Journal of Memory and Language 62 (2010) 19–34

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Individual differences in false recall: A latent variable analysis

Nash Unsworth *, Gene A. Brewer

Department of Psychology, University of Georgia, Athens, GA 30602-3013, United States

ARTICLE INFO

Article history:
Received 4 May 2009
revision received 18 August 2009
Available online 27 September 2009

Keywords:
False recall
Individual differences

ABSTRACT

The relation between intrusions in several different recall tasks was examined in the current study. Intrusions from these tasks were moderately correlated and formed a unitary intrusion factor. This factor was related to other cognitive ability measures including working memory capacity, judgments of recency, and general source-monitoring ability. Individual differences in source-monitoring ability fully mediated the relation between working memory capacity, judgments of recency, and intrusions in recall. Theoretically, individual differences in false recall may result, in part, from differences in preretrieval and post-retrieval source-monitoring processes in addition to lure activation. Future models of false recall should integrate theses source-monitoring mechanisms to fully account for various intrusions that occur in standard recall tasks.

Introduction

The ability to accurately recall prior experiences has long interested memory and individual differences researchers alike. Students of memory have extensively studied the conditions that lead to accurate recall of lists of words including factors such as retention interval, interpolated distractor activity, list-length, word frequency, presentation duration, and presence or absence of retrieval cues. At the same time, individual differences researchers have studied the extent to which differences in recall abilities are related to performance on other memory tasks as well as to broader intellectual functioning (Carroll, 1993; Underwood, Boruch, & Malmi, 1978). More recently, researchers have become interested in inaccurate or false memories (see Gallo, 2006 for a review). In particular, researchers have become interested in situations in which individuals falsely recall information that was not presented. These intrusion errors have been shown to be highly systematic and diagnostic of a number of neurological disturbances including Alzheimer's Disease, Parkinson's Disease, frontal lobe damage (e.g., Balota et al., 1999; Helkala, Laulumaa, Soininen, & Riekkinen, 1989; Rouleau, Imbault, Laframboise, & Bédard, 2001), as well as differences in healthy aging (e.g., Balota et al., 1999; Kahana, Dolan, Sauder, & Wingfield, 2005; Lövdén, 2003; Norman & Schacter, 1997). The aim of the current study was to examine different types of intrusions in free and cued recall, individual differences in false recall, and their relation to other cognitive constructs in a young, healthy sample.

False recall

In most studies of free recall, intrusion errors (items not presented on the current list) are not analyzed given that they rarely occur. Yet when they are analyzed a number of interesting and systematic findings emerge. Intrusion errors can be broken down into two types: previous-list and extra-list intrusions. On the one hand, previous-list intrusions (PLIs) represent words that were not presented on the current list that participants are trying to remember, but were presented on previous lists. Extra-list intrusions (ELIs), on the other hand, represent words that were not presented on any of the lists. When examining these two intrusion types separately a number of systematic patterns are found. For instance, it is found that PLIs

* Corresponding author.
E-mail address: nunswor@uga.edu (N. Unsworth).

0749-596X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2009.08.002

predominantly come from the immediately preceding list and the recency gradient for PLIs tends to fall off monotonically for lists further back (Murdock, 1974; Unsworth & Engle, 2007; Zaromb et al., 2006). These PLIs also tend to come predominantly from primacy and recency positions on the lists they were presented on (Unsworth, 2008). This is likely due to the fact that PLIs typically are words that were initially recalled correctly on their respective lists. Very few unrecalled items appear as PLIs in later lists. Furthermore, ELIs tend to be either semantically or phonologically related to one of the target words in the current list (e.g., Craik, 1968; see also Watson, Balota, & Sergent-Marshall, 2001; Zaromb et al., 2006). Thus, information that was just presented tends to intrude more than information presented further back in time and information that is related to the current information also tends to intrude.

Research has also found that both types of intrusions tend to occur late in the recall period (Craik, 1968; Unsworth, 2008), with roughly 60% of both types of intrusions occurring at one of the last three output positions (Unsworth, 2008). Additionally, when an intrusion is recalled (either a PLI or an ELI) the next response tends to be another intrusion (Zaromb et al., 2006). Specifically, if a PLI is recalled, the next item recalled tends to be another PLI. If an ELI is recalled the next item tends to be another ELI (Zaromb et al., 2006). Finally, for both types of intrusions, participants are generally quite accurate in overtly editing and monitoring these responses which is one reason why they tend to be so rare (Kahana et al., 2005; Unsworth & Brewer, in press a). Thus, although intrusions (both PLIs and ELIs) do not occur often, it is clear that when they do occur they are associated with a number of systematic findings.

Perhaps the most commonly studied of these intrusions are ELIs that occur in lists of associated words (Deese, 1959; Roediger & McDermott, 1995). In the Deese–Roediger–McDermott (DRM) paradigm participants are presented with a list of words which are all related to a common word (e.g., sleep) and are asked to recall all of the words that were presented. Critically, the non-presented word (sleep) tends to be falsely recalled with a high probability. That is, a very particular ELI is recalled with a high probability due to the fact that it is related to all of the list items. Like other ELIs, research suggests that these errors tend to occur late in the recall period (Roediger & McDermott, 1995), and arise due to both semantic and phonological associations (Watson et al., 2001). Furthermore, these ELIs tend to be associated with lower confidence ratings than words actually presented on the lists (Jou, 2008).

A number of factors have been shown to either increase or decrease the rate of false recalls in the DRM paradigm. For instance, the rate of false recall tends to increase with fast presentation durations, but decreases with slower presentation durations (McDermott & Watson, 2001). Furthermore, Hicks and Marsh (1999) found that presenting information from multiple sources led to a reduction in false recalls as did the knowledge of a subsequent source monitoring test. Warning participants prior to the presentation of the lists has been shown to reduce false recalls (Watson, Bunting, Poole, & Conway, 2005) as has giving participants multiple study-test trials (McDermott, 1996). The combination of warnings and multiple study-test trials has been shown to lead to almost complete elimination of false recalls in younger adults (Watson, McDermott, & Balota, 2004). Thus, it is clear that false recalls in the DRM paradigm are highly systematic (similar to PLIs and ELIs more generally) and these false recalls can be reduced and even eliminated with various manipulations.

In order to explain ELIs in the DRM paradigm, Roediger, McDermott, and colleagues (Roediger & McDermott, 1995; Roediger, Watson, McDermott, & Gallo, 2001) have proposed an activation-monitoring account of the data. In this theory it is assumed that false recalls arise from both encoding and retrieval processes that influence not only the activation of the intruding item, but also the monitoring processes that operate post-retrieval to determine if the generated item is in fact a correct item. It is assumed that during encoding and subsequent retrieval of the list items each item activates its representation in the lexical-semantic network and activation spreads throughout the network activating nearby associates. The stronger the association between the list items and the critical item (as measured by backward associative strength for example), the higher the probability that the critical item will become activated and subsequently falsely recalled (Roediger et al., 2001).

In addition to activation processes, monitoring processes are also thought to be important for the generation of false recalls in the DRM paradigm. Like activation processes, monitoring processes are also thought to operate at both encoding and retrieval. At encoding, monitoring processes can be used to guard against potential false recalls by ensuring that participants pay close attention to only words actually presented (i.e., via warnings; Watson et al., 2005) or by paying attention to the specific qualitative characteristics of the words which are presented (Hicks & Marsh, 1999). At retrieval, monitoring processes also work to ensure that the information recalled was in fact presented in the context of the experiment. That is, source-monitoring processes (Johnson, Hashtroudi, & Lindsay, 1993) are used after an item has been retrieved to check and make sure that there is sufficient overlap in contextual features between the recalled item and the experimental context (i.e., overlap in temporal–contextual features, as well as features specific to the experiment, such as the experimenter and the experimental room). If there is a sufficient degree of overlap then the word is deemed to be correct and actually output (i.e., recalled). If there is a low degree of overlap then the word is considered as incorrect and withheld (i.e., not recalled).

According to the activation-monitoring theory of the DRM paradigm, the critical item can be implicitly activated during both encoding and retrieval by either the presentation of strong associates or the recall of strong associates. Without sufficient warning during encoding, special attention will not be paid to certain contextual features of the presented items and thus, during retrieval, monitoring processes will be needed to correctly discriminate between items actually presented and items that are simply primed via their associative strength. Similar to generate-edit models of free recall, this account suggests that in order to falsely recall an item, first ELIs in the DRM paradigm must be generated (based on their backward association to the list items) and then monitoring and editing processes must fail to acknowledge the item as incorrect and instead allow it to be recalled. Thus, both processes are needed to account for false recalls.

Recently, Kimball, Smith, and Kahana (2007) have extended the search of associative memory (SAM) model (Raaijmakers & Shiffrin, 1980) to account for false recalls in a manner generally similar to the activation-monitoring account (i.e., fSAM). In the fSAM model it is assumed that items which are coactive in the short-term store tend to activate each others lexical-semantic representations at encoding. Thus, the stronger the association between items in the short-term store and the critical item, the higher the probability of activating the critical item will be at retrieval due to increased lure activation at encoding. Additionally, at retrieval, the most recently retrieved 3–4 items (including both corrects and intrusions) are used as retrieval cues to cue the next item. Similar to the activation-monitoring view, the critical item can become activated based on associations during the presentation of list items, during the retrieval of the list items, or both. Importantly, the fSAM model provides a detailed quantitative account of false recalls not only in the DRM paradigm, but also ELIs and PLIs more broadly. That is, the model is able to simulate the high number of critical ELIs in the DRM paradigm as well as the lower number of other ELIs and PLIs in more traditional recall tasks. This occurs because fSAM incudes not only semantic associations between items, but it also includes contextual associations between items in each list and the state of context at the time of test. By assuming that the association between test context and list items decreases with recency, the model predicts the occurrence of PLIs and the fact that PLIs tend to come from the most recently presented lists (Sirotin, Kimball, & Kahana, 2005). Thus, the various types of intrusions seem to arise from similar mechanisms.

One difference between the activation-monitoring view of false recall and the fSAM model is the fact that the activation-monitoring view explicitly suggests that monitoring processes are important, whereas the fSAM model does not currently include any post-retrieval processes. Although as pointed out by Kimball et al. (2007) such an implementation is possible (e.g., Sirotin et al., 2005). Thus, the fSAM model accounts for many of the findings in the DRM literature, but seems to lack an important component in terms of monitoring processes which are likely important in terms of not only experimental manipulations that lead to the incidence of false recalls but also individual differences in false recall.

Individual differences in false recall

In addition to basic memory studies on false recalls, a number of studies have been done to examine individual differences in false recalls. Most of these studies have been concerned with possible variation in the rate of false recalls in the DRM paradigm. For instance, Norman and Schacter (1997) found that older adults not only recalled fewer correct target items than younger adults, but they also recalled more critical items in the DRM paradigm than younger adults. Similarly, Watson et al. (2001) found that older adults were more likely to recall the critical item in the DRM paradigm than younger adults for lists composed of semantic associates, phonological associates, and for both phonologically and semantically associated lists. Furthermore, Watson et al. (2004) found that these age differences could be drastically attenuated when both younger and older adults were given prior warnings, multiple study-test trials, and slow presentation rates.

These results suggest that older adults are more likely to have critical ELIs in the DRM paradigm than younger adults, and these effects seem to be largely driven by differences in the strategic monitoring component in the activation-monitoring theory (Watson et al., 2004). Indeed, research has suggested that high ability older adults (high scorers on a putative battery of frontal-lobe sensitive tests) have near equivalent levels of correct and false recalls as younger adults (Butler, McDaniel, Dornburg, Price, & Roediger, 2004). Subsequent correlational research has suggested that age and ability (again based on a putative composite of frontal functioning) account for unique variance in false recalls and this occurs for both older and younger adults (Chan & McDermott, 2007). It is clear that there are age differences in false recall, and these age differences are likely due in part to differences in strategic monitoring abilities as well as other abilities which change over the lifespan.

Recent work by Lövdén (2003) has extended this work and examined the extent to which aging effects in false memories are mediated by other cognitive processes. Specifically, Lövdén (2003) had 146 adults (ages 20–80) perform a number of tasks thought to elicit false memories including category cued recall, DRM, and a picture recognition task. Using confirmatory factor analysis, Lövdén formed a common false memory factor and examined the relation between this factor and age as well as processing speed, inhibition, and episodic memory factors. Consistent with prior work, Lövdén found that false memory was strongly related to age and this effect was primarily mediated by age differences in episodic memory abilities, which were strongly related to inhibitory abilities. These results suggest that age differences in false memory are due in large part to age differences in basic episodic memory abilities. This could be due to the need to utilize accurate source-monitoring processes in both false memory and episodic memory paradigms as suggested by the activation-monitoring account (Roediger et al., 2001). The results of this study provide important initial evidence that false memories across paradigms are reliably related to one another and form a distinct factor (see also Blair, Lenton, & Hastie, 2002). Furthermore, this factor was reliably related to other important cognitive constructs thought to be related to the ability to accurately monitor and edit out false memories.

In addition to aging work in the DRM paradigm, other studies have also examined individual differences in false recalls. For instance, Winograd, Peluso, and Glover (1998) examined the correlation between false recall and false recognition in the DRM paradigm with a variety of cognitive and personality measures. Winograd et al. found that self-reports on the Dissociative Experiences Scale, the Vividness of Visual Imagery Questionnaire, and the Subjective Memory Questionnaire were all correlated with measures of false recognition in the DRM paradigm. Winograd et al. suggested that the relations found among the self-report measures and susceptibly to false memories arose due to failures in source monitoring. Those individuals who were likely to report a greater frequency of dissociative experiences had greater vivid mental imagery abilities and reported more everyday memory failures. Consistent with the activation-monitoring view of false recall and recognition (Roediger et al., 2001) this variation is likely due to problems in discriminating between events that actually occurred and internally generated events, with lower ability individuals having poorer source-monitoring abilities, which lead to greater susceptibility to false memories.

Further evidence consistent with this notion comes from work examining the relation between working memory capacity (WMC) and susceptibility to false memories in the DRM paradigm. For instance, Watson et al. (2005) had high and low WMC individuals (based on a quartile split on operation span performance) perform a standard DRM free recall task with either a prior warning or no prior warning in their Experiment 1. Watson et al. found that high and low WMC individuals produced the same number of critical word intrusions in the no warning condition, but that high WMC individuals recalled fewer critical word intrusions in the warning condition. Watson et al. suggested that high WMC individuals were using their superior goal maintenance abilities (Kane & Engle, 2002) to actively maintain the task goal (i.e., only recall presented words) and avoid the automatic activation of the false memories. In their second experiment high and low WMC individuals again performed the DRM task with warnings or no warnings, but this time they were also given five study-test trials. Consistent with their Experiment 1, high WMC individuals recalled fewer critical word intrusions than low WMC individuals, and this occurred only in the warning condition. Furthermore, Watson et al. found that multiple study-test trials reduced the rate of false recalls for both high and low WMC individuals. Watson et al. suggested that different mechanisms were responsible for the reduction in false recall for warning manipulations and study-test trial manipulations. From the activation-monitoring theory point of view, this would suggest that high and low WMC differences in false recall are primarily due to differences in activation whereby high WMC individuals are better at preventing or resisting the activation of the critical word intrusion than low WMC individuals. Given that both WMC groups were helped by the study-test manipulation, individual differences in source-monitoring abilities at retrieval do not seem likely. However, recent work by Gerrie and Garry (2007) has suggested that individual differences in WMC and false memory are in fact related to differences in source-monitoring abilities. Thus, it remains an open question as to whether WMC differences in false memories are due, in part, to differences in basic source-monitoring abilities.

In addition to work specifically examining individual differences in the DRM paradigm, recent work has been devoted to examining individual differences in intrusions (both PLIs and ELIs) more broadly. For instance, Kahana et al. (2005) found that older adults recalled more PLIs and ELIs than younger adults in a standard free recall task. In their second experiment, Kahana et al. used a variant of externalized free recall in which participants were instructed to recall all the target words and any other words that came to mind. In addition, in order to examine monitoring and editing abilities, participants were instructed to press a key for each word that they recalled but knew was incorrect. Kahana et al. found that not only were older adults more likely to recall intrusions than younger adults, but they were also less likely to correctly reject those intrusions. This result suggests that older adults susceptibly to intruding information is likely due, in part, to differences in their source-monitoring abilities.

Recently, we (Unsworth & Brewer, in press a) have found similar results with high and low WMC individuals using the same externalized free recall task as Kahana et al. (2005). Specifically, we found that low WMC individuals were more likely to recall both PLIs and ELIs than high WMC individuals in standard delayed free recall and low WMC individuals were poorer at monitoring these intrusions than high WMC individuals in the externalized free recall task (see also Unsworth, 2007). Similar effects have also recently been demonstrated in cued recall with low WMC individuals recalling more intrusions than high WMC individuals (Rosen & Engle, 1998; Unsworth, 2009a).

Finally, in a recent latent variable analysis we found that intrusions across multiple free recall tasks were related and formed a single intrusion factor (Unsworth, 2009b). This factor was negatively related to overall recall performance, positively related to recall latency, and negatively related to both WMC and fluid intelligence. We have suggested that individual differences in WMC and false recalls were partially due to differences in monitoring and editing abilities, in which low WMC individuals were poorer at discriminating between correct target items and intruding related items. Collectively these results suggest that there is systematic variation in intrusions (PLIs, ELIs, and DRM critical word intrusions) and these intrusions are related to individual differences in a number of areas.

The present study

The aim of the present study was to examine individual differences in false recalls from a latent variable perspective. In particular, two main questions were addressed. First, to what extent are different types of intrusions and intrusions across multiple tasks related? Specifically, do PLIs and ELIs represent the same underlying construct or are there sufficient differences between the two types of errors to suggest that they are different. Intuitively one might suspect that because both are intrusions that it is only natural for a single factor to account for both. However, some work suggests that PLIs and ELIs are different. Specifically, as noted previously PLIs by their very nature reflect intruding items from recently presented lists (i.e., proactive interference) whereas ELIs tend to be phonologically or semantically related to one of the target items. Furthermore, we have found that some manipulations tend to selectively influence one intrusion type but not the other (e.g., Unsworth, 2009a; Unsworth & Brewer, in press a). Thus, it is possible that the two error types are in fact different.

Additionally, given the large amount of work that has been done examining ELIs in the DRM paradigm, one might question whether these intrusions are unique to that particular experimental paradigm or whether they are simply a specific type of ELI. Finding that DRM critical intrusions are the same or different as other types of intrusions would have implications not only for work on the DRM paradigm, but also on the extent to which theories of the DRM paradigm can generalize to explain other types of intrusions more broadly. Clearly, given the very specific nature of the DRM paradigm and the very specific nature of these ELIs one might expect some slight differences, but overall it seems likely that DRM critical intrusions are simply another manifestation of ELIs and intrusions overall.

Finally, we can ask to what extent intrusions across multiple tasks are related. Previous research has suggested that intrusions in various free recall tasks are related and form a distinct factor (Unsworth, 2009b), but it is not known how these intrusions in standard delayed free recall are related to intrusions in other recall tasks. That is, are non-critical word intrusions in the DRM the same as other intrusions in standard free recall tasks? Furthermore, are intrusions in standard delayed free recall the same as intrusions in externalized free recall? Previous research that has utilized externalized free recall to make inferences on the nature of intrusions (e.g., Kahana et al., 2005; Unsworth & Brewer, in press a) has implicitly assumed that they are the same. However, this has not been empirically demonstrated. Likewise, we can ask to what extent are intrusions in free and cued recall the same or different. We have found similar WMC differences in intrusions in free (Unsworth, 2007) and cued recall (Unsworth, 2009a) and have assumed that these reflect variation in largely similar processes, yet this has not been directly shown.

The second main question addressed in the current study was to what extent are intrusions (either as a unitary construct or separate constructs) related to other cognitive processes such as veridical recall, source monitoring, WMC, judgments of recency, and overall verbal knowledge? For instance, prior work has shown that veridical and false recalls are negatively related at both the list (Roediger et al., 2001) and the individual level (Lövdén, 2003). Furthermore, as noted throughout, source-monitoring abilities are generally seen as one primary factor that accounts for false recalls and individual variation in false recalls, yet there is no direct evidence that, at an individual level, source monitoring is related to false recall. Although experimental work has implicated source monitoring in false recall (Hicks & Marsh, 1999) and failures in source monitoring have long been considered a driving factor in individual differences in false recall (Watson et al., 2001) and false recall more broadly (Roediger & McDermott, 1995), there is no study to our knowledge that has shown a direct correlation between source monitoring and false recalls. If false recalls are due, in part, to source monitoring failures, we should find that individual differences in source-monitoring abilities are negatively related to false recalls.

Additionally, given that prior research has demonstrated a link between WMC and false recalls not only in the DRM paradigm (Watson et al., 2005) but in intrusions more generally (Unsworth, 2009b; Unsworth & Brewer, in press a) we should find that WMC abilities are also negatively related to intrusions. However, if the reason that WMC is related to false recall is because of differences in source-monitoring abilities (Unsworth, 2007; Unsworth & Brewer, in press a), then source monitoring should mediate the relation between WMC and intrusions. If, however, WMC is related to false recalls because of individual differences in the ability to resist automatic spreading activation (e.g., Watson et al., 2005), then source monitoring will not fully mediate the relation between WMC and intrusions.

Finally, the relation between false recalls and other cognitive constructs such as judgments of recency and overall verbal knowledge were examined to see if these processes had a direct effect on individual differences in false recall. Specifically, one might expect that judgments of recency should be related to false recalls to the extent that participants rely on recency information when deciding whether an item is from the most recently presented list or not. Participants who perform well on judgments of recency tasks and are able to accurately discriminate between two recent pieces of information should be better at discriminating between target and non-target items that have been generated during retrieval. Thus, judgments of recency can be seen as a specific instantiation of source-monitoring abilities discussed previously, but here they are restricted only to judgments based on the recency of items and not on overall contextual features. Likewise, overall verbal knowledge may be related to susceptibility to false recalls to the extent that overall verbal knowledge represents a proxy for connectivity within a participant's semantic network. Participants with higher verbal knowledge (based on vocabulary) might have stronger connectivity within the network leading to a higher probability of activating a non-target item (an ELI).

In order to address these questions, we relied on a latent variable analysis. This was done because previous results may have been found due to the fact that only a single task was used and thus, may not provide the best evidence for more general constructs. Furthermore, most individual differences studies that have been done typically examined only extreme groups of participants, and thus it is not clear whether the relation holds across a full range of participants. In order to derive latent variables for the constructs of interest, multiple indicators of each construct were used. Specifically, each participant performed six free and cued recall tasks including a standard delayed free recall task with unrelated words, a delayed free recall task in which words were semantically related on three consecutive trials and then the semantic category switched on the fourth trial, an externalized free recall task (Kahana et al., 2005; Unsworth & Brewer, in press a), the DRM paradigm (Roediger & McDermott, 1995), a cued recall task with word–word pairs, and cued recall task with number–word pairs. For each recall task correct recalls and intrusions (PLIs, ELIs, and DRM critical intrusions in the DRM task) were recorded. In addition to the recall tasks participants performed two source monitoring tasks, three WMC tasks, two judgments of recency tasks, and two vocabulary measures. Performance on these tasks was then used to build latent variables for the constructs of interest in order to address the two main questions put forth previously.

Method

Participants

A total of 177 participants were recruited from the subject-pool at the University of Georgia. Participants were between the ages of 18 and 35 and received course credit for their participation. Each participant was tested individually in two laboratory sessions lasting approximately 2 h each.

Materials and procedure

After signing informed consent, all participants completed operation span, symmetry span, reading span, delayed free recall with semantically related words, paired associates with word-word pairs, judgments of recency with words, and judgments of recency with pictures in Session 1. In Session 2, all participants completed delayed free recall with unrelated words, paired associates with number–word pairs, gender source recognition, picture source recognition, the DRM task, a version of externalized free recall, a synonym vocabulary test, and an antonym vocabulary test. All tasks were administered in the order listed above.

Tasks

Recall tasks

Delayed free recall unrelated words. Participants recalled six lists of 10 words each. All words were common nouns that were presented for 1 s each. After list presentation, participants engaged in a 16 s distractor task before recall: Participants saw 8 three-digit numbers appear for 2 s each, and were required to write the digits in ascending order. After the distractor task participants typed as many words as they could remember from the current list in any order they wished. Participants had 45 s for recall. Total number correct, number of PLIs, and number of ELIs were recorded.

Delayed free recall semantically related words. Participants recalled six lists of 10 words each broken down into two blocks (three lists per block). All words in each block came from the same semantic category. Following the last word in a list participants were required to count backwards by three's as quickly and accurately as possible from a three-digit number onscreen for 15 s and to write the numbers down as they go. After the distractor task participants typed as many words as they could remember from the current list in any order they wished. Participants had 45 s for recall. Total number correct, number of PLIs, and number of ELIs were recorded.

DRM. Participants recalled six lists of 15 words with each list being comprised of semantically associated words that were all related to a critical word. The lists were taken from Roediger and McDermott (1995) Experiment 1 with the six critical words being chair, mountain, needle, rough, sleep, and sweet. Following the presentation of the last word in each list participants saw ??? and were instructed to type the words from the most recent list in order they wished. Participants had 45 s for recall. Total number correct, number of PLIs, number of ELIs, and number of critical word intrusions was recorded.

Externalized free recall. The externalized free recall task was exactly the same as the delayed free recall with unrelated words task except that a separate set of nouns was used and instructions for the recall phase indicated that recall should be uninhibited. Specifically, participants were instructed to not only type all of the items from the most recent list as they could, but to also type any other words that came to mind during the recall phase even if they knew that the word was not presented on the most recent list. Furthermore, participants were instructed that if they typed a word that they knew was incorrect then they should press the spacebar to indicate that it was incorrect prior to recording their response. Total number correct, number of PLIs, and number of ELIs were recorded. Note, for both PLIs and ELIs only intrusions that were not associated with a spacebar press were counted as intrusions.

Paired associates with words. Participants were given three lists of 10 word pairs each. All words were common nouns and the word pairs were presented vertically for 2 s each. Participants were told that the cue would always be the word on top and the target would be on bottom. After the presentation of the last word participants saw the cue word and ??? in place of the target word. Participants were instructed to type in the target word from the current list that matched cue. Cues were randomly mixed so that the corresponding target words were not recalled in the same order as they were presented. Participants had 5 s to type in the corresponding word. Proportion correct, number of PLIs, and number of ELIs were recorded.

Paired associates with numbers. This task was exactly the same as the cued recall task with words, but instead of word-word pairs, a random three-digit number served as the cue that was paired with the target word. Proportion correct, number of PLIs, and number of ELIs were recorded.

Source monitoring tasks

Gender source recognition. Participants heard words (30 total words) in either a male or a female voice. Participants were explicitly instructed to pay attention to both the word as well as the voice the word was spoken in.[1] At test participants were presented with 30 old and 30 new words and were required to indicate if the word was new or old and, if old, what voice it was spoken in via key press. Participants had 5 s to press the appropriate key to enter their response. A participant's score was the proportion of correct responses.

Picture source recognition. Participants were presented with a picture (30 total pictures) in one of four different quadrants onscreen for 1 s. Participants were explicitly instructed to pay attention to both the picture as well as the quadrant it was located in. At test participants were presented with 30 old and 30 new pictures in the center of the screen. Participants indicated if the picture was new or old and, if old, what quadrant it was presented in via key press. Participants had 5 s to press the appropriate key to enter their response. A participant's score was the proportion of correct responses.

Working memory capacity tasks

Operation span (Ospan). Participants solved a series of math operations while trying to remember a set of unrelated letters. Participants were required to solve a math operation and after solving the operation they were presented with a letter for 1 s. Immediately after the letter was presented the next operation was presented. At recall, letters from the current set were recalled in the correct order by clicking on the appropriate letters. For all of the span measures, items were scored if the item was correct and in the correct position. The score was the number of correct items in the correct position.

Symmetry span (Symspan). Participants were required to recall sequences of red squares within a matrix while performing a symmetry-judgment task. In the symmetry-judgment task participants were shown an 8×8 matrix with some squares filled in black. Participants decided whether the design was symmetrical about its vertical axis. The pattern was symmetrical half of the time. Immediately after determining whether the pattern was symmetrical, participants were presented with a 4×4 matrix with one of the cells filled in red for 650 ms. At recall, participants recalled the sequence of red-square locations in the preceding displays, in the order they appeared by clicking on the cells of an empty matrix. The same scoring procedure as Ospan was used.

Reading span (Rspan). Participants were required to read sentences while trying to remember a set of unrelated letters. Participants read a sentence and determined whether the sentence made sense or not. Half of the sentences made sense while the other half did not. Nonsense sentences were made by simply changing one word from an otherwise normal sentence. After participants gave their response they were presented with a letter for 1 s. At recall, letters from the current set were recalled in the correct order by clicking on the appropriate letters. The same scoring procedure as Ospan was used.

Judgments of recency (JOR) tasks

JOR words. This task was modeled after Konishi et al. (2002). Participants were presented with three lists of ten words each. Each word was presented individually onscreen for 2 s each. At test, participants were given two words from the current list and were required to indicate which word had been presented more recently than the other via key press. Participants had 5 s to press the appropriate key to enter their response. A participant's score was proportion correct.

JOR pictures. This task was exactly the same as JOR words, but instead of words, pictures were presented. A participant's score was proportion correct.

Vocabulary tasks

Synonym vocabulary. In this task participants were given 10 vocabulary words and were required to select the best synonym (out of five possible choices) that best matched the target vocabulary word (Hambrick, Salthouse, & Meinz, 1999). Participants were given 2 min to complete the 10 items. A participant's score was the total number of items solved correctly.

Antonym vocabulary. In this task participants were given 10 vocabulary words and were required to select the best antonym (out of five possible choices) that best matched the target vocabulary word (Hambrick et al., 1999). Participants were given 2 min to complete the 10 items. A participant's score was the total number of items solved correctly.

Results

The results are divided into two primary sections. The first section examined the extent to which different types of intrusions (PLIs, ELIs, and DRM critical intrusions) were related to one another and related across tasks. The second section examined the extent to which intrusions were related to overall veridical recall as well as to other cognitive constructs such as source monitoring, WMC, judgments of recency, and vocabulary.

Relation among intrusions

The first primary set of analyses concerned the extent to which PLIs, ELIs, and DRM critical intrusions were re-

[1] *Note*: we specifically made all participants aware of the fact that a source test would be required because it is likely that after the first source test was given some participants would guess the nature of the subsequent source test. This would likely introduce idiosyncratic variance in terms of who guessed the nature of the test. Therefore, we decided to control for this by letting all participants know the nature of the test ahead of time. By letting participants know up front about the nature of the test it is likely that performance is largely driven by controlled processes operating at both encoding and retrieval. More work is clearly needed to examine the extent to which individual differences in controlled encoding and retrieval processes affect normal memory processes as well susceptibility to false memories.

Table 1
Correlations among the different intrusion types.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. DFRUP	1.00												
2. DFRUE	0.56	1.00											
3. DFRRP	0.35	0.40	1.00										
4. DFRRE	0.12	0.26	0.34	1.00									
5. DRMP	0.15	0.09	−0.05	−0.02	1.00								
6. DRME	0.38	0.40	0.29	0.23	0.05	1.00							
7. DRMC	0.24	0.35	0.18	0.18	0.10	0.31	1.00						
8. EFRP	0.51	0.46	0.26	0.11	0.18	0.36	0.25	1.00					
9. EFRE	0.40	0.39	0.15	0.06	0.08	0.32	0.12	0.40	1.00				
10. PAWP	0.21	0.32	0.18	0.13	0.00	0.10	0.16	0.05	0.16	1.00			
11. PAWE	0.21	0.51	0.20	0.29	0.04	0.23	0.15	0.24	0.32	0.30	1.00		
12. PANP	0.31	0.43	0.21	0.23	0.15	0.23	0.18	0.29	0.27	0.28	0.34	1.00	
13. PANE	0.33	0.46	0.23	0.14	−0.04	0.18	0.09	0.14	0.19	0.27	0.43	0.27	1.00

Note: DFRU = delayed free recall unrelated words; DFRR = delayed free recall related words; DRM = Deese–Roediger–McDermott paradigm; EFR = externalized free recall; PAW = paired associates with word–word pairs; PAN = paired associates with number–word pairs; P = previous-list intrusion; E = extra-list intrusion; C = critical word intrusion.

lated to one another and related across various recall (both free and cued) tasks. For each of the six recall tasks we computed the total number of PLIs and ELIs for all of the lists for each individual. In addition, in the DRM task we also computed the number of critical intrusions that an individual made for all of the lists (for a total possible of six). The correlations between all of the PLIs, ELIs, and DRM critical intrusions for all of the tasks are shown in Table 1. As can be seen most of the intrusions were positively related to one another, suggesting that individuals who were prone to one type of intrusion were also likely to make the other type of intrusion and this occurred for all of the tasks. Note that some of the intrusions demonstrated fairly weak correlations with the other intrusions, but this is likely due to the fact that in these situations very few participants made that type of intrusion. For instance, DRM PLIs were weakly related to all of the other types of intrusions, and this effect is likely due to the fact that given the nature of the lists used in the DRM task, it is unlikely that participants would recall an item associated with *chairs* when they were supposed to be recalling items related to *sleep*.[2] For the most part, however, the correlations suggest moderate relations among all of the intrusions.

Next, we utilized confirmatory factor analysis to better examine the relations among the different types of intrusions. Specifically, we tested whether PLIs and ELIs were best conceptualized as a single unitary factor, or whether there were sufficient differences between the two intrusions types to suggest two separate, yet correlated factors. To examine this we specified two models. Note for these initial models DRM critical intrusions were not included. In the first model we specified all of the intrusions (both PLIs and ELIs) to load onto a single factor. The fit of the model was acceptable, $\chi^2(54) = 117.36$, $p < .01$, RMSEA = .08, SRMR = .07, NNFI = .92, CFI = .94.[3] Shown in Fig. 1a is the resulting model. As can be seen all of the intrusion types except DRM PLIs loaded significantly and moderately onto the single factor.

In the second model we specified that all of the PLIs from the six recall tasks loaded onto one factor and all of the ELIs across the tasks loaded onto a second factor. The fit of the model was acceptable, $\chi^2(53) = 116.31$, $p < .01$, RMSEA = .08, SRMR = .07, NNFI = .92, CFI = .94. Shown in Fig. 1b is the resulting model. As can be seen, each of the PLIs (except for DRM PLIs) loaded significantly and fairly strongly on the PLI factor, while all of the ELIs loaded significantly onto the ELI factor. Furthermore, the PLI and ELI factors were highly correlated ($r = .94$), suggesting that the two error types were driven by similar underlying mechanisms. In fact, the fit of the two factor intrusion model was not significantly different from the unitary intrusion model, $\Delta\chi^2(1) = 1.05$, $p > .30$. Thus, the simpler unitary model was retained as the preferred model. From an individual differences perspective this suggests that those participants who are likely to output many PLIs are also likely to output many ELIs. This does not mean that there are not important differences between PLIs and ELIs that can arise from specific experimental manipulations (e.g., Unsworth, 2009a, 2009b; Unsworth & Brewer, in press a), rather this suggests that there is a least one important common factor (and perhaps several common factors) between PLIs and ELIs that determines why some

[2] *Note*: despite the low correlations associated with DRM PLIs, they were kept in the models for completeness. Excluding DRM PLIs from the models led to virtually identical results as those reported.

[3] The chi-square statistic reflects whether there is a significant difference between the observed and reproduced covariance matrices. Therefore, nonsignifcant values are desirable. However, with large sample sizes even slight deviations can result in a significant value, therefore the ratio of chi-square to the number of degrees of freedom is also reported. Ratios of two or less usually indicate acceptable fit. Also reported are the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) both of which reflect the average squared deviation between the observed and reproduced covariances. In addition, the nonormed fit index (NNFI) and the comparative fit index (CFI) both of which reflect the proportion of the observed covariance matrix explained by the model are reported. NNFI, and CFI values greater than .90 and SRMR values less than .08 are indicative of acceptable fit (Kline, 1998).

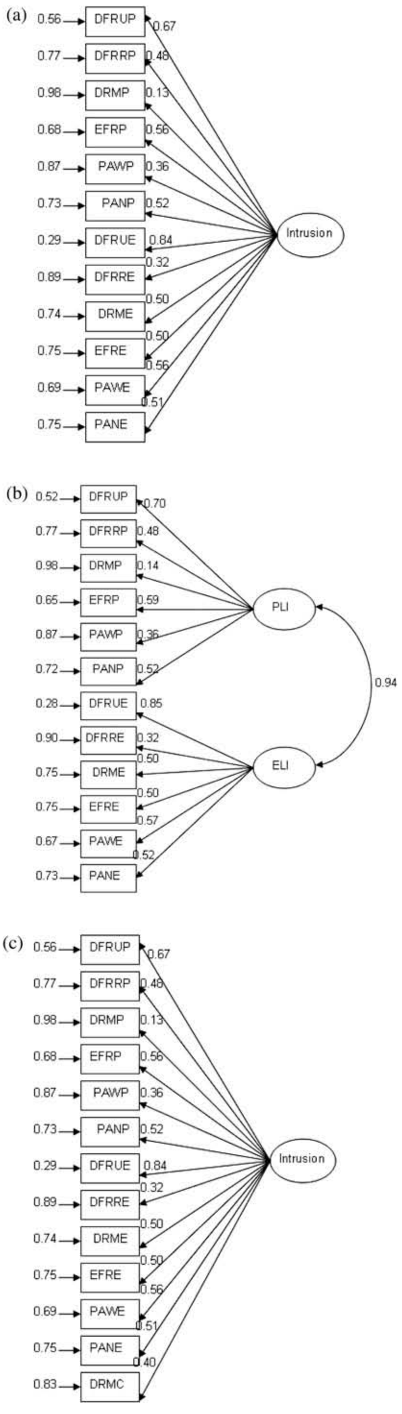

Fig. 1. (a) Confirmatory factor analysis for unitary intrusion model for previous-list and extra-list intrusions; (b) confirmatory factor analysis for separate previous-list and extra-list intrusion model; (c) confirmatory factor analysis for unitary intrusion model with DRM critical intrusions included.

individuals are more likely to emit intrusions than other individuals.

Next, we examined the extent to which DRM critical intrusions would be related to the other types of intrusions. That is, we examined whether DRM critical intrusions are simply another manifestation of more general intrusions or whether DRM critical intrusions should be considered as a separate type of intrusion. We added DRM critical intrusions to the unitary intrusion model. If DRM critical intrusions are related to the other types of intrusions and can be considered the same, then the DRM critical intrusions should load on the intrusion factor without harming the model fit and these intrusions should load with a similar magnitude as the other intrusions. The fit of this model was acceptable, $\chi^2(65) = 128.99$, $p < .01$, RMSEA = .08, SRMR = .07, NNFI = .93, CFI = .94, suggesting that adding DRM critical intrusions to the model did not harm model fit. Shown in Fig. 1c is the resulting model. As can be seen, DRM critical intrusions loaded significantly on the intrusion factor and with a magnitude similar to the other intrusion types. Thus, it would seem that DRM critical intrusions are simply a manifestation of other intrusions and should not necessarily be considered as a completely separate type of intrusion. Rather, whatever mechanism(s) that given rise to PLIs and ELIs more generally, also seem to give rise to DRM critical intrusions. This suggests that theories of false recall need to be able to account for all types of intrusions and account for the fact that these intrusions are all interrelated.[4]

Relations among intrusions and other cognitive constructs

The next set of analyses was done in order to examine the extent to which cognitive abilities such as source monitoring, WMC, judgments of recency (JOR), and vocabulary were related to false recall. Given that the prior analyses suggested that PLIs and ELIs largely measure the same underlying construct, for the following analyses PLIs were combined with ELIs in each task. That is, for each task (except for the DRM) we computed the total number of intrusions that were made. For the DRM task we combined PLIs and ELIs into a single measure but kept DRM critical intrusions separate in order to gauge how these intrusions would be related to the other constructs given the large amount of work that has been done examining DRM critical intrusions. Shown in Table 2 are the descriptive statistics for all of the measures used in these analyses.

As can be seen, most of the measures had generally acceptable values of internal consistency and most of the measures were approximately normally distributed with values of skewness and kurtosis under the generally accepted values (i.e., skewness < 2 and kurtosis < 4; see Kline, 1998). Correlations, shown in Table 3, were generally moderate. Note that some of the descriptive statistics and cor-

[4] *Note*: DRM critical intrusions were kept out of the initial models in order to examine the factor structure of intrusions found in standard recall tasks. Next, in order to determine if DRM critical intrusions were special, they were added to the unitary intrusion factor model. However, we also tested the dual-factor model with DRM critical intrusions included. Similar to the initial intrusion models that are reported, the dual-factor model with DRM critical intrusions did not fit significantly better than the unitary model with DRM critical intrusions, $\Delta\chi^2(1) = .71$, $p > .39$, and the correlation between PLI and ELI factors was exactly the same as in the initial models (i.e., $r = .94$). Thus, the inclusion of DRM critical intrusions into the models did not change any of the results.

Table 2
Descriptive statistics and reliability estimates for all of the measures.

Measure	*M*	*SD*	Skew	Kurtosis	α
DFRUI	3.46	4.13	2.63	10.19	.72
DFRRI	1.81	2.04	1.60	2.73	.52
DRMI	1.27	2.19	4.20	26.43	.50
DRMC	2.53	1.58	.57	.36	.52
EFRI	5.61	9.93	4.95	30.93	.60
PAWI	7.21	5.63	.93	.30	.50
PANI	8.53	5.65	.92	.72	.49
DFRU	134.07	8.87	.22	−.24	.79
DFRR	36.04	6.13	.01	1.29	.71
DRM	54.36	10.53	−.32	−.34	.84
EFR	28.59	8.41	−.05	−.09	.76
PAW	.24	.22	.90	−.28	.83
PAN	.13	.11	.98	.37	.68
GenSour	.64	.14	−.30	−.11	.62
PicSour	.80	.15	−1.92	5.41	.72
Ospan	60.80	11.76	−.38	−.61	.73
Symspan	30.46	7.30	−1.11	−.37	.86
Rspan	57.91	14.14	−1.91	.64	.80
JORW	.57	.26	−.82	1.03	.78
JORP	.72	.28	−1.09	.33	.82
Syn	4.52	2.26	.13	−.68	.60
Ant	4.45	2.06	.38	−.37	.62

Note: DFRU = delayed free recall unrelated words; DFRR = delayed free recall related words; DRM = Deese–Roediger–McDermott paradigm; EFR = externalized free recall; PAW = paired associates with word–word pairs; PAN = paired associates with number–word pairs; I = intrusion; C = critical word intrusion; GenSour = gender source recognition; PicSour = picture source recognition; Ospan = operation span; Symspan = symmetry span; Rspan = reading span; JORW = judgments of recency with words; JORP = judgments of recency with pictures; Syn = synonym vocabulary test; Ant = antonym vocabulary test.

relations have previously been presented elsewhere (Unsworth & Brewer, in press b). Importantly, none of the analyses concerning either intrusions or any of the primary analyses have been reported previously.

Next we used confirmatory factor analysis to examine a measurement model of all of the measures and to determine how each of the putative factors were related to one another. Specifically, we specified one factor as an intrusion factor with intrusions from each of the tasks loading onto that factor and only that factor. We also specified a recall factor based on the six recall tasks. Separate factors were also formed for source memory (made up of the two source monitoring tasks), WMC (made up of the three complex span tasks), JOR (made up of the two JOR tasks), and vocabulary (made up of the two vocabulary measures). All of these six factors were allowed to correlate. Thus, this model tests the extent to which different measures can be grouped into separate yet correlated factors, and examines the latent correlations among the factors. The fit of the model was acceptable, $\chi^2(194) = 282.84$, $p < .01$, RMSEA = .05, SRMR = .06, NNFI = .93, CFI = .94, suggesting that the specified model provided a good description to the underlying pattern of data. The factor loadings for each task and the interfactor correlations are shown in Table 4. As can be seen, each of the measures loaded moderately and significantly on their respective factors. An examination of the interfactor correlations suggested that intrusions in the recall tasks were negatively correlated with veridical recall as has been reported previously (e.g., Roediger et al., 2001; Unsworth, 2009b). Furthermore, source-monitoring abilities were moderately related to false recalls, whereas the other cognitive abilities (WMC, JOR, and vocabulary) demonstrated weaker correlations with intrusions. Finally, the other cognitive abilities were weakly to moderately related with one another. For instance, recall and source monitoring were fairly highly correlated to one another and both were related to WMC. Source monitoring was moderately related to JORs, whereas recall demonstrated a weaker relation, and WMC and vocabulary were not related to JORs. Finally, vocabulary was only related to intrusions and WMC. These results suggest that many of the cognitive abilities were related to one another and were related to individual differences in false recall.

Next, we used structural equation modeling to determine which of the cognitive ability factors predicted intrusions. Specifically, we specified a model in which source monitoring, WMC, JOR, and vocabulary predicted intrusions. Note, veridical recall was not included in the model given that we were primarily interested in how other cognitive abilities would account for individual differences in intrusions and because the same tasks were used to obtain estimates of veridical recall and false recall. Thus, the correlation could be due to shared method variance (Lövdén, 2003). Furthermore, given the strong correlation (.78) between the source monitoring factor and the recall factor, it is likely that much of the much variance predicted in intrusions would be due to shared variance between source monitoring and recall. In fact, given the stronger relation between source monitoring and intrusions (−.58) than the relation between recall and intrusions (−.49), source monitoring should account for the relation between recall and intrusions. Indeed, the correlation between source monitoring and intrusions remained significant even after partialling out recall, $pr\ (177) = -.36$, $p < .01$, but the correlation between recall and intrusions was no longer significant and near zero after partialling out source monitoring, $pr\ (177) = -.07$, $p > .32$. Thus, the relation between recall and intrusions seemed to be due to the substantial shared variance with source-monitoring abilities.

In the specified structural equation model each of the four cognitive ability factors (excluding recall) were allowed to correlate with one another based on the previous confirmatory factor analysis, and all were allowed to predict intrusions. This model examines the extent to which each of the cognitive ability factors accounts for unique or shared variance in predicting intrusions. As shown in the previous confirmatory factor analysis each of the cognitive ability factors was significantly related to intrusions, but it is not known whether these correlations represent unique relations or whether the correlations represent shared variance. That is, is the relation between WMC and intrusions due to specific WMC processing such as goal maintenance, or is this relation really due to shared variance with source-monitoring abilities? If this variance is due to unique processes, then WMC should have a direct link to intrusions. If, however, the relation is due to shared variance with source-monitoring abilities, then the path from WMC to intrusions should not be significant, but the path between source monitoring and intrusions should

Table 3
Correlations for all of the measures.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. DFRUI	1.00																					
2. DFRRI	0.42	1.00																				
3. DRMI	0.45	0.31	1.00																			
4. DRMC	0.34	0.22	0.32	1.00																		
5. EFRI	0.51	0.17	0.36	0.15	1.00																	
6. PAWI	0.46	0.31	0.22	0.19	0.33	1.00																
7. PANI	0.53	0.32	0.27	0.18	0.32	0.50	1.00															
8. DFRU	−0.42	−0.16	−0.15	−0.20	−0.11	−0.28	−0.17	1.00														
9. DFRR	−0.11	−0.28	−0.08	−0.17	−0.03	−0.17	0.00	0.27	1.00													
10. DRM	−0.26	−0.13	−0.24	−0.15	−0.14	−0.38	−0.16	0.56	0.44	1.00												
11. EFR	−0.23	−0.05	−0.14	−0.19	−0.24	−0.24	−0.05	0.51	0.23	0.51	1.00											
12. PAW	−0.22	−0.14	−0.10	−0.11	−0.07	−0.43	−0.12	0.36	0.27	0.48	0.26	1.00										
13. PAN	−0.25	−0.06	−0.09	−0.14	−0.16	−0.31	−0.28	0.43	0.15	0.32	0.38	0.31	1.00									
14. GenSour	−0.15	−0.06	−0.12	−0.06	−0.08	−0.32	−0.10	0.25	0.20	0.25	0.29	0.20	0.24	1.00								
15. PicSour	−0.27	−0.17	−0.16	−0.06	−0.20	−0.43	−0.24	0.28	0.17	0.36	0.36	0.28	0.21	0.28	1.00							
16. Ospan	−0.15	−0.11	−0.13	−0.15	−0.11	−0.29	−0.14	0.09	0.23	0.20	0.26	0.16	0.16	0.12	0.15	1.00						
17. Symspan	−0.16	−0.16	−0.09	−0.07	−0.13	−0.18	−0.18	0.04	0.17	0.18	0.22	0.03	0.12	0.18	0.18	0.48	1.00					
18. Rspan	0.03	−0.06	−0.11	−0.01	−0.03	−0.20	0.00	0.06	0.15	0.22	0.31	0.19	0.11	0.30	0.14	0.52	0.42	1.00				
19. JORW	−0.19	−0.10	−0.04	−0.04	−0.15	−0.21	−0.02	0.09	0.10	0.06	0.09	0.13	0.12	0.21	0.13	−0.02	0.10	0.03	1.00			
20. JORP	−0.19	−0.11	−0.04	−0.02	−0.11	−0.24	−0.03	0.15	0.08	0.12	0.03	0.15	0.14	0.17	0.20	0.06	0.11	0.01	0.62	1.00		
21. Syn	−0.10	−0.04	−0.02	−0.01	−0.11	−0.16	−0.10	0.00	0.03	0.10	0.14	0.16	0.00	0.02	0.05	0.15	0.07	0.09	0.11	0.00	1.00	
22. Ant	−0.14	−0.07	−0.16	−0.08	−0.14	−0.23	−0.14	0.00	0.04	0.09	0.05	0.21	0.00	0.06	0.09	0.19	0.07	0.14	0.06	0.00	0.53	1.00

Note. DFRU = delayed free recall unrelated words; DFRR = delayed free recall related words; DRM = Deese-Roediger-McDermott paradigm; EFR = externalized free recall; PAW = paired associates with word-word pairs; PAN = paired associates with number-word pairs; I = intrusion; C = critical word intrusion; GenSour = gender source recognition; PicSour = picture source recognition; Ospan = operation span; Symspan = symmetry span; Rspan = reading span; JORW = judgments of recency with words; JORP = judgments of recency with pictures; Syn = synonym vocabulary test; Ant = antonym vocabulary test.

Table 4
Confirmatory factor analysis for intrusions, recall, source, WMC, JOR and vocabulary measures.

Measure	Latent Factor					
	Intrusions	Recall	Source	WMC	JOR	Vocabulary
DFRUI	.83					
DFRRI	.50					
DRMI	.52					
DRMC	.40					
EFRI	.56					
PAWI	.63					
PANI	.63					
DFRU		.71				
DFRR		.45				
DRM		.79				
EFR		.66				
PAW		.55				
PAN		.51				
GenSour			.47			
PicSour			.58			
Ospan				.77		
Symspan				.62		
Rspan				.67		
JORW					.75	
JORP					.83	
Syn						.60
Ant						.88
Interfactor Correlations						
Intrusions	–					
Recall	−.49*	–				
Source	−.58*	.78*	–			
WMC	−.26*	.36*	.45*	–		
JOR	−.25*	.20*	.42*	.08	–	
Vocabulary	−.25*	.12	.15	.24*	.04	–

* $p < .05$.

be significant. The same rationale applies to the other cognitive abilities.

Shown in Fig. 2 is the resulting structural equation model. The fit of the model was good, $\chi^2(94) = 120.60$, $p < .05$, RMSEA = .04, SRMR = .06, NNFI = .95, CFI = .96, suggesting that the specified model provided a good account of the data. As can be seen in Fig. 2, only source monitoring uniquely predicted variation in intrusions. The other three cognitive ability constructs (WMC, JOR, and vocabulary) did not account for any unique variance in intrusions. This suggests that individual differences in source-monitoring abilities are one reason (and the primary reason in the current study) why individuals differ in false recall. Furthermore, these results suggest that source-monitoring abilities mediate the relation between WMC and intrusions and the relation between JORs and intrusions. That is, these results suggest that previous work which has suggested a link between WMC and intrusions (Unsworth, 2007; Watson et al., 2005) may be due to differences in source-monitoring abilities. Although the current model clearly suggests this, we decided to explicitly test mediation models to see if source-monitoring abilities fully mediated the relations between WMC and intrusions and JORs and intrusions.

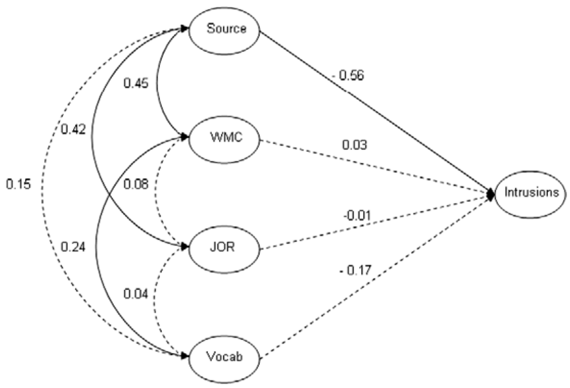

Fig. 2. Structural equation model predicting intrusions with Source, WMC, JOR, and Vocab. Single-headed arrows connecting latent variables (circles) to each other represent standardized path coefficients indicating the unique contribution of the latent variable. Double headed arrows connecting the memory factors represent the correlations among the factors. Solid lines are significant at the $p < .05$ level and dotted lines are not significant at the $p < .05$ level.

In the first mediation model we specified a model in which WMC predicted source monitoring, WMC predicted intrusions, and source monitoring predicted intrusions. If source monitoring accounts for the relation between WMC and intrusions, then we should see that WMC is related to source monitoring, source monitoring is related to intrusions, but WMC does not have a direct link to intrusions. If source monitoring does not fully mediate the relation between WMC and intrusions, then the direct path between WMC and intrusions should be significant. Shown in Fig. 3a is the resulting model. The fit of the model was acceptable, $\chi^2(51) = 91.58$, $p < .01$, RMSEA = .07, SRMR = .06, NNFI = .92, CFI = .94. As shown in Fig. 3a, WMC predicted source monitoring, and source monitoring predicted intrusions, but WMC did not have a direct effect on intrusions once source monitoring was taken into account. Thus, the relation between WMC and intrusions was fully mediated by source-monitoring processes.

A similar analysis was conducted examining whether source monitoring would mediate the relation between JORs and intrusions. In this model JORs predicted source monitoring, source monitoring predicted intrusions, and there was a direct path from JORs to intrusions. The fit of this model was acceptable, $\chi^2(41) = 66.89$, $p < .01$, RMSEA = .06, SRMR = .06, NNFI = .94, CFI = .95. As shown in Fig. 3b, JORs predicted source monitoring, source monitoring predicted intrusions, but the path from JORs to intrusions was not significant once source monitoring was taken into account. Like the WMC model, this result suggests that the relation between JORs and intrusions was fully mediated by source-monitoring processes. Collectively these results suggest that several cognitive constructs are related to individual differences in false recall, but the majority of these constructs are related to false recalls through their shared relation with source monitoring. Thus, within the current study, the primary determinant of one's susceptibility to falsely recall was a deficit in source-monitoring abilities.

Discussion

The current study examined the nature of intrusion errors (false recalls) in multiple free and cued recall tasks. It was found that although intrusions can usefully be classified into two primary categories (PLIs and ELIs) these two intrusion types were highly correlated at a latent level and were best accounted for by a single underlying factor. Furthermore, critical intrusions in the DRM paradigm were also related to these other types of intrusions and loaded with a similar magnitude as the other intrusions on the unitary intrusion factor. These results suggest that intrusion errors, of all types, arise from deficits in the same set of processes regardless of the exact nature of the intrusion (i.e., whether they were from a previous list or were based on shared semantic or phonological features). This common factor likely reflects variation in post-retrieval editing and monitoring processes that are needed for PLIs, ELIs, and DRM critical intrusions. Importantly, it should also be noted that although PLIs and ELIs were strongly correlated and could be accounted for by a single factor, this does not mean that there are not also important differences between PLIs and ELIs. Specifically, it is clear that the factors that lead to the generation of PLIs and ELIs are usually quite different, with PLIs being generated based on shared temporal contextual overlap with target items and ELIs being generated based on shared semantic and phonological overlap with target items. Thus, although there are clearly similarities between PLIs and ELIs leading to the strong association between the two, there are also clearly differences between them. Just because two variables are highly correlated does not preclude the possibility of finding subtle differences between them as well.

Important evidence that the correlation between PLIs and ELIs is due to the shared need for post-retrieval editing comes from a detailed examination of intrusions in the externalized free recall task. In this task participants were instructed to not only recall target items from the list, but to also emit all items that came to mind (i.e., intrusions) during the recall period. Additionally, in this variant of the task participants were instructed to press the space bar each time they emitted an item they knew was incorrect (Kahana et al., 2005). Thus, this manipulation allows for an examination of the specificity of the post-retrieval editing process. Importantly, as noted in "Methods" section, intrusion errors from this task were only those intrusions that were not associated with a space bar press indicating that they arose from a failure in post-retrieval editing. As shown in Table 1, PLIs and ELIs in this task correlated at .40. However, when we examine all PLIs and ELIs regardless of whether they were associated with a space bar press the correlation was only .07. This suggests that the correlation between PLIs and ELIs is driven by shared variation in post-retrieval editing. As noted above, this does not preclude the possibility of important differences between PLIs and ELIs that are due to mechanisms other than post-retrieval editing.

A related issue is how criterion placement can affect the rate of intrusions and individual differences in false recalls. Specifically, it is likely that at least some of the shared variance between PLIs and ELIs is due to individual variation in criterion placement. Some individuals may have a more liberal criterion thereby producing many PLIs and ELIs, whereas other individuals will likely have a more conservative criterion producing few, if any, intrusions. Thus, variation in criterion placement may partially drive some of the relation between PLIs and ELIs resulting in a common factor. Importantly, as noted by Koriat and Goldsmith (1996) overall monitoring abilities and criterion placement will interact to determine whether a response will be given

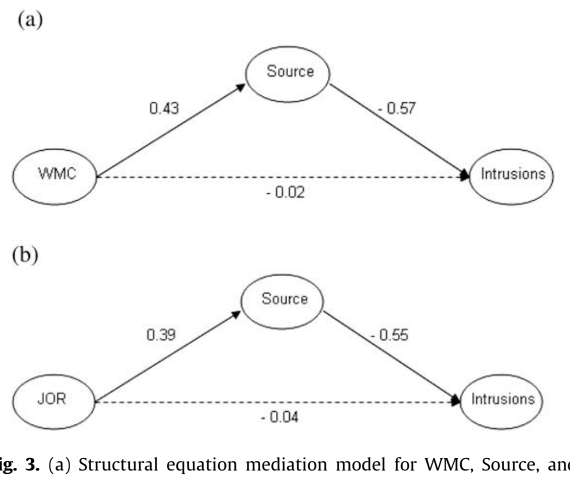

Fig. 3. (a) Structural equation mediation model for WMC, Source, and Intrusions; (b) structural equation mediation model for JOR, Source, and Intrusions. Solid lines are significant at the $p < .05$ level and dotted lines are not significant at the $p < .05$ level.

or withheld. Thus, the post-retrieval editing process involves both an evaluation of each response as correct or incorrect (likely based on source and contextual evidence) as well as a decision to report or withhold that response based on the current placement of the criterion. Clearly more work is needed to examine individual differences in monitoring and criterion placement and how these differences may influence variation in false recall.

An examination of the relation between multiple cognitive constructs and intrusions suggested that intrusions were negatively related to all of the cognitive constructs in the current study. In particular, it was found that intrusions were negatively related to veridical recall on the same tasks suggesting that those individuals who are better at recalling the correct target items were less likely to falsely recall items than participants who are worse at recalling correct items (Unsworth, 2009b). Additionally, it was found that a source monitoring latent variable was moderately correlated with intrusions such that individuals with superior source-monitoring abilities were less susceptible to false recalls than individuals with poor source-monitoring abilities. It was also found that latent variables of WMC, JORs, and vocabulary knowledge were all weakly negatively related to intrusions. The negative correlation between WMC and intrusions is consistent with prior work demonstrating a negative relation in the DRM task in particular (Watson et al., 2005) and in free recall intrusions more broadly (Unsworth, 2009b). Likewise, the negative relation between JORs and intrusions suggests that participants may rely on recency information when making decisions about whether to recall an item or whether to edit it out. Individuals who are better at relying on recency information should be better at editing out intrusions based on the fact that the item was presented too far back in time to be member of the current list in the case of PLIs or was not presented at all in the case of ELIs.

Follow up structural equation models qualified some of these effects, however, by suggesting that source-monitoring abilities accounted for the relations between some of the cognitive variables and intrusions. Specifically, simultaneously examining how source monitoring, WMC, JORs, and vocabulary would predict intrusions, suggested that only source monitoring accounted for unique variance in intrusions. Specific mediation models further suggested that the relation between WMC and intrusions and the relation between JORs and intrusions was entirely mediated by individual differences in source-monitoring abilities. Thus, although WMC and JORs were found to be related to false recalls, these latent constructs were only related to false recalls due to the shared relation with source-monitoring abilities. In terms of WMC, this suggests that the relation between WMC and false recalls is likely not due to differences in goal maintenance abilities (e.g., Watson et al., 2005), rather these differences are due to differences in source monitoring or context discrimination abilities (Unsworth, 2007). Although given that we did not provide explicit warnings to participants (Watson et al., 2005) it is possible that need for goal maintenance was weak. Future work should examine the possibility that when participants are provided with explicit warnings both WMC and source monitoring are needed to account for false recalls, or whether only source monitoring is needed as demonstrated in the current study.

In terms of the JORs, the current results suggest that not only are participants relying on recency information to edit out their intrusions, but they are relying on more global source and context based information of which recency is only one part. That is, it is likely that when deciding whether an item was a member of the current target set, recency information is used, but other sources of information and decision processes will also be important (Johnson et al., 1993). Importantly, it should be noted that although source-monitoring abilities were the primary predictor of intrusions in the current study, this does not mean that intrusions are caused by a single mechanism. Rather, whether an individual is likely to emit an intrusion is likely driven by multiple mechanisms that operate both prior to retrieval and after retrieval. Thus, we are not advocating a single mechanism approach; rather we are suggesting that source-monitoring abilities are one important factor that contributes to intrusions and individual differences in false recalls. Furthermore, these source-monitoring abilities can themselves be broken down into multiple sub-components (Johnson et al., 1993), each of which may be important in explaining false recalls.

Implications for models of false recall

The current results have a number of implications for current theories of false recall. For instance, the current finding that all intrusions are related and seem to arise from the same set of underlying mechanisms is consistent with the fSAM model (Kimball et al., 2007). In this model it is assumed that during item presentation 3–4 items are held in the short-term store. While in the short-term store it is assumed that coactive items activate each others lexical-semantic representations and are associated with similar contextual features. At retrieval, it is assumed that the most recently retrieved 3–4 items (including both corrects and intrusions) are used as retrieval cues to cue the next item. Importantly, this model predicts the high number of DRM critical intrusions in the DRM task while also predicting lower levels of other ELIs and PLIs. The reason for differential rates of intrusion errors is because DRM critical intrusions have strong semantic relations to all of the words on a list, whereas other ELIs are usually only related to a few items on the list, and PLIs are usually related based on shared temporal–contextual features. Thus, this model suggests that the same set of processes gives rise to DRM critical intrusions, other ELIs, and PLIs, which is consistent with the current findings which suggest that the same underlying factor is responsible for PLIs, ELIs, and DRM critical intrusions across multiple tasks. Given these results, it is clear that models of false recall need to be able to account not only for DRM critical intrusions, but for intrusions more broadly. Furthermore, as demonstrated by fSAM, these models need to be able to account for the different frequencies associated with different intrusion types that change as a function of the task and word lists used. Models that concentrate on explaining only one type of intrusion will only tell part of the overall story.

At the same time, although the fSAM model should be able to account for the current results in terms of the overall relations among the intrusions, it is currently limited by the fact that it does not implement any post-retrieval decision or source-monitoring processes. That is, the fSAM model would not be able to account for the fact that source-monitoring abilities seem to be the primary determinant of whether someone is likely to falsely recall an item. The activation-monitoring theory of Roediger, McDermott and colleagues (Roediger & McDermott, 1995; Roediger et al., 2001), on the other hand, naturally accounts for fact that source monitoring is an important component of false recalls. Specifically, the activation-monitoring theory suggests that not only are activation processes important, but monitoring the products of the activation processes are also important. These monitoring processes are thought to operate at both encoding and retrieval, but it seems clear that they are predominantly important at retrieval to judge whether an item that has just been retrieved is in fact a correct item or whether it is an intrusion. Because of the emphasis on monitoring processes, the current results are very much in line with the activation-monitoring view and the notion that individual and age differences in false recall are due, in large part, to failures in source monitoring (Lövdén, 2003; Unsworth, 2007; Watson et al., 2001).

Given the importance of source-monitoring processes to false recalls, one remaining question is when do these processes operate? As noted by Roediger et al. (2001) these monitoring processes may operate at either encoding or retrieval. Furthermore, it is possible that during retrieval these source-monitoring processes operate at both preretrieval and post-retrieval. That is, it is usually assumed that source-monitoring processes of the type discussed thus far operate after an item has been retrieved and work as an editing processes as in generate-edit models of free recall. In this view monitoring works as a post-retrieval processes in which the contextual/source features of retrieved items are examined to determine if the current item was part of the current list or whether it was part of a previous list (a PLI) or whether it is simply related to one of the current list items (an ELI). As noted previously, recency information and other source information will be needed during this post-retrieval phase to make an assessment on whether the item should be recalled. If it is determined that the item shares many contextual features with the current list items then the item will be recalled. If it is determined that the item does not share enough contextual features then it will be edited out and not recalled. Thus, these post-retrieval processes work as a late correction filter to ensure that incorrect items are not recalled (e.g., Jacoby, Kelley, & McElree, 1999).

Conversely it is possible that these source-monitoring processes operate preretrieval to constrain or focus the search of items to only those representations that share many contextual features with the retrieval cue. That is, source monitoring or contextual discrimination processes would be needed to specify the cues that would be used to search for items prior to the search actually taking place. This would result in a source constrained search of memory (Jacoby, Shimizu, Daniels, & Rhodes, 2005) in which intrusions would not be included in the search set because of very specific context cues. In accounting for individual differences in WMC and intrusions, our prior work has suggested just this type of preretrieval process as being important in accounting for the relation between WMC and intrusions (Unsworth, 2007). Specifically, we suggested that individuals high in WMC were better at using contextual cues to focus the search on only the most recent list of items. Individuals low in WMC, however, utilize noisier contextual cues which results in the search sets being composed of both target items and items from previous lists (PLIs). Thus, this account suggests that early selection processes (Jacoby et al., 1999) are important for differences in false recalls, with some participants being better at relying on specific cues at the front end than other participants.

In terms of explaining false recalls and individual differences in false recalls it is likely that both early selection and late correction processes will be important. That is, an examination of both source constrained search processes at the front end and source based monitoring and editing processes at the back end will be important in explaining the rate of false recalls and individuals differences in false recalls. For instance, using the same externalized free recall task as used in the current study we have recently found that high and low WMC individuals differ not only in the number of intrusions they generate, but they also differ in the number of intrusions that they correctly monitor (Unsworth & Brewer, in press a). We suggested that high and low WMC individuals differed in both the ability to constrain the search at the front end and monitor for potential errors at the back end. Given that the current study found that the relation between WMC and intrusions was fully mediated by source-monitoring abilities, this suggests that the prior WMC differences found in both early selection and late correction were both due to deficits in source monitoring.

Collectively, the current results suggests that combining a generation or activation process similar to that utilized in the fSAM model with specific source-monitoring processes as found in the activation-monitoring model should be able to account for intrusions and individual differences in false recall. The generation or activation component of the fSAM model should be able to account for the fact that all intrusion types seem to be due to the same set of underlying processes and the activation-monitoring model should be able to account for the fact that poor source monitoring is a primary reason for false recalls and individual differences in false recalls seen across different types of intrusions from different recall tasks. That is, as suggested by the activation-monitoring account, both generation/activation and source-monitoring processes are needed to fully account for false recalls. Furthermore, it is likely that both preretrieval and post-retrieval source-monitoring processes will be needed to fully account for false recalls and for individual differences in the susceptibility to falsely recall.

Acknowledgment

Thanks to Greg Spillers, Mike Kahana, and two anonymous reviewers for comments on an earlier version of the article.

References

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D. A., Roediger, H. L., McDermott, K. B., et al. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimers type. *Cognitive Neuropsychology, 16*, 361–384.

Blair, I. V., Lenton, A. P., & Hastie, R. (2002). The reliability of the DRM paradigm as a measure of individual differences in false memories. *Psychonomic Bulletin & Review, 9*, 590–596.

Butler, K. M., McDaniel, M. A., Dornburg, C. C., Price, A. L., & Roediger, H. L. (2004). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin & Review, 11*, 921–925.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Chan, J. C. K., & McDermott, K. B. (2007). The effects of frontal lobe functioning and age on veridical and false recall. *Psychonomic Bulletin & Review, 14*, 606–611.

Craik, F. I. M. (1968). Types of error in free recall. *Psychonomic Science, 10*, 353–354.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22.

Gallo, D. A. (2006). *Associative illusions of memory*. New York: Psychology Press.

Gerrie, M. P., & Garry, M. (2007). Individual differences in working memory capacity affect false memories for missing aspects of events. *Memory, 15*, 561–571.

Hambrick, D. Z., Salthouse, T. A., & Meinz, E. J. (1999). Predictors of crossword puzzle proficiency and moderators of age–cognition relations. *Journal of Experimental Psychology: General, 128*, 131–164.

Helkala, E. L., Laulumaa, V., Soininen, H., & Riekkinen, P. (1989). Different error pattern of episodic and semantic memory in Alzheimer's Disease and Parkinson's Disease with Dementia. *Neuropsychologia, 27*, 1241–1248.

Hicks, J. L., & Marsh, R. L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1195–1209.

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection vs late correction. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 383–400). NY: Guilford.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12*, 852–857.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28.

Jou, J. (2008). Recall latencies, confidence, and output positions of true and false memories: Implications for recall and metamemory theories. *Journal of Memory and Language, 58*, 1049–1064.

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in episodic recall: Age differences in editing of overt responses. *Journal of Gerontology: Psychological Sciences, 60B*, 92–97.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence. An individual differences perspective. *Psychonomic Bulletin & Review, 9*, 637–671.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review, 114*, 954–993.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Konishi, S., Uchida, I., Okuaki, T., Machida, T., Shirouzu, I., & Miyashita, Y. (2002). Neural correlates of recency judgment. *The Journal of Neuroscience, 22*, 9549–9555.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.

Lövdén, M. (2003). The episodic memory and inhibition accounts of age-related increases in false memories: A consistency check. *Journal of Memory and Language, 49*, 268–283.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35*, 212–230.

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language, 45*, 160–176.

Murdock, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Lawrence Erlbaum and Associates.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*, 838–848.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. Bower (Ed.). *The psychology of learning and motivation* (Vol. 14). New York: Academic Press.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 803–814.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*, 385–407.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language, 39*, 418–436.

Rouleau, I., Imbault, H., Laframboise, M., & Bédard, M.-A. (2001). Pattern of intrusions in verbal recall: Comparison of Alzheimer's disease, Parkinson's disease and frontal lobe d ementia. *Brain & cognition, 46*, 244–249.

Sirotin, Y. B., Kimball, D. R., & Kahana, M. J. (2005). Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin & Review, 12*.

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General, 107*, 393–419.

Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 1020–1034.

Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal–contextual search. *Journal of Memory and Language, 59*, 223–236.

Unsworth, N. (2009a). Examining variation in working memory capacity and retrieval in cued recall. *Memory, 17*, 386–396.

Unsworth, N. (2009b). Variation in working memory capacity, fluid intelligence, and episodic recall: A latent variable examination of differences in the dynamics of free recall. *Memory & Cognition, 37*, 837–849.

Unsworth, N., & Brewer, G.A. (in press a). Variation in working memory capacity and intrusions: Differences in generation or editing? European Journal of Cognitive Psychology.

Unsworth, N., & Brewer, G.A. (in press b). Examining the relationships among item recognition, source recognition, and recall from an individual differences perspective. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*, 104–132.

Watson, J. M., Balota, D. A., & Sergent-Marshall, S. D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with dementia of the Alzheimer Type. *Neuropsychology, 15*, 254–267.

Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. A. (2005). Individual differences in susceptibility to false memory in the Deese–Roediger–McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 76–85.

Watson, J. M., McDermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in the Deese–Roediger–McDermott paradigm: Assessing the combined influence of practice and warnings in young and old adults. *Memory & Cognition, 32*, 135–141.

Winograd, E., Peluso, J. P., & Glover, T. A. (1998). Individual differences in susceptibility to memory illusions. *Applied Cognitive Psychology, 12*, S5–S27.

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., et al. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 792–804.