal of Memory and Language. 1996; 35:261–285.

Reyna VF, Brainerd CJ. Fuzzy processing in transitivity development. Annals of Operations Research. 1990; 23:37–63.

Reyna VF, Brainerd CJ. Fuzzy-trace theory: An interim synthesis. Learning and Individual Differences. 1995; 7:1–75.

Reyna VF, Kiernan B. Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. Developmental Psychology. 1994; 30:178–191.

Reyna VF, Kiernan B. Children's memory and metaphorical interpretation. Metaphor and Symbolic Activity. 1995; 10:309–331.

Roediger, HL.; Balota, DA.; Watson, JM. Spreading activation and arousal of false memories. In: Roediger, HL.; Nairne, JS.; Neath, I.; Surprenant, AM., editors. The nature of remembering: Essays in honor of Robert G Crowder. Washington, DC: American Psychological Association; 2001. p. 95-115.

Roediger HL, McDermott KB. Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory, & Cognition. 1995; 21:803–814.

Roediger HL, McDermott KB. Tricks of memory. Current Directions in Psychological Science. 2000; 9:123–127.

Roediger, HL.; McDermott, KB.; Robinson, KJ. The role of associative processes in creating false memories. In: Conway, MA.; Gathercole, SE.; Cornoldi, C., editors. Theories of memory. Vol. 2. Hove, UK: Psychology Press; 1998. p. 187-245.

Roediger HL, Watson JM, McDermott KB, Gallo DA. Factors that determine false recall: A multiple regression analysis. Psychonomic Bulletin & Review. 2001; 8:385–407. [PubMed: 11700893]

Russell, WA.; Jenkins, JJ. Tech. Rep. No.11, Contract N8 ONR 66216 Office of Naval Research. Minneapolis: University of Minnesota; 1954. The complete Minnesota norms for responses to 100 words from the Kent–Rosanoff Word Association Test.

Seamon JG, Luo CR, Gallo DA. Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. Psychological Science. 1998; 9:20–26.

Stadler MA, Roediger HL, McDermott KB. Norms for word lists that create false memories. Memory & Cognition. 1999; 27:494–500.

Tabachnick, BG.; Fidell, LS. Using multivariate statistics. 4. Needham Heights, MA: Allyn & Bacon; 2001.

Toglia, MP.; Battig, WF. Handbook of semantic word norms. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1978.

Toglia MP, Neuschatz JS, Goodwin KA. Recall accuracy and illusory memories: When more is less. Memory. 1999; 7:233–256. [PubMed: 10645381]

Tse CS, Neely JH. Assessing activation without source monitoring in the DRM paradigm. Journal of Memory and Language. 2005; 53:532–550.

Tulving E. Memory and consciousness. Canadian Psychologist. 1985; 26:1–12.

Underwood BJ. False recognition produced by implicit verbal responses. Journal of Experimental Psychology. 1965; 70:122–129. [PubMed: 14315122]

WordNet. WordNet: A Lexical Database for English (Version 1.7.1). Princeton University; NJ: 2001. Retrieved from http://wordnetcode.princeton.edu/1.7.1/

Wu LL, Barsalou LW. Perceptual simulation in conceptual combination: Evidence from property generation. Acta Psychologica. 2009; 132:173–189. [PubMed: 19298949]

Zeelenberg R, Boot I, Pecher D. Activating the critical lure during study is unnecessary for false recognition. Consciousness & Cognition. 2005; 14:316–326. [PubMed: 15950885]

Zeelenberg R, Pecher D. False memories and lexical decision: Even twelve primes do not cause long-term semantic priming. Acta Psychologica. 2002; 109:269–284. [PubMed: 11881903]

APPENDIX A. Slightly modified version of the Wu and Barsalou (2009) feature taxonomy

The critical nonpresented word is in *italics*, and the list word is in <angled brackets>.

The major knowledge types are in **bold**.

Synonym: *thief* <pcrook> ; *trash* <garbage>

Antonym: *cold* <hot> ; *long* <short>

Taxonomic relations

Superordinate: *bread* <food> ; *butterfly* <insect>

Subordinate: *fruit* <apple> ; *pen* <fountain>

Individual: *king* <George> ; *river* <Mississippi>

Coordinate: *pen* <crayon> ; *car* <truck>

Entity features

External component: A three-dimensional component of an entity that, at least to some extent, normally resides on its surface. *chair* <legs> ; *shirt* <button>

External surface feature: An external feature of an entity that is not a component, and that is perceived on or beyond the entity's surface, including shape, colour, pattern, texture, size, touch, smell, taste. *carpet* <red> ; *mountain* <steep>

Internal surface feature: An internal feature of an entity that is not normally perceived on the entity's exterior surface, and that is only perceived when the entity's interior surface is exposed. *fruit* <juice>

Entity behaviour: An intrinsic action that is characteristic of an entity's behaviour, and that is not an entity's normal function for an external agent. *thief* <steal> ; *lion* <roar>

Entity made-of: A specification of the materials of which the entity is made. *chair* <wood> ; *bread* <flour>

Quantity: A numerosity, frequency, or intensity of an entity or its features. *trash* <pile>

Associated abstract entity: An abstract entity associated with the target entity and external to it. *flag* <freedom>

Systemic feature: A global systemic feature of an entity or its parts, including states, conditions, abilities, traits. *lion* <fierce> ; *girl* <young>

Larger whole: A whole to which an entity belongs. *lion* <pride> ; *citizen* <United States>

Spatial relation: A spatial relation between two or more properties within an entity, or between an entity and one of its properties. *spider* <poison>

Situation features

Function: A typical role that an entity serves for an agent. *needle* <injection> ; *pen* <write>

Action: An action that a participant performs in a situation. *citizen* <vote> ; *foot* <walk>

Participant: A person in a situation who typically uses an entity or performs an action on it and/or interacts with other participants. *justice* <jury> ; *mountain* <climber>

Location: A place where an entity can be found, or where people engage in an event or activity. *fruit* <basket> ; *car* <garage>

Origin: How or where an entity originated. *bread* <dough> ; *smoke* <fire>

Time: A time period associated with a situation or with one of its features. *black* <night> ; *butterfly* <summer>

Manner: The manner in which action or behaviour is performed. *whistle* <blow> ; *smell* <sniff>

Associated entity: An entity in a situation that contains the target concept. *city* <streets> ; *cup* <saucer>

Spatial relation: A spatial relation between two or more things in a situation. *thief* <gun> ; *music* <sound>

State of the world: State of a situation or any of its components except entities. *city* <crowded> ; *justice* <truth>

Introspective features

Affect/emotion: An affective or emotional state toward the situation or one of its components by either the subject or the participant. *needle* <hurt> ; *spider* <fright>

Evaluation: A positive or negative evaluation of a situation or one of its components by either the subject or a participant. *smoke* <stink> ; *girl* <beautiful>

Contingency: A contingency between two or more aspects of a situation, including if, enable, cause, because, depends, requires. *trouble* <help>

Representational state: Representational state in the mind of a situational participant, including beliefs, goals, ideas, etc. *man* <mouse> ; *health* <happiness>

Quantity: A numerosity, frequency, or intensity of an introspection or one of its properties. *anger* <rage>

Negation: An explicit mention of the absence of something, with absence requiring a mental state that represents the opposite. *health* <sickness>

Table 1

Summary of multiple regression analysis using knowledge types to predict BAS in Experiment 1

Independent variable	Partial r	β	t value	Significance
Synonyms	.27	.318	1.98	.053
Antonyms	.30	.338	2.43	.019
Taxonomic relations	.34	.505	2.78	.008
Entity features	.20	.289	1.65	.106
Situation features	.40	.581	3.25	.002

Note: BAS = backward associative strength.

Table 2

Summary of multiple regression analysis using knowledge types and BAS to predict false recall in Experiment 1

Independent variable	Partial r	β	t value	Significance
BAS	.71	.75	7.02	.000
Synonyms	.15	.14	1.08	.29
Antonyms	2.04	2.03	20.31	.76
Taxonomic relations	2.16	2.17	21.15	.26
Entity features	.07	.07	0.50	.62
Situation features	2.03	2.03	20.22	.83

Note: BAS = backward associative strength.

Table 3

Summary of multiple regression analysis using weighted knowledge types to predict false recall in Experiment 1

Independent variable	Partial r	β	t value	Significance
Synonyms	.41	.35	3.11	.003
Antonyms	.17	.14	1.22	.23
Taxonomic relations	.39	.30	2.89	.006
Entity features	.39	.32	2.97	.005
Situation features	.31	.26	2.28	.027
Introspective features	.28	.21	2.00	.051

Table 4

Means and standard deviations for the weighted knowledge types as a function of intrusion rate in Experiment 1

	Knowledge types weighted by BAS											
	Synonyms		Antonyms		Taxonomic relations		Entity features		Situation features		Introspective features	
Intrusion rate	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Low	.07	.15	.05	.18	.08	.15	.04	.12	.14	.33	.05	.11
Medium	.18	.34	.16	.33	.86	.91	.32	.43	.50	.62	.05	.09
High	.58	.74	.29	.44	.65	.90	.27	.50	1.04	.99	.17	.38

Note: BAS = backward associative strength.

Table 5

Mean probability of critical-word response for the 12 situation lists and the 6 DRM lists in Experiment 2

List type	List	Mean	SD
Situation feature			
	Accident	.53	.51
	Airport	.51	.51
	Arrest	.05	.21
	Bathroom	.56	.51
	Breakfast	.81	.39
	Casino	.86	.35
	Concert	.84	.37
	Farm	.70	.47
	Funeral	.72	.45
	Gym	.72	.45
	Theatre	.60	.50
	War	.77	.43
DRM			
Low false recall	Fruit	.65	.48
	King	.05	.21
	Lion	.31	.47
High false recall	Sleep	.58	.50
	Smell	.37	.49
	Window	.37	.49

Note: DRM = Deese–Roediger–McDermott.

Table 6

Mean proportion of falsely recalling the nonpresented critical word and mean number of gist-consistent intrusions for the 12 situation lists in Experiment 3

	Critical word		Gist-consistent	
Gist list	**Mean**	**SD**	**Mean**	**SD**
Accident	.03	.18	.15	.40
Airport	.05	.22	.23	.46
Arrest	.02	.13	.33	.70
Bathroom	.03	.18	.16	.42
Breakfast	.03	.18	.03	.18
Casino	.11	.32	.11	.37
Concert	.02	.13	.25	.47
Farm	.11	.32	.13	.34
Funeral	.03	.18	.23	.46
Gym	.02	.13	.10	.30
Theatre	.02	.13	.20	.44
War	.08	.28	.13	.39

Table 7

Proportions of veridical and false recall and recognition in Experiment 4.

	Veridical recall		Veridical recognition		False recall		False recognition	
List type	*Mean*	SD	*Mean*	SD	*Mean*	SD	*Mean*	SD
Situation	.67	.11	.86	.14	.15	.24	.65	.31
Strong	.62	.07	.79	.14	.52	.30	.75	.30
Mid-gist	.63	.09	.74	.15	.05	.14	.36	.34
Low-gist	.52	.10	.70	.16	.09	.19	.30	.27

Table 8

Remember/know proportions for veridical and false recognition in Experiment 4

	Veridical remember		Veridical know		False remember		False know	
List type	*Mean*	SD	*Mean*	SD	*Mean*	SD	*Mean*	SD
Situation	.37	.36	.49	.38	.30	.31	.34	.30
Strong	.34	.30	.45	.31	.36	.33	.39	.34
Mid-gist	.35	.27	.39	.26	.17	.26	.19	.27
Low-gist	.31	.25	.39	.25	.17	.26	.13	.23

Q J Exp Psychol (Hove). Author manuscript; available in PMC 2011 November 30.

APPENDIX B

Number of each knowledge type in the DRM lists

List	Synonyms	Antonyms	Taxonomic relations	Entity features	Situation features	Introspective features	Miscellaneous
anger	3	1	5	0	2	3	0
army	0	0	5	5	5	0	0
beautiful	3	1	1	0	9	1	0
bitter	1	0	12	0	0	1	1
black	1	1	6	1	2	3	0
bread	0	0	2	4	9	0	0
butterfly	0	0	4	4	5	1	1
cabbage	0	0	5	4	6	0	0
Car	1	0	10	0	4	0	0
carpet	1	0	0	6	6	1	1
chair	0	0	9	4	2	0	0
citizen	0	2	9	2	2	0	0
City	1	0	6	3	4	1	0
Cold	1	3	0	0	11	0	0
command	2	0	1	0	10	2	0
cottage	0	0	5	8	1	1	0
Cup	0	0	4	2	7	0	0
doctor	1	0	4	0	10	0	0
Flag	1	0	4	6	3	1	0
Foot	0	0	4	3	7	1	0
Fruit	0	0	9	2	4	0	0
Girl	1	1	4	2	4	3	0
health	0	0	0	0	5	10	0
High	0	1	0	0	10	4	0
justice	0	0	0	0	11	3	1
King	0	1	5	4	4	1	0
Lamp	0	0	0	9	5	1	0
Lion	0	0	4	4	7	0	0
Long	1	1	2	0	0	10	1

List	Synonyms	Antonyms	Taxonomic relations	Entity features	Situation features	Introspective features	Miscellaneous
Man	1	2	6	2	1	3	0
mountain	0	0	0	6	9	0	0
music	0	0	2	3	10	0	0
mutton	0	0	7	5	2	0	1
needle	0	0	4	3	6	2	0
Pen	0	0	8	2	4	0	1
River	1	0	4	5	5	0	0
rough	5	1	1	0	6	2	0
rubber	0	0	1	13	1	0	0
Shirt	0	0	5	6	4	0	0
Sleep	5	2	0	0	7	1	0
Slow	1	2	7	0	3	2	0
smell	2	0	2	0	8	3	0
smoke	0	0	0	0	12	1	0
Soft	3	2	2	0	7	1	0
spider	1	0	3	5	2	4	0
stove	1	0	4	7	3	0	0
sweet	0	2	7	0	1	4	0
swift	1	1	2	0	6	4	0
Thief	4	0	2	1	7	1	0
Trash	7	0	2	1	5	0	0
trouble	3	0	0	0	5	6	1
whiskey	0	0	10	0	4	1	0
whistle	0	0	1	4	10	0	0
window	0	0	0	10	4	0	0
Wish	3	0	0	0	0	8	1
Total	56	24	200	146	287	91	9

Note: DRM = Deese–Roediger–McDermott.

Q J Exp Psychol (Hove). Author manuscript; available in PMC 2011 November 30.

APPENDIX C

Situation lists used in experiments 2, 3, and 4

Critical item	List item	BAS	Critical item	List item	BAS
Airport	chairs	0	Accident	ambulance	.01
	luggage	.02		blood	.001
	passports	.01		cars	0
	pilots	0		crash	.13
	planes	0		damage	.05
	restaurants	0		fireman	0
	runway	.03		glass	0
	security	0		injuries	.04
	taxis	.01		paramedic	.01
	terminal	.34		police	0
	tickets	0		road	0
	travelers	.01		victims	0
	Mean BAS	.03		Mean BAS	.02
Arrest	bail	.05	Breakfast	bacon	.05
	charges	.01		cereal	.03
	court	0		food	0
	crime	0		coffee	0
	detective	0		eggs	.04
	evidence	0		fruit	0
	handcuffs	.07		juice	0
	jail	0		milk	0
	lawyer	0		pancakes	.12
	sheriff	0		plate	0
	suspect	.02		muffin	.11
	warrant	.45		toast	.07
	Mean BAS	.05		Mean BAS	.03
Bathroom	bath	.04	Casino	alcohol	0
	mirror	0		blackjack	.08
	shampoo	0		cards	0

Q J Exp Psychol (Hove). Author manuscript; available in PMC 2011 November 30.

Critical item	List item	BAS	Critical item	List item	BAS
	shower	0		chips	.02
	sink	.01		craps	.18
	soap	0		dealer	.08
	toilets	.34		gambler	.14
	toothbrush	.01		Las Vegas	.17
	towel	.02		lights	0
	tub	.01		money	0
	urinal	.15		slots	.41
	water	0		tables	0
	Mean BAS	.05		Mean BAS	.09
Concert	band	0	Farm	barn	.12
	fans	.02		cow	.06
	lighters	.01		fence	0
	microphone	.04		field	.04
	music	0		hay	.06
	pyrotechnics	.01		house	0
	roadies	.12		mud	0
	rockstar	.06		pig	.06
	smoke	0		pitchfork	.05
	speaker	.01		rooster	.02
	stage	.03		stables	.02
	ticket	.12		tractor	.29
	Mean BAS	.03		Mean BAS	.06
Funeral	cemetery	.03	Gym	athletes	.02
	coffin	.05		basketball	0
	death	.02		bike	0
	dirt	0		exercise	.20
	eulogy	.26		locker	.15
	flowers	0		mats	.01
	grave	.01		stairmaster	.06
	gravestone	.01		sweat	0

Q J Exp Psychol (Hove). Author manuscript; available in PMC 2011 November 30.

Critical item	List item	BAS	Critical item	List item	BAS
	hearse	.23		trainer	.19
	pallbearers	.38		treadmill	.04
	priest	0		running	.01
	undertaker	.06		weights	.05
	Mean BAS	.09		Mean BAS	.06
Theatre	actors	.01	War	army	.04
	audience	.04		artillery	.26
	costumes	.01		bombs	.12
	curtains	0		countries	.01
	darkness	0		guns	0
	director	.01		missiles	.02
	movies	.16		opponent	0
	popcorn	.02		peace	.37
	projector	0		soldier	.20
	screen	.02		tanks	.17
	seats	.05		trenches	.08
	ushers	.14		weapons	.05
	Mean BAS	.04		Mean BAS	.11

Note: BAS = backward associative strength.

Q J Exp Psychol (Hove). Author manuscript; available in PMC 2011 November 30.