Source Memory Revealed Through Eye Movements and Pupil Dilation

by

Megan H. Papesh

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2012 by the
Graduate Supervisory Committee:

Stephen D. Goldinger, Chair
Gene A. Brewer
Erik D. Reichle
Donald Homa
Arthur M. Glenberg

ARIZONA STATE UNIVERSITY

May 2012

ABSTRACT

Current theoretical debate, crossing the bounds of memory theory and mental imagery, surrounds the role of eye movements in successful encoding and retrieval. Although the eyes have been shown to revisit previously-viewed locations during retrieval, the functional role of these saccades is not known. Understanding the potential role of eye movements may help address classic questions in recognition memory. Specifically, are episodic traces rich and detailed, characterized by a single strength-driven recognition process, or are they better described by two separate processes, one for vague information and one for the retrieval of detail? Three experiments are reported, in which participants encoded audio-visual information while completing controlled patterns of eye movements. By presenting information in four sources (i.e., voices), assessments of specific and partial source memory were measured at retrieval. Across experiments, participants' eye movements at test were manipulated. Experiment 1 allowed free viewing, Experiment 2 required externally-cued fixations to previously-relevant (or irrelevant) screen locations, and Experiment 3 required externally-cued new or familiar oculomotor patterns to multiple screen locations in succession. Although eye movements were spontaneously reinstated when gaze was unconstrained during retrieval (Experiment 1), externally-cueing participants to re-engage in fixations or oculomotor patterns from encoding (Experiments 2 and 3) did not enhance retrieval. Across all experiments, participants' memories were well-described by signal-detection models of memory. Source retrieval was characterized by a continuous process, with evidence that source retrieval occurred following item memory failures, and additional evidence that participants partially recollected source, in the absence of specific item retrieval. Pupillometry provided an unbiased metric by which to compute receiver operating characteristic (ROC) curves, which were consistently curvilinear (but linear in *z*-space), supporting signal-detection predictions over those from dual-process theories. Implications for theoretical views of memory representations are discussed.

DEDICATION

To my parents, Stephen and Kathleen Papesh, who have provided endless support (and top-notch dog-sitting!) throughout my education. Thank you.

ACKNOWLEDGMENTS

I am grateful to the members of my committee, Arthur ("TGU") Glenberg, Donald Homa, Erik ("EZ-Rider") Reichle, Gene Brewer, and Stephen Goldinger, for their feedback on this dissertation, and other milestone projects. I owe a particular debt of gratitude to Steve Goldinger, for his constant mentorship, encouragement, and patience throughout my graduate training. Thank you, also, to the many research assistants involved in collecting the data reported here, including Melissa Miola, who managed almost all laboratory operations, Anne Kuni, Hannah Messick, Tresa Marchi, Kelsey Grodner, Rachelle Friedman, Suhani Mehrotra, Jessie Moecia, Natalie Wall, and Laura Berman. Finally, I thank my colleagues, Michael Hout, Anthony Barnhart, and Whitney Hansen, each for different reasons, but all for their support and friendship. This dissertation was funded in part by a dissertation completion fellowship awarded from the Graduate College of Arizona State University, and in part by the Robert B. Cialdini Dissertation Project Prize in Psychology.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

Figure Page

Source Memory Revealed Through Eye Movements and Pupil Dilation

The nature and structure of recognition memory have been debated since Tulving (1983) introduced the term “episodic memory” to describe memory for everyday experiences. Episodic memories, however, are not always concrete and are often experienced as vague feelings of knowledge. For example, when shopping at the grocery store, you may run into work colleagues, old friends, family members, or a vaguely familiar clerk from a different store. You recognize most of these people with ease, accurately accessing your memory of each person and, more specifically, from where you know them. But memory is not always so specific; sometimes, as in Mandler’s (1980) famous example, we see our local butcher on the bus, and experience a feeling of familiarity that can range from vague (“I think I know that guy”) to strong (“Seriously – how do I know that guy?!”). Importantly, this experience is not necessarily resolved with recollected details; you cannot name the butcher, but know that you know him. Such apparent memory failures provide important anecdotal evidence about the everyday function of memory; memory can go undetected and under-appreciated until we experience such salient and frustrating failures. An interesting facet of such everyday memory failures, however, is what they reveal about the general function and architecture of the recognition memory system: Memories can be strong, weak, detailed, or partial, yet they are all “memory.” The aim of this dissertation is to critically examine several competing hypotheses regarding the components of episodic memory, including the nature of recollection and the functional role of eye movements across encoding and retrieval.

Recognition Memory

In broadly conceptualizing recognition memory, in particular the experience of recollection, two dominant and opposing views have persisted for decades, continuous and threshold models. Variations of each model have been proposed, each with varying degrees of success, but for present purposes, they will be dichotomized by their theoretical treatment of recollection. Hybrid models, which combine select elements from continuous and threshold models, will also be discussed. All of the models assume that memory decisions are based on the retrieval of some degree of evidence (for consistency, I will call that evidence “strength;” this is for clarity, not a theoretical stance).

Threshold Models

Threshold models are often traced back to Fechner’s psychophysical research (Boring, 1929), because the assumption inherent in these models is that a single "evidence" threshold must be exceeded before an item is detected as previously encountered. *High-threshold* model is perhaps the simplest of such views: According to high-threshold model, memory strength is characterized by two distributions (often visually depicted as square distributions, but this is not a theoretically-constrained assumption), one representing target item strength and one representing foil item strength. Because targets have more inherent strength, the target distribution is centered to the right of the foil distribution. A critical assumption of this model regards the tails of the foil distribution. Specifically, high-threshold model assumes that the lower tail of the foil distribution *does not* extend beyond the lower tail of the target distribution; it is entirely encapsulated by the target distribution. Further, high-threshold model predicts that observers adopt a single criterion, which divides the distribution of memory strength into dichotomous old/new responses (see Macmillan & Creelman, 2005). If the criterion is set exactly at the threshold, performance will be perfect (all hits and no false alarms). Observers rarely select the ideal criterion, however, and in cases in which old items go undetected, this model assumes that the observer should have no conscious feelings of memory.

Similarly, *high-low threshold model* (sometimes referred to simply as *two-high threshold model*, see Hilford et al., 2002; Yonelinas, 2002) assumes that memory is a categorical process, but it includes a method by which new items can be "recognized" as new (e.g., "I would have remembered the word *waffle* because my dog's name is Eggo"). Specifically, this model also proposes two overlapping distributions of memory strength for targets and foils, but, in contrast to standard high-threshold model, it assumes that the foil distribution extends beyond the lower tail of the target distribution. Observers still set a single criterion, reflecting all-or-none memory retrieval, but they are also assumed to "recognize" new items from the portion of the foil distribution that does not overlap the targets.

Continuous Models

In contrast to threshold models, continuous models do not make the assumption that recollection is a threshold, all-or-none process. Rather, they posit a continuous stream of evidence, capable of eliciting a range of recollected details. As summarized by Wixted (2007; Wixted & Mickes, 2010), continuous models are based on the principles of signal-detection theory, and suggest that recognition memory decisions are made by comparing the strength of the retrieved memory signal to a decision criterion. As in standard signal-detection theory, this view proposes the theoretical existence of two Gaussian distributions (see Figure 1), one reflecting target strength and one reflecting foil strength. During a recognition test, any item that yields memory strength exceeding the decision criterion is judged "old," whereas items with lower strengths are judged "new." Although equal-variance models (left panel of Figure 1) usefully illustrate general signal-detection-based models (and, in fact, have been incorporated into hybrid models, see below), abundant evidence supports *unequal* variance distributions, wherein the target distribution is wider than the foil distribution (right panel of Figure 1; see, Ratcliff, Sheu, & Gronlund, 1992; Wixted & Stretch, 2004). Critically, regardless of the distributional assumptions, continuous models all propose that recognition decisions are based on a concept of *continuous memory strength* (Wixted & Stretch, 2004), and not on a threshold memory process.

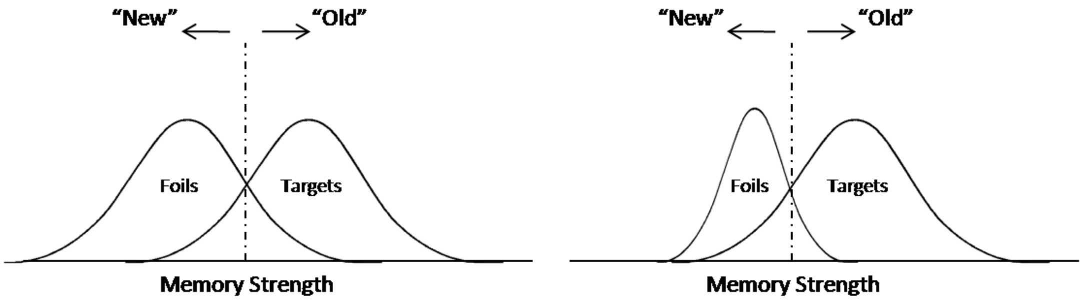


Figure 1. Single-process, signal-detection models of recognition memory. The left set of distributions corresponds to equal-variance models and the right set corresponds to unequal-variance models. Adapted from Wixted (2007).

Hybrid Models

Contemporary threshold theories, such as dual-process signal detection theory (DPSD, see Yonelinas, 1994, 2001), incorporate elements from signal-detection theory. In similar fashion, recent continuous theories, such as continuous dual-process signal-detection (CDP) theory (Wixted & Mickes, 2010), incorporate aspects of dual-process theory, resulting in different takes on hybrid models.

Hybrid threshold theories, henceforth, "*dual-process theories*," are intuitively appealing. In the "butcher on the bus" example described above, people have the strong sense that they are not recollecting episodic detail; instead the butcher is merely familiar. This feeling of knowing can be easily contrasted with more richly detailed memories, such as not only recognizing your butcher, but also remembering that he recently gave you a great recipe for grilled salmon. This intuitive feeling of mnemonic dissociation encapsulates the nature of dual-process theories. Rather than assume that the same processes support detailed and relatively vague memories, as does not subjectively seem true, dual-process theories propose that recognition memory is served by two distinct, independent processes, *recollection* and *familiarity* (Jacoby, 1991). Whereas recollection is assumed to occur by consciously controlled processing, reflecting a person's ability to recall the specific details of the encoding event, familiarity is said to operate quickly and automatically, reflecting a vague "feeling of knowing" that the encoding event had occurred.

An early version of a hybrid DPSD model was the *two-criterion model,* wherein familiarity-based decisions are made quickly on the basis of whether they fall above a high criterion or below a low criterion (Atkinson & Juola, 1973, 1974, as cited in Wixted, 2007). According to this model, recollection is only initiated if the memory strength falls between the two criteria, acting, in essence, as a backup plan. Later, Yonelinas (1994, 2001) provided another attempt to combine the two models, with the DPSD model. According to DPSD, recollection is a high-threshold, categorical process; it either occurs or does not. Familiarity, on the other hand, is generally viewed as a continuous, ahistorical memory strength variable, capable of ranging in strength from low to high (Mandler, 1980; Yonelinas, 1994). Familiarity, being described by strength of the signal, is therefore compatible with an equal-variance signal detection model. DPSD differs from the two-criterion model of Atkinson and Juola primarily in its order of operations (and its added quantitative detail). Whereas the two-criterion model assumes that recollection is initiated as a backup for failures of familiarity, DPSD assumes that familiarity-based decisions are initiated following failures of recollection. That is, if recollection does not occur, responses are based on familiarity.

The majority of evidence in favor of separate recollection and familiarity processes comes from functional process dissociations, or manipulations that affect the contributions of each system independently. For example, several studies have indicated that responses based on familiarity are faster than those based on recollection (e.g., Hintzman & Caulton, 1997; Hintzman, Caulton, &

Levitin, 1998). Furthermore, it has been demonstrated that this fast familiarity process leads to increased false alarms immediately after the presentation of a new item, but that the false alarm probability drops off with increased time, reflecting slow recollective retrieval dynamics (e.g., Gronlund & Ratcliff, 1989; Hintzman & Curran, 1994; Jacoby, 1999; McElree, Dolan, & Jacoby, 1999). Using event related potentials (ERPs), many studies have also found distinct electrophysiological correlates for recognition memory responses based on recollection versus familiarity (Curran, 2000; Duarte, Ranganath, Winward, Hayward, & Knight, 2004; Guo, Duan, Li, & Paller, 2006; Klimesch, Dopplemayr, Yonelinas, Kroll, Lazzara, Rohm, & Gruber, 2001) and some have found opposite effects on recollection and familiarity following hippocampal damage (Sauvage, Fortin, Owens, Yonelinas, & Eichenbaum, 2008). These dissociations (and others) strongly suggest that two separate neural substrates underlie recognition memory.

The recent CDP model (Wixted & Mickes, 2010) combines select elements of dual-process theory with key aspects of signal-detection theory, yielding a signal-detection-based model capable of explaining subjective feelings of recollection versus familiarity. According to this model, separate (non-independent) recollection and familiarity components exist, but memory decisions are still based on a continuous stream of evidence, comprised of the additive strength from each component. Perceivers are privy to the predominant source of evidence, either recollection or familiarity, and therefore possess different states of awareness associated with each component (e.g., an item feels familiar if the predominant source of evidence is the familiarity component). Evidence in favor of this model primarily comes from the Remember/Know task (see below).

Model Comparison

Although there are many methods by which to compare and contrast single- and dual-process theories, two of the most commonly-used methods are analysis of receiver operating characteristic (ROC) curves, and the Remember/Know paradigm (Tulving, 1983). Evaluating the shapes of ROCs (and of their *z*-transformed counterparts, *z*-ROCs) has informed the dual- versus single-process debate for years, leading to refinement of theories and the development of newer models (e.g., DPSD and CDP).

Stated broadly, ROC curves are plots of the cumulative hit rate versus the cumulative false alarm rate at various levels of bias or confidence (Macmillan & Creelman, 2005; Yonelinas & Parks, 2007). When examined in normalized space, *z*-ROCs represent the ratio of the standard deviation of the foil distribution to the standard deviation of the target distribution ($\sigma_{\text{foil}}/\sigma_{\text{target}}$). Recall that most single-process views of memory adopt the unequal variances assumption. Specifically, they predict that the standard deviation of the target distribution will be 1.25 times that of the foil distribution, yielding a *z*-ROC slope of 0.80. Whereas models with a high-threshold component (i.e., models predicting all-or-none memory) predict linear ROCs and curvilinear *z*-ROCs, single-process models predict curvilinear ROCs and linear *z*-ROCs (Wixted, 2007). This curvilinear prediction is directly related to the assumption of a continuous memory process; accurate recognition memory can be observed at various levels of confidence, which is graphically represented by a curvilinear function. On the other hand, if recollection is an either/or categorical process, memory retrieval will only be successful at the highest level of confidence or bias, yielding a linear function in standard plots and a curved function in *z*-space.

Overwhelming support for the ROC predictions of signal-detection models, relative to high-threshold models, has resulted in near abandonment of high-threshold models (see Yonelinas & Parks, 2007; Wixted, 2007). Although dual-process theories still incorporate the high-threshold assumption for recollection, they are able to make distinct ROC predictions because of the signal-detection familiarity process. Manipulations thought to affect recollection and familiarity independently have been shown to influence the shape of the ROC, which is commonly interpreted as evidence for dual-process theories (Yonelinas & Parks, 2007). Additionally, when recollection and familiarity are put in opposition to one another (as in the exclusion condition of Jacoby's (1991) *process-dissociation procedure*), ROCs are curvilinear and negative in slope, while *z*-ROCs are linear (although they take on a pronounced, inverted U-shaped function when recollection is more prominent; see Yonelinas, 1994; Yonelinas, Regehr, & Jacoby, 1995).

Other research, however, supports a single-process signal-detection account of memory, with more evidence in support of UVSD models over equal-variance signal-detection models, because equal variance models, like high-threshold models, predict linear ROCs with a slope of 1.0. Curvilinear ROCs have been repeatedly observed throughout the literature (Glanzer, Kim, Hilford, &

Adams, 1999; Ratcliff, Sheu, & Gronlund, 1992; Wixted, 2007), and UVSD has been successfully extended to describe performance in several recognition memory paradigms (Glanzer, Hilford, & Kim, 2004; Hilford, Glanzer, Kim, & DeCarlo, 2002). In fact, as summarized by Wixted (2007), every memory study between the years of 1958 and 1997 yielded curvilinear ROCs. Further, meta-analyses of confidence-based ROCs demonstrate that *z*-ROCs are typically linear, with slopes of less than 1.0, and that the slopes increase as performance decreases (Glanzer et al., 1999; Ratcliff et al., 1992). This is important for single-process models that adopt the assumption of unequal variances, because a common finding is that the slope of the *z*-ROC is 0.80 (Ratcliff et al., 1992), supporting the notion that the standard deviation of the target distribution is 1.25 times that of the foil distribution.

A second common method used to adjudicate between single- and dual-process theories is the *Remember/Know* (RK) paradigm, which makes use of subjects' subjective feelings of the relative specificity of their memory. Although Tulving (1985) initially intended for the procedure to differentiate between states of awareness associated with subjective experiences of memory, it has more recently been used to support dual-process theories of recognition memory (see Wixted & Mickes, 2010). In this paradigm, participants are assumed to appreciate why they make old/new recognition decisions. If an item is judged "old," there are three possible routes to this decision, an item can be "remembered," "known" (henceforth R and K) or simply guessed (although not all studies use a "guess" option). Jacoby, Yonelinas, and Jennings (1997) suggested that remember responses reflect episodic retrieval, and the function of conscious recollection, whereas know responses reflect familiarity, or the recognition of an item's status as "old" without concomitant recollection of its earlier presentation (see also, Yonelinas, 2002).

Results from several studies (e.g., Engelkamp & Dehn, 1997; Gardiner, 1988; Gardiner, Java, & Richardson-Klavehn, 1996; Gardiner & Java, 1991; Gardiner & Parkin, 1990; Rajaram, 1993) indicate that different experimental manipulations selectively enhance or diminish remember/know response frequencies, supporting Gardiner's (1988) reports of functional dissociations between R and K responses. These dissociations imply that recollection and familiarity are distinct, independent processes, and their existence is typically interpreted within a dual-process framework (Gardiner, 2001, as cited by Dunn, 2004).

This interpretation has been criticized, however, as several researchers have proposed that the response types reflect *confidence* in memory more than they reflect the function of two separate memory processes (Donaldson, 1996; Hirshman, 1998). Instructions to respond "remember" or "know," according to this view, are interpreted by participants as a requirement to adopt a more conservative or liberal response criterion, respectively (Dunn, 2004). Donaldson (1996) approached the RK task from a single-process viewpoint, and suggested that participants complete the task by adopting two decision criteria, one (high criterion) for remember responses and one (low criterion) for know responses. He argued that, although participants were issuing responses that appeared to reflect two different memory systems, they were responding in line with their decision criteria, which reflected memory strength, and not two different processes (cf. Knowlton & Squire, 1995). Meta-analyses of decades of data and critical tests of recent data have been taken to support both single-process (Donaldson, 1996; Dunn, 2004, 2008; Hirshman & Master, 1997; Wixted, 2007; Wixted & Stretch, 2004) and dual-process models (Conway, Dewhurst, Pearson, & Sapute, 2001; Gardiner, Ramponi, & Richardson-Klavehn, 1998, 2002; Yonelinas, 2002).

The recent hybrid CDP model (Wixted & Mickes, 2010) was developed in response to the dual-process interpretation of RK data. Although the model is entirely based upon signal-detection theory, it assumes that participants are able to determine whether their memorial experiences are based on the function of recollection or familiarity. In short, CDP assumes that recognition decisions reflect the combined influence of recollection and familiarity, such that decisions are still based on a ‘strength of evidence’ dimension, but that participants have access to the source of the predominant strength. Evidence for this model comes from RK tasks in which participants provide both confidence estimates and source discriminations. During encoding, participants study words presented at either the top or bottom of the screen, in red or blue font. When prompted to retrieve an item (with a centrally-presented, black-ink item), participants provide RK judgments, confidence estimates, and responses regarding original location and color (source discriminations). Wixted and Mickes found that, although “know” responses can be associated with a high degree of confidence, corresponding source accuracy was lower, relative to “remember” responses.

As discussed above, neuropsychological evidence has consistently been interpreted within a dual-process framework, and this is largely because of the use of RK in neuroimaging studies. In many experiments, researchers have observed elevated activity in the hippocampus for R judgments, and almost no activity for K judgments (e.g., Aggleton, Vann, Denby, Dix, Mayes, Roberts, et al., 2005; Eldridge, Engel, Zeineh, Bookheimer, & Knowlton, 2000; Holdstock, Mayes, Gong, Roberts, & Kapur, 2005; Moscovitch & McAndrews, 2002; Uncapher & Rugg, 2005; Verfaellie, Rajaram, Fossum, & Williams, 2008; Yonelinas, Kroll, Quamme, Lazzara, Sauve, Widaman et al., 2002). One difficulty in interpreting these effects, however, is that R judgments are typically associated with higher confidence, relative to K judgments (e.g., Dunn, 2004, 2008; Rotello & Zeng, 2008; Wixted & Mickes, 2010; Wixted & Stretch, 2004), making it impossible to determine whether the effect arises from differences in fundamental processes, or differences in retrieved strength. This "strength confound" was originally pointed out by Wixted (2009), who noted that recollective detail is almost never entirely absent from know judgments (see Mickes et al., 2009; Wais, et al., 2008). In fact, when subjective retrieval strength was equated, Wais, Squire, and Wixted (2009) observed similar levels of hippocampal activity during putatively recollection-based and familiarity-based memories (see also Kirwan et al., 2008; Wais, 2008; Wais, Wixted, Hopkins, & Squire, 2006). In order to interpret RK data, it seems, one needs to take an additional step and collect overt, metacognitive estimates of confidence.

The experiments reported here assessed the characteristics of recognition memory using ROC analyses and the *source-monitoring* framework. Whereas the RK paradigm relies upon participants' subjective states of awareness, source-monitoring paradigms allow researchers to estimate the degrees of specificity in retrieved memories, particularly when paradigms are designed so as to allow multiple sources (or dimensions) of information to be recalled.

Source Monitoring

To this point, the discussion of memory has focused solely on *item memory*, the ability to recognize that information has been previously encountered (see Malmberg, 2008, for a review). But, as the RK procedure makes apparent, memory is typically more elaborate than merely recognizing previously-acquired information. Often, people are able to identify the characteristics of the learning event, such as who taught it, or where it was learned. This is akin to remembering both that you were promised a raise (item memory), and that your immediate supervisor was the one who promised it (source memory). The *source monitoring framework* (Johnson, Hashtroudi, & Lindsay, 1993), an extension of the reality-monitoring framework (Johnson & Raye, 1981), was developed to assess mnemonic decision processes in the retrieval of specific details from memory. During standard source memory experiments, participants' memory for the qualitative characteristics (e.g., perceptual, contextual, semantic, and affective details) of the encoding event are queried (often following an assessment of item memory, but see DeCarlo, 2003, for an exception). In addition to the perceptual, contextual, semantic, and affective details that may serve as diagnostic source information, sources can also include the cognitive operations performed during the course of encoding. Although they share many characteristics, the source monitoring framework differs from reality monitoring in that the source monitoring framework considers the underlying decision processes of source recollection (Johnson et al., 1993).

According to Johnson et al. (1993), source monitoring decisions are based on the richness of the memory trace. For example, whereas memories for imagined events are likely to be characterized by an emphasis on cognitive operations, memories for performed actions are more likely to be characterized by detailed perceptual, temporal, and spatial information, with less emphasis on cognitive operations. This information is not necessarily the product of a consciously-controlled search process during remembering; often source information accompanies memories automatically. When source memory decisions are more deliberate (i.e., when source information does not immediately accompany a memory), they are often slow and characterized by logical assessment (Johnson et al., 1993). Regardless of the method by which they are retrieved, the source details must be weighed against decision criteria (or criterion, depending on method). If enough details are retrieved to support one source judgment over another, then the item is ascribed to that particular source (Johnson et al., 1993). Depending on methodological changes or individual differences, source decisions can be made more or less quickly, and with more or less deliberation (see Dodson & Johnson, 1993; Johnson & Raye, 1981; Marsh & Hicks, 1998).

One of the most debated questions in source monitoring regards whether the paradigm reveals the functioning of a separate, threshold recollection process. Clearly, retrieving source content involves recollecting some form of detail, leading several researchers to suggest that source monitoring critically involves an independent recollection process (Guttentag & Carroll, 1997; Perfect, Mayes, Downes, & Van Eijk, 1996; Quamme, Frederick, Kroll, Yonelinas, & Dobbins, 2002; Yonelinas, 1999). Demonstrations of partial source memory (e.g., Bink, Marsh, & Hicks, 1999; Dodson, Holland, & Shimamura, 1998; Hicks, Marsh, & Ritschel, 2002), and of preserved source memory in the absence of item memory (e.g., Cook, Marsh, & Hicks, 2006; Starns, Hicks, Brown, & Martin, 2008), have recently been taken to suggest that source memories are characterized by retrieval from a continuous (as opposed to all-or-none) recollection process. By this logic, any degree of recollection can support any response to an assessment of item memory (although stronger evidence should still naturally yield higher item accuracy and better source memory). Alternatively, if recollection is conceived of as a threshold process, failing to accumulate enough evidence to support the more fundamental item memory response should not result in full or partial recollection of source details, even if one assumes that partial source recollection follows a threshold process operating on an imperfectly encoded memory trace (e.g., Parks & Yonelinas, 2007).

In standard source memory experiments, participants give a source judgment following “old” item responses (or, in some cases, they respond “Source 1,” “Source 2,” “new”). Unless the participant claims to have studied the item, they are not asked to make a source judgment. This practice reflects an assumption inherent in multinomial models of source memory, which have been used to describe the relationships among various components of item and source memory decisions (e.g., Batchelder & Riefer, 1990): Specifically, source discrimination is not possible for items that are not first recognized. Applying such models to source memory requires the assumption of discrete cognitive states (e.g., you either recognize an item with probability *D*, or you fail to recognize it, with probably 1 – *D*; see Bayen, Murnane, & Erdfelder, 1996; Dodson et al., 1998; Dodson & Shimamura, 2000; Meiser & Bröder, 2002). These models are consistent with the assumption of an all-or-none recollection process.

Recently, models have been proposed which loosen the all-or-none assumption, and instead suggest that source retrieval follows a continuously-distributed, signal-detection process (e.g., Banks, 2000; DeCarlo, 2003; Glanzer, Hilford, & Kim, 2004; Hilford, Glanzer, Kim, & DeCarlo, 2002; Qin, Raye, Johnson, & Mitchell, 2001; Slotnick & Dodson, 2005; Slotnick, Klein, Dodson, & Shimamura, 2000; Starns et al., 2008). Such models propose the existence of separate strength distributions for targets from each of the studied sources, and a distribution for new items. Starns and colleagues (2008) provided compelling evidence for the existence of a signal-detection process by demonstrating that their multivariate signal detection model adequately described the observation of accurate source memory in the absence of accurate item memory. According to their model, for each item presented on a recognition test, participants retrieve a certain amount of evidence that the item was studied and a certain amount of evidence that the item belongs to one source or the other. Even if item strength fails to surpass the recognition criterion, the source evidence may still surpass a (liberal) source criterion. Cook, Marsh, and Hicks (2006) reported a similar finding; even when participants could not produce studied items on cued recall, they could nonetheless demonstrate accurate source discrimination.

The notion that source memory is not an all-or-none process is supported by findings of source memory in the absence of item memory, and also by demonstrations of partial source recollection. Dodson et al. (1998) observed that, in the absence of correct source discriminations, participants can base source judgments on partial evidence. In their study, participants studied words spoken by one of four speakers, either all males or half males, each identified by face and name. During recognition, participants first gave old/new responses to printed items. Contingent upon "old" responses, participants identified the original source (i.e., speaker name) of the learned item. By examining response frequencies, Dodson et al. (1998) determined that source memories were *specific* or *partial*. Specific source memories were observed when participants correctly identified the original speaker, and partial source memories were observed when participants made within-gender source misattributions. Critically, these within-gender source misattributions were made with above-chance frequency, reflecting partial evidence supporting source judgments. This finding provides strong evidence in favor of a continuous source recognition process. Consistent evidence has also been observed in the absence of item recognition: When people are unable to retrieve studied items, they are nevertheless able to provide partial information about the source of the material (e.g., Koriat, Levy-Sadot, Edry, & de Marcas, 2003; Kurilla & Westerman, 2010).

The well-known tip-of-the-tongue (TOT) phenomenon also suggests that recollection is graded. In a TOT state, people cannot retrieve sought-after information from long-term memory, despite being positive that they possess the information (e.g., when you cannot name an actor in a movie, but know that you know his name; see Brown & McNeill, 1966; Schwartz, 2002). Although Google has helped to eliminate the struggle to internally resolve TOTs, they are still often characterized by the ability to recall partial information. For example, in the example just given, you may remember that the actor has a weird name with a K in it, and that he was in the show ‘24,’ but you may be unable to resolve that you are thinking of Kiefer Sutherland. Extensive evidence has revealed that TOT states involve recollection of partial details of the encoding event (see Maril, Simons, Weaver, & Schacter, 2005, for a review). These details include the first and last letters of the word (Koriat & Lieblich, 1974), the number of syllables (Rubin, 1975), and synonyms for the word (Cohen & Faulkner, 1986), among others. Critically, although these recollections are clearly only partial, they are typically accompanied by high confidence that the item will be retrieved. A threshold-based recollection process would have difficulty explaining these phenomena.

The necessity of an independent recollection process during source retrieval has often been debated by combining source monitoring with the RK procedure. Recall that many researchers suggest that R responses reflect recollection (e.g., Jacoby et al., 1997; Yonelinas, 2002; but see Wixted & Mickes, 2010). If this is true, then researchers should observe similar effects of various manipulations on the rates of “remembering” and retrieving accurate source information. For example, encoding manipulations that enhance source recognition accuracy also increase the frequency of R responses (Conway & Dewhurst, 1995; Dewhurst & Hitch, 1999; Donaldson, MacKenzie, & Underhill, 1996). In a further demonstration of the similarity of the two processes, Rugg et al. (1998) observed that the pattern of electrophysiological brain activity observed during R responses is the same as the pattern observed during successful source retrieval. Abundant evidence also suggests that source retrieval is more successful following R, relative to K, responses (Dewhurst & Hitch, 1999; Meiser & Bröder, 2002; Perfect, Mayes, Downes, & Van Eijk, 1996; Wixted & Mickes, 2010). On the other hand, however, inconsistent data have been observed by researchers who have found that source memory is just as, if not more, accurate following K responses, relative to R responses (Conway & Dewhurst, 1995; Hicks, Marsh, & Ritschel, 2002). In fact, when Rotello, Macmillan, Reeder, and Wong (2005) instructed participants that they might need to justify their R responses (e.g., with recollected details), the frequency of R responses decreased. R response frequencies were also influenced by bias manipulations, suggesting that they are not process-pure reflections of an all-or-none recollection process.

The debate over the processes that support episodic memory, recollection and familiarity, can be addressed through the source monitoring paradigm. Although the results have been equivocal, with respect to supporting one theory over another, single- and dual-process theories make clear predictions about the role of recollection in source retrieval. Whereas dual-process theorists have argued that threshold processes can account for partial source recollection (by assuming that only some proportion of information about the encoded event surpasses threshold; Parks & Yonelinas, 2007), this explanation is unsatisfying, because responses based on all-or-none recollection should be accompanied by high confidence in the decision, which is not consistently observed (Wixted, 2007). As with item memory, source memory is amenable to analysis via ROC curves, because single- and dual-process theories make competing predictions regarding their linearity or curvilinearity. Linear ROCs (with curved *z*-ROCs), which imply a threshold recollection process consistent with dual-process theories, have been observed repeatedly (Parks & Yonelinas, 2007; Yonelinas, 1999; Yonelinas & Parks, 2007). Yet curved ROCs (with linear *z*-ROCs), consistent with single-process, signal-detection views, have been observed across many other studies (e.g., Dodson, Bawa, & Slotnick, 2007; Glanzer, Hilford, & Kim, 2004; Hilford, Glanzer, Kim, & DeCarlo, 2002; Qin, Raye, Johnson, & Mitchell, 2001; Slotnick & Dodson, 2005; Slotnick, Klein, Dodson, & Shimamura, 2000). Attempts have been made to reconcile these disparate findings by appealing to a “unitized familiarity” process, wherein item and source information are contextually bound during encoding, yielding source judgments based on familiarity (see Diana, Yonelinas, & Ranganath, 2008). Most evidence, however, has suggested that a continuous recollection process is the more plausible account for the findings. For example, Mickes, Johnson, and Wixted

(2010) observed that increases in presentation frequency (a manipulation assumed to influence recollection) were associated with increases in ROC curvilinearity, the frequency of R responses, and cued recall accuracy. This relationship suggests that recollection plays a critical role, not unitized familiarity.

To address the single-/dual-process debate, the current experiments adopted a modified source monitoring paradigm from Dodson et al. (1998), aimed at investigating the existence of partial source recollection. Recall that dual-process theories do not predict partial source retrieval, as recollection is a threshold process. By monitoring eye movements, a commonly used index of attention, as participants encoded material, I was able to guard against any overt differences in attention across trials. While recording eye movements, pupil size measures were also continuously recorded, which allow estimates of memory strength without soliciting overt metacognitive estimates of memory strength or confidence (see below).

Eye Movements in Memory

Although saccadic eye movements are central to visual processing (Ross & Ma-Wyatt, 2003), they account for only a small portion of eye movements that people generate. Further, saccadic eye movements are unique among other types of eye movements (e.g., smooth pursuit, vergence, vestibular) in that they are the only type to direct gaze toward new sources of visual information (Richardson, Dale, & Spivey, 2007). In his classic work on eye movements, Yarbus (1967, as cited in Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010) demonstrated that different task instructions (e.g., free-viewing versus making an inference about the scene) yielded different patterns of viewing behavior as people inspected visual scenes; eye movements were guided by strategic cognitive processes, and not solely by bottom-up perceptual influences. Since that time, researchers have shown that saccadic eye movements are influenced by top-down processes, or the expectations of the perceiver (Godijn & Theeuwes, 2002), and ongoing cognitive activity (e.g., Ferreira, Apel, & Henderson, 2008). Saccadic eye movements have been used to document the time course of cognitive processing across domains, including speech perception (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), face perception (e.g., Goldinger, He, & Papesh, 2009; Henderson, Williams, & Falk, 2005; Mäntylä & Holm, 2006), reading (Rayner, 1998; Reichle, Warren, & McConnell, 2009), and scene processing (e.g., Althoff & Cohen, 1999; Henderson, Weeks, & Hollingworth, 1999; Parker, 1978). In many cases, however, these eye movements occur in the absence of visual information. Such non-visually guided saccades are often directed to spatial locations that once contained relevant information, but are now empty (e.g., Edelman & Goldberg, 2001). These non-visual gaze patterns (NVGPs), which include saccades and fixations (Micic, Ehrlichman, & Chen, 2010), are not initiated for the benefit of social communication: NVGPs occur during phone conversations (Beattie & Barnard, 1979), in darkness (Ehrlichman & Barrett, 1983), and when the eyes are closed (Ehrlichman, Micic, Sousa, & Zhu, 2007).

Although the neural locus of NVGPs remains obscure (although see Micic et al., 2010), they appear closely tied to memory retrieval. Early evidence for this comes from Bergstrom and Hiscock (1988), who observed higher eye movement rates during tasks that involved greater memory search (e.g., unconstrained tasks, “name a five-letter word with three consonants and two vowels”), relative to tasks involving less memory search (e.g., vowel counting). Subsequent work by Ehrlichman and colleagues (Ehrlichman et al., 2007; Micic et al., 2010) confirmed this observation: Participants engage in far more NVGPs during tasks requiring extensive memory search and access to long-term memory (e.g., recalling lists of words, naming synonyms), relative to less long-term memory demanding tasks (e.g., reciting the alphabet, n-back). These eye movements are not sensitive to variations in levels of processing, but occur during all forms of long-term retrieval (Micic et al., 2010). Further, word recall performance is unaffected by suppression of NVGPs, suggesting that these eye movements are not functional for retrieval (but see Lyle, Logan, & Roediger, 2008[1]).

Other work on patterns of motor behaviors and their relation to memory suggests that motoric and spatial processing are closely related, and that repetition of a motor program (behaved or observed) facilitates recognition memory (Engelkamp, Zimmer, Mohr, & Sellen 1994; Fendrich, Healy, & Bourne, 1991; Fendrich, 1998; Zimmer, 1998). Research on the rate of NVGPs (e.g., Ehrlichman et al., 2007; Micic et al., 2010) has not closely examined the direction, or target, of saccadic eye movements; instead, such research is concerned with the

---

[1] Lyle and colleagues (Lyle et al., 2008a, b; Lyle & Jacobs, 2010) have documented ‘saccade-induced retrieval enhancement,’ whereby fast bilateral saccades are said to enhance memory retrieval by increasing inter-hemispheric processing in the intraparietal sulcus. This brain region has not been implicated in the phenomena under investigation in the current research, and is inconsistent with research on eye movement desensitization techniques in people with post-traumatic stress disorder (e.g., van den Berg & van der Gaag, 2012). Further, data from our laboratory has failed to document the effect in hundreds of participants, casting doubt on the generality of the effect.

frequency of saccades in the absence of visual information. Although research on NVGPs clearly establishes a link between saccadic eye movements and memory processes, one of the key questions addressed in the present research is whether NVGPs can reflect the *reinstatement* of oculomotor behaviors, and whether that reinstatement has a functional role for memory retrieval. To date, research on this topic has been equivocal.

Abundant research has demonstrated that patterns of eye movements tend to be reinstated across learning and retrieval (see Ballard, Hayhoe, & Pelz, 1995; Holm & Mäntylä, 2007; Richardson & Spivey, 2000; Ryan, Hannula, & Cohen, 2007; Spivey & Geng, 2000; Yarbus, 1967, as cited in Tatler et al., 2010), but only a few studies suggest a functional role for these reinstated eye movements (see Ferreira et al., 2008 for a review). Those that do suggest variants of Noton and Stark's (1971) "scanpath hypothesis," noting that if the position of the eyes is assumed to represent a self-generated context during encoding, then reinstating this context during test may facilitate memory. This is essentially the oculomotor equivalent to myriad demonstrations of context-dependent memory (e.g., Godden & Baddeley, 1975; Winograd & Church, 1988), wherein reinstated contexts, either external or self-generated, facilitate memory. Although some researchers have documented evidence against this hypothesis, such that sampling behaviors (i.e., fixations, fixation durations, etc.) *decrease* as a function of familiarity or memory (Althoff & Cohen, 1999, but see Melcher & Kowler, 1999), others find evidence that increased frequency of saccades reflects memory. For example, saccades are automatically directed to the location of changes in a visual scene, even in the absence of conscious awareness of the change (Parker, 1978; Ryan & Cohen, 2004a, 2004b; Smith & Squire, 2008). The question remains, however, whether reprocessing or reinstated eye movements ever facilitate memory retrieval.

Whether or not eye movements facilitate memory, it is clear that they at least *reflect* memory. Several researchers have found evidence that fixation patterns on a blank screen during retrieval mimic those during encoding (e.g., Ballard, Hayhoe, & Pelz, 1995; Gnadt, Bracewell, & Andersen, 1991; Johansson, Holsanova, & Holmqvist, 2006; Laeng & Teodorescu, 2002). Spivey and Geng (2000), for example, presented participants with four colored shapes in the four corners of a 3 x 3 display, followed by a blank screen. Moments later, a display was presented in which one of the items was missing and participants were asked a question about the missing item. In 30% - 50% of such trials, participants made a saccade to the location in which the missing item had been presented. This finding, that memory retrieval automatically accessed the spatial location from encoding, and yielded a saccade to that location, was also observed when participants in another study learned semantic information presented in one of four screen quadrants (the “Hollywood Squares” paradigm, see Richardson & Spivey, 2000). During this experiment, participants learned factual information auditorily, while a visual stimulus (a face or a spinning cross) was displayed in one of the four quadrants. When subsequently tested on the studied information in front of a blank display, participants’ spontaneously directed their gaze to the quadrant “where the information was learned” significantly more often than they directed it to other quadrants. Despite this finding, and others like it (see Hoover & Richardson, 2008), in no case did the researchers find that saccades to formerly relevant locations facilitated memory retrieval. In fact, Laeng et al. (2007) observed a similar effect in amnesic patients; patients spontaneously refixated previously viewed locations during a memory test, in the absence of overt memory for the encoding event. The lack of this effect has led researchers to conclude that eye movements, and the reinstatement of their patterns, are obligatory effects in recognition memory, but that they are not necessarily functional (Ryan et al., 2007).

Further evidence for eye movement-based memory effects comes from studies of relational memory and scene processing. For example, Hannula and colleagues (Hannula, Ryan, Tranel, & Cohen, 2007; Hannula & Ranganath, 2009) had participants learn face-scene pairs by superimposing individual faces upon specific scene images. Later, when shown a preview of a previously-learned scene prior to the onset of a 3-face lineup containing the learned associate face, eye movements were preferentially drawn to the originally-paired face within 500 ms of its onset. This effect occurred even in the absence of accurate behavioral selection of the originally-paired face (Hannula & Ranganath, 2009; Hollingworth et al., 2001). More recently, Hannula, Baym, Warren, and Cohen (2012) demonstrated that this effect is robust to manipulations of similarity, which are known to negatively affect behavioral performance. Participants in their study encoded target faces, and were later asked to pick studied faces out of 3-face lineups consisting of foil faces. Foils were either highly similar or less similar to the target face. When the distorted faces were highly similar to the targets, behavioral performance suffered. Eye movements, however, were impervious to the similarity manipulation. Regardless of the behavioral response, fixations were preferentially directed to studied faces within 1000-ms of the onset of the lineup. Again, however, there was no evidence for a facilitative effect of eye movements on memory.

Eye movements may reveal more about memory than just whether or not someone has previously experienced something. In a study on the reprocessing effect, refixations across study and test were associated with increased probability of retrieval (Holm & Mäntylä, 2007; Mäntylä & Holm, 2006). Mäntylä and Holm (2006) conducted a series of face perception and recognition experiments wherein participants' eye movements were controlled via gaze-contingent displays during study and/or test. Using the RK procedure, Mäntylä and Holm (2006) observed that the frequency of R responses was significantly diminished when eye movements were constrained during either study or test, but that K responses were unaffected. Additionally, relative to K responses, R responses were associated with more refixations across encoding and retrieval, suggesting that recollection, or at least strong memory strength, is related to the degree to which saccades are reinstated across learning and retrieval. A similar pattern was observed with refixations during complex scene processing (Holm & Mäntylä, 2007).

Additional evidence in favor of a functional role for eye movements in memory retrieval comes from a study of mental imagery and memory (Laeng &

Teodorescu, 2002). Participants in this study examined modified checkerboard-type patterns and were later asked to imagine them while looking at a blank screen. Across encoding and imagining, participants' eye movements were highly correlated. Even when required to fixate centrally during encoding, participants spontaneously fixated centrally during mental imagery. Further, participants' memory for the patterns was related to their ability to reinstate their eye movements. Participants who were allowed to freely move their eyes during encoding, but were required to fixate centrally during retrieval, showed a decreased ability to recall the studied pattern, relative to participants who were allowed to move their eyes during retrieval. In short, the debate surrounding the functional role of eye movements in memory retrieval is currently unresolved, and the evidence for such a role has been equivocal.

Better understanding the (putatively) functional role of eye movements in memory retrieval will help address a theoretical debate relevant to students of memory. Whereas some researchers (e.g., Hoover & Richardson, 2008; Richardson & Spivey, 2000) interpret reprocessing effects in terms of O'Regan's (1992) concept of the world as an "external memory store," others (e,g., Ferreira et al., 2008; Laeng & Teodorescu, 2002) view reprocessing effects as evidence of rich, multi-modal memory traces, capable of enhancing memory retrieval. According to the 'external memory store' view, the constancy of the visual world is a perceptual illusion; eye movements are made to new locations in order to "access" this store, and to avoid relying upon internal memory. As quoted by O'Regan (1992, p. 463), "There is no need for an internal representation that is a faithful metric-preserving replica of the outside world in the head." The ideas are very much in line with ecological scientists (e.g., Gibson, 1950, 1966; Turvey, 1977), and with recent suggestions within the field of embodied cognition that cognition and action are intimately related (e.g., Barsalou, 1999; Glenberg, 1997; Wilson, 2002). By this view, saccades are directed to once-occupied locations because there is no internal representation of what was once there; the visual system directs an eye movement to accumulate information, as it normally does during standard perception.

On the other hand, researchers have argued that eye-movement-based memory effects reflect integrated memory representations consisting of the spatial, verbal, auditory, and action-based information present during encoding (Ferreira et al., 2008). According to this view, reactivating any part of that multi-modal episodic trace enhances the probability that other parts of that trace will also become active. This is similar to Hintzman's (1986, 1988) MINERVA 2, such that a memory probe will activate similar traces in episodic memory, eventually resulting in the retrieval of the most similar trace. Relevant to eye movements, suppose that a person is cued to recall an item that was encoded while her eyes were moving around the environment. The retrieval cue partially activates the episodic representation of that item, which consists, among other things, of the oculomotor behaviors present during encoding. Because retrieving one aspect of the episode increases the probability of retrieving other aspects, people are more likely to engage in refixations, even in the absence of visual information. This proposal is related to the ideas of 'event files' (Hommel, 2004) and ‘object files’ (Kahneman et al., 1992), which are essentially episodic traces indexed by spatio-temporal information. Substantial evidence suggests that memory retrieval is facilitated when spatio-temporal context is continuous, or maintained (e.g., Gordon & Irwin, 1996; Henderson, 1994; Hommel, 2004).

The present research examined the functional utility of saccadic eye movements during memory retrieval. As noted by Richardson, Altmann, Spivey, and Hoover (2009), in order to determine the functional role of eye movements in memory retrieval, eye movements need to be manipulated at the time of retrieval, and they should have an appreciable effect on behavior. If eye movements are epiphenomenal, and reflect the function of an external memory store, there should be no relationship between refixations during retrieval, and the behavioral manifestation of retrieval. On the other hand, if eye movements are functional for retrieval, they should be associated with some behavioral benefit, either in accuracy, detail, or response time. To fully support the latter position, eye movements that are externally directed to previously-viewed locations should enhance recognition memory (assuming, however, that externally-cued fixations are not disruptive to behavior, as they are in working memory tasks, see Godijn & Threeuwes, 2012). In short, the current experiments examined whether, and to what extent, reinstated eye movement patterns are associated with changes in memory processing.

Pupil Dilation as a Measure of Cognitive Effort

Pupillometry, the measurement of the diameters of the eyes’ pupils, has been used for centuries to examine visual and cognitive processing (e.g., Fontana, 1765). Although it is well-known that the pupils dilate in response to changes in ambient lighting, it is less well-known that the pupils also dilate in response to non-visual stimuli, such as emotions and thoughts (Goldwater, 1972; Loewenfeld, 1993). This distinction characterizes two independent types of pupillary reflex, *tonic* changes, which occur in response to general factors, such as emotional arousal, stress, and anxiety, and *phasic* changes, which occur following the onset of stimuli for cognitive processing (Karatekin, Couperus, & Marcus, 2004). These cognitively-evoked pupillary reflexes occur following inhibition of the parasympathetic nervous system's Edinger-Westphall nucleus (Steinhauer, Siegle, Condray, & Pless, 2004), which is controlled by the locus coeruleus-norepinephrine (LC-NE) system. The LC is a subcortical brain system that contains the noradrenergic system, which is the sole source of the neurotransmitter NE. This system has been shown to play a critical role in the control of attention (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; for a review, see Laeng, Sirois, & Gredebäck, 2012). A role for the LC-NE system in memory consolidation has been determined by documenting LC-NE activity during memory retrieval (Sterpenich et al., 2006) and slow-wave sleep (Eschenko & Sara, 2008). Relevant to the current experiments, the LC-NE system is also critically involved in the pupillary reflex (Koss, 1986). In combined single-cell recording and pupillometry studies with monkeys, researchers have documented a tight correspondence between pupillary reflexes and activity in cells within the LC-NE system (Rajkowski, Kubiak, & Ashton-Jones, 1993; Rajkowski, Majczynski, Clayton, & Ashton-Jones, 2004).

Further neurophysiological evidence for the influence of cognitive processing on pupil size comes from examining the muscles that control pupillary reflexes. Two muscles are known to control pupil dilation and constriction, the dilator and the sphincter; these muscles are differentially affected by activation in the sympathetic and parasympathetic systems (Steinhauer, Siegle, Condray, & Pless, 2004). As discussed above, inhibition of the parasympathetic nervous system has been attributed to dilation resulting from cognitive processing. This system also controls the activity of the sphincter muscle; when the parasympathetic system is inhibited, activity on the sphincter muscle decreases (Steinhauer et al., 2004). These autonomic pathways hold reciprocal connections with the central nervous system (CNS), so it has been suggested that they can modulate, or be modulated by, CNS structures related to cognition (Gianaros, Van der Veen, & Jennings, 2004; Steinhauer, Siegle, Condray, & Pless, 2004). Investigations into the neural mechanisms of successful learning and memory in animals have revealed a close correspondence between accurate performance and the involvement of the autonomic system (Croiset, Nijsen, & Kamphuis, 2000). Such findings are paralleled by recent findings from human experiments, wherein increased autonomic responses (e.g., skin conductance) are positively correlated with memory strength for emotional words (Buchanan, Etzel, Adolphs, & Tranel, 2006). Additionally, stimulation of the vagus nerve (a parasympathetic pathway known to carry signals to the brain) is associated with memory formation and consolidation (Clark, Naritoku, Smith, Browning, & Jensen, 1999).

As noted above, phasic changes in pupil diameter occur following the onset of cognitive processing. These reflexes are observed independently of tonic changes; in dark-adapted conditions, which inhibit the parasympathetic system, the pupils reliably dilate in response to cognitive demand (Steinhauer & Hakerem, 1992), leading them to be referred to as *task-evoked pupillary responses* (TEPRs). Although Hess is often credited for initiating the psychological study of pupillary reflexes (cf., Hess, 1965; Hess & Polt, 1964; Hess, Seltzer, & Shlien, 1965), his research focused almost exclusively on the pupillary reflex as it reflected "emotionality" (Hess, 1965, p. 46), a tonic response. Since then, the "emotional" component of the pupillary reflex and the "cognitive" component have been clearly dissociated (Stanners, Coulter, Sweet, & Murphy, 1979). Kaheneman and Beatty (1966) are best known for initiating interest in TEPRs, and have even suggested that TEPRs reflect a "summed index" of brain activity during cognitive processing. Their early work demonstrated that pupil dilations are time-locked to cognitive processing, and that differences between and within tasks are observable via pupillometry (Kahneman, 1973; Kahneman, Beatty, & Pollack, 1967). For example, in a digit recall task, as participants were given more numbers to retain, their pupils became larger; as the digits were recalled, the pupils constricted with each additional item (Kahneman & Beatty, 1966). Although pupillometry fell out of favor for some time, it has been used to infer cognitive effort in a variety of domains, such as lexical decision (Kuchinke, Võ, Hofmann, & Jacobs, 2007), attention (Kahneman, 1973; Karatekin, Couperus, & Marcus, 2004), word processing (Papesh & Goldinger, 2012), working memory

(Granholm, Asarnow, Sarkin, & Dykes, 1996; Van Gerven, Paas, Van Merriënboer, & Schmidt, 2004), face perception (Goldinger, He, & Papesh, 2009), general cognitive processing (Granholm & Verney, 2004), and memory (Kafkas & Montaldi, 2011; Papesh, Goldinger, & Hout, 2012).

In the study of memory, pupillometry can be likened to ERP waveforms (Beatty, 1982); enlarged pupils are typically associated with increased cognitive demand (Porter, Troscianko, & Gilchrist, 2007). Comparing neurophysiological measures across study and test has been used to differentiate the neural activity associated with subsequently remembered versus forgotten information in both fMRI (e.g., Ranganath et al., 2004) and ERP investigations (e.g., Cansino & Trejo-Marales, 2008; Duarte, Ranganath, Winward, Hayward, & Knight, 2004; Guo, Duan, Li, & Paller, 2006). The logic underlying such studies is that encoding should utilize the same set of processes and neural substrates that are subsequently recruited during successful retrieval. Moreover, the strength and type (e.g., recollection or familiarity) of memory should be observable from different patterns of activation during both encoding and retrieval. In the current investigation, pupillometry was used to examine retrieval effort, acting in place of metacognitive confidence estimates.

Although early TEPR investigations have been criticized on the grounds that now-standard experimental controls were not implemented (Võ et al., 2008), recent work incorporates strict experimental control to eliminate the unwanted influence of tonic reflexes. Because pupils dilate reflexively to changes in luminance, color, or the spatial frequency composition of the visual input, care must be taken to equate, as much as possible, stimulus characteristics in experimental designs that utilize pupillometry (Porter et al., 2007). Porter and Troscianko (2003) identified several methodological approaches that minimize unwanted pupillary reflexes, including the use of relatively low stimulus contrast, avoiding colored stimuli, and using relatively long stimulus exposure durations. Goldinger and Papesh (2012) recently added to this list of constraints by suggesting the use of relatively long (e.g., 1000-ms or more) inter-trial intervals (ITIs) and baseline-correction procedures. Both suggestions guard against *carryover effects*, as when the difficulty of trial *n* influences the waveform of trial *n* + 1. Recent work has taken such precautions into careful consideration, including several relevant studies on the pupillary reflex and memory.

One of the first demonstrations of a long-term memory effect in pupil dilation was observed by Võ and colleagues (2008), who were motivated by the similarity between pupillary waveforms and ERP waveforms, which are known to reflect memorial processes (Dietrich et al., 2000; Johanson et al., 2004). In their study, participants studied a series of positively and negatively valenced words, followed by a speeded old/new recognition test. Although they found effects of emotionality (e.g., better memory for emotional words, smaller pupils to negative, relative to positive, words), the most relevant finding was a “pupillary old/new effect,” wherein pupils were larger during trials leading to hits, relative to correct rejections. The authors interpreted this effect within a dual-process framework (as in Yonelinas, 2001, 2002), suggesting that the increased pupil size observed for hits was directly related to the occurrence of recollection, because it is suggested to be a more cognitively demanding, slow process.

A similar conclusion was drawn by Kafkas and Montaldi (2011), who investigated the nature of incidental memory for images using a modified RK procedure. In their study, participants viewed images without instruction to remember them, and were later given a surprise memory test in which they distinguished memories based on degrees of familiarity (e,g., F1 – F3) or a single recollection response. Although they did not observe an effect differentiating recollection from familiarity, pupil size during encoding differed based on subsequent memory strength: As subsequent memory increased, pupil diameter decreased. These findings are suggestive, but should be interpreted with caution. Kafkas and Montaldi's (2011) procedure eliminated TEPRs during the encoding phase, leaving only tonic changes free to vary. In an investigation of phasic pupillary changes and memory strength, Papesh, Goldinger, and Hout (2012) observed the opposite pattern. After studying a series of spoken words, participants issued old/new memory decisions and overt confidence estimates. Examining pupil size by subsequent memory performance revealed a clear relationship between confidence and pupil size: As subsequent confidence increased, so too did pupil size.

The studies of Kafkas and Montaldi (2011) and Papesh et al. (2012) are difficult to compare, both due to differences in encoding instructions and stimuli. Evidence consistent with Papesh et al. (2012), however, has been documented in other pupillometry investigations using the RK procedure. Otero, Weekes, and

Hutton (2011) observed no difference in pupil size based on remembering or knowing using visual and acoustic materials. Further, this difference was not due to changes in conscious control: Heaver and Hutton (2011) observed pupillary old/new effects regardless of whether participants feigned amnesia or were told to report all items as “new.” Similar effects were reported by Papesh and Goldinger (2011), who found a pupillary old/new effect across study and test presentations of auditory high and low frequency words. Specifically, when participants studied words that were subsequently remembered, the second presentation of the item was associated with enlarged pupils, as compared to subsequently forgotten and new items. This pattern was especially strong for low frequency words, suggesting that the act of remembering, coupled with the cognitive operations usurped in processing low frequency words (see Papesh & Goldinger, 2012), resulted in an overall increase in the cognitive demand of the task.

In the present investigation, pupillometry was used to infer underlying memory strength, in lieu of overt confidence estimates. In earlier, as-yet-unpublished research (Papesh, in prep), I observed a relationship between pupil size and confidence that supported the use of *pupillary ROCs*. In that experiment, participants (n = 17) studied and were tested on words spoken in an unaccented female voice. During test, old/new recognition responses were accompanied by overt confidence estimates. Because of differences in response frequencies (e.g., participants very rarely respond that they are ‘very sure’), the pupil data were examined by three confidence estimates, estimates of 2, 5, and 10 (low, medium, and high, respectively). Figure 2 shows the relationship between pupil size and confidence during test, separately for high- and low-frequency words. This relationship was observed for old and new items at test, and also for items during the study phase (i.e., pupil diameter also accurately tracked subsequent confidence).

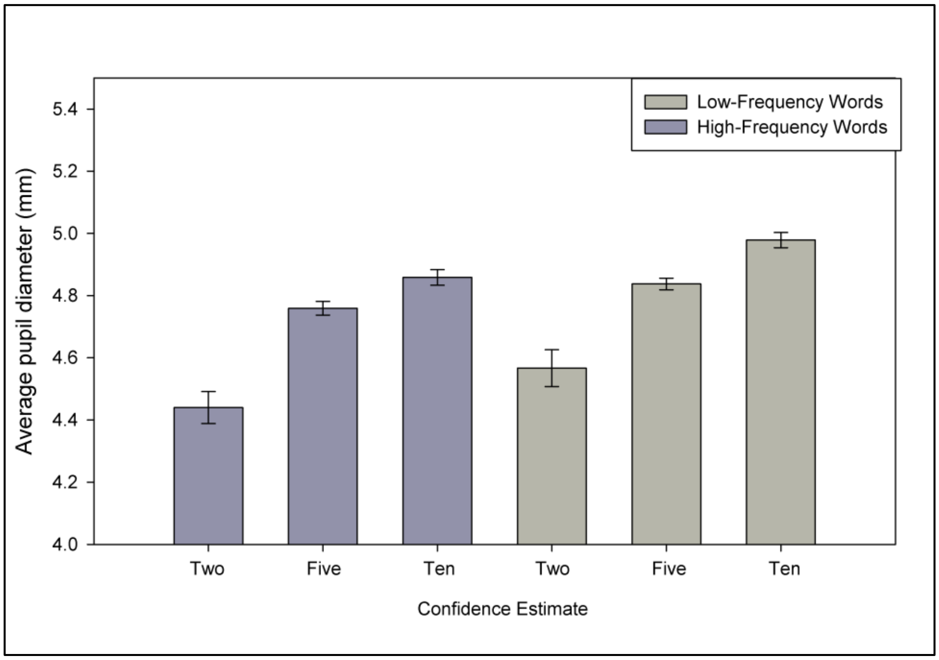


Figure 2. Pupil diameter as a function of confidence and word frequency during recognition test. Error bars represent standard error of the mean. From Papesh (in prep).

To document the feasibility of pupillary ROCs, I constructed two sets of ROC/*z*-ROC pairs, one based on overt confidence estimates and one based on pupillary confidence estimates. To do so, the mean and range of pupil diameters were calculated. This distribution was then segmented into 10 “confidence” bins. Values of 1 and 10 were the extreme upper and lower tails of the distribution, respectively; intermediate values each represented 1/2 of the standard deviation.

Figure 3 displays the behavioral (top row) and pupillary (bottom row) ROCs (left panel) and *z*-ROCs (right panel).

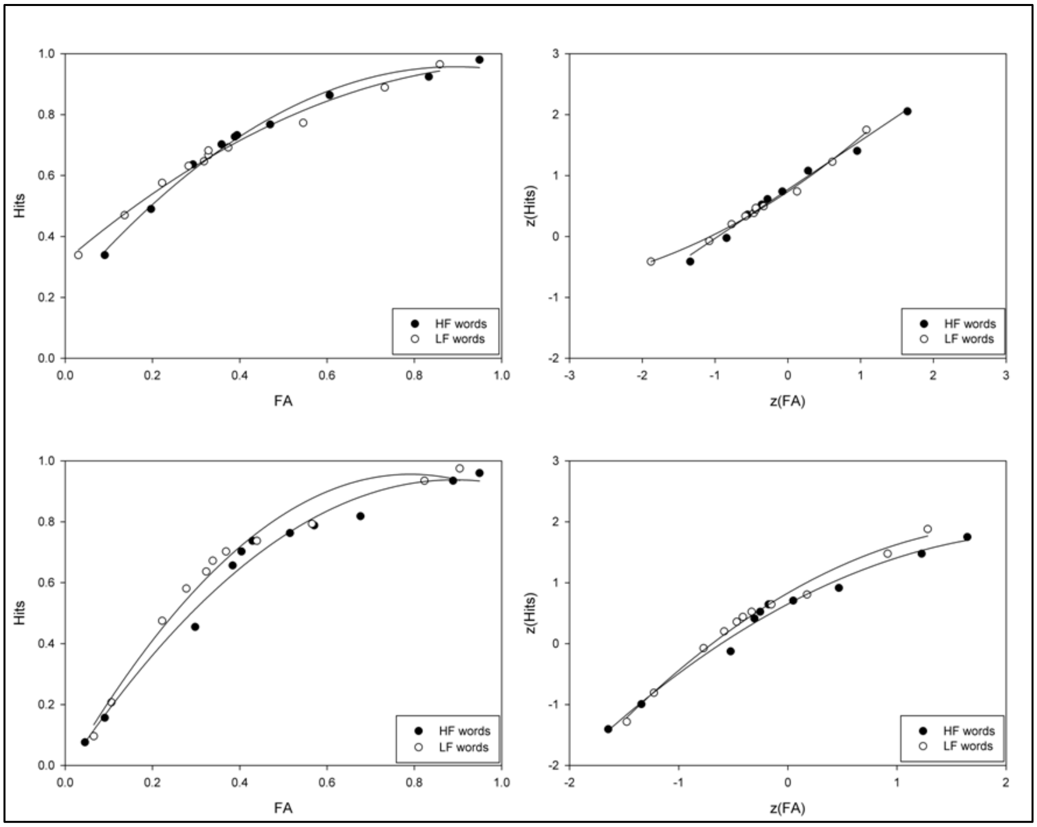


Figure 3. Behavioral (upper graphs) and pupillary (lower graphs) ROCs (left panels) and z-ROCs (right panel) fit with quadratic equations. From Papesh (in prep).

As can be seen by visually comparing the behavioral and pupillary graphs, the pupillary graphs provided a more reasonable description of the data. A common finding in recognition memory is that low-frequency (LF) words are remembered better than high-frequency (HF) words (Glanzer & Adams, 1990). Although the behavioral ROC does not truly capture this, the pupillary ROC depicts lines that are clearly separated for the two word types, suggesting that pupil size provides a sensitive, and accurate, estimate of memory strength.

The Current Experiments

The current experiments build upon previous work, addressing several theoretical issues of episodic memory. Specifically, the present experiments examined audiovisual source memory via behavioral and physiological indices. The research presented here was designed to address four research aims. Research Aim 1 was to examine the influence of source similarity on item and source memory judgments. Specifically, the voices used to present the material had a measured similarity relationship, both to one another and to the (different) voices used to present the test material. If memories are richly-detailed, and if these details are brought to mind automatically during memory retrieval, participants' item and source memories should be more accurate during "low," relative to "high," similarity conditions. Research Aim 2 was to assess the degree of specificity in source memory. As in Dodson et al. (1998), participants in the current experiments could have specific or partial source memories. Unlike Dodson et al., specific and partial source memory were also examined during failures of item memory. As discussed above, dual-process theories have trouble predicting both partial source memory and the existence of source memory in the absence of the more fundamental item memory. Single-process views, particularly multivariate signal-detection models (e.g., Starns et al., 2008), are capable of predicting and explaining such findings. To further examine the nature of recognition memory, each of the following experiments was accompanied by

ROC analyses, using pupil diameter as an index of “confidence” (Research Aim 3). Whereas dual-process theories predict linear ROCs and curvilinear *z*-ROCs, single-process views predict curvilinear ROCs and linear *z*-ROCs. Lastly, by manipulating eye movements across learning and retrieval, I will provide evidence that eye movements are spontaneously reinstated across study and test, but that externally cueing those eye movements does not necessarily improve memories (Research Aim 4), which has implications for two theories of eye movement behavior during memory retrieval.

General Method

Overview

The experiments reported here shared several methodological characteristics, and only differed in the testing procedures. Each experiment involved identical encoding procedures, and four between-subjects conditions. The conditions differed in the similarity of the voices used to present the study items (see **Materials**), and were broadly defined as “high similarity” (HS) and “low similarity” (LS), as in Dodson, Holland, and Shimamura (1998). Within each HS and LS group are the two conditions, HS1, HS2, LS1, and LS2. These are described below.

Materials

Ten speakers (five male) with no discernible accents or speech errors (e.g., lisps or pronunciation errors) recorded a list of 48 questions and 1100 high and low frequency words, from which, 80 words were selected for the familiarization task (see below) and 136 were selected for the main experiment (although word frequency was not intentionally manipulated, see Table 1 for a summary of lexical characteristics, from Balota, Yap, Cortese, Hutchinson, Kessler et al., 2007; stimuli are reported in Appendix A). Eighteen volunteers (none of whom participated in any of the experiments reported here) multidimensionally scaled the voices using direct similarity ratings to all pairs of voices, yielding the "psychological space" shown in Figure 4.

Table 1

Summary Statistics for the Stimulus Items

	Means				
Word Type	n	KF[†]	Subtitle	Letters	Syllables
High Frequency	68	250.8	355.8	4.26	1.16
Low Frequency	68	19.1	13.2	4.26	1.15

[†] Kuçera & Francis (1967)

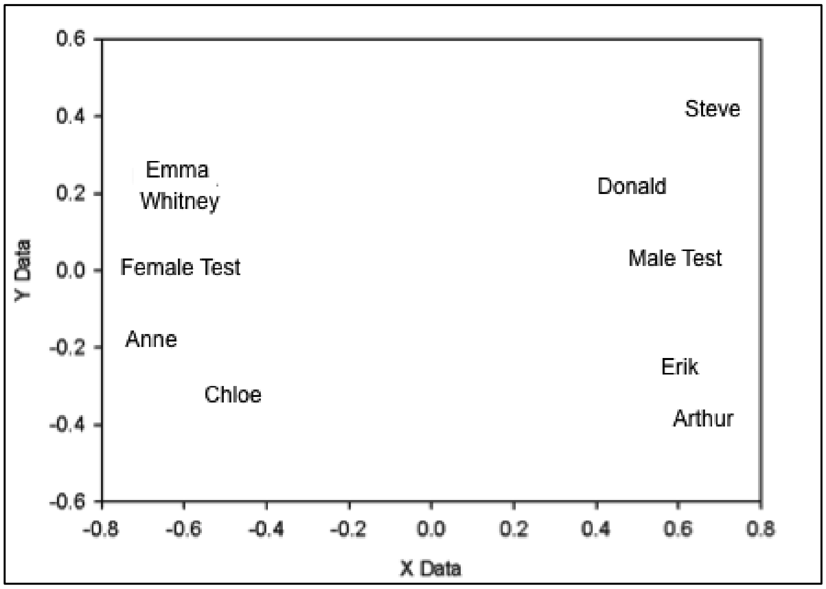


Figure 4. Multidimensional scaling solution of stimulus voices (*note*: names are fictional, and do not reflect the actual name of the speaker).

From this distribution, voices were grouped into 4 similarity conditions (HS1, HS2, LS1, LS2). High-similarity conditions were comprised of all same-sex speakers; HS1 used all male speakers and HS2 used all female speakers. Low-similarity conditions were composed of 2 males and 2 females, selected such that the same-gender speakers were not "close" in psychological space. The voices in LS1 were Emma, Anne, Steve, and Erik and those in LS2 were Whitney, Chloe, Donald, and Arthur. The male and female test voices were selected to represent the "average" male and female voices, so as to not be too confusable with any particular study speaker. The faces used to represent the speakers were selected from Ekman and Matsumoto (1993). All of the people in the database were White, and photographed under constant lighting conditions with the same visible clothing.

Study Procedures

Participants were tested individually in a dimly-lit, sound-attenuated booth. A chin rest maintained head position and viewing distance at 60 cm. Eye movements and pupil dilation were continuously recorded from both eyes at 50 Hz by a Tobii 1750 eye tracker and behavioral responses were recorded by an SRBox. E-Prime software managed both the presentation of the experiment and data collection (Psychology Software Tools, 2006). The luminance of the screen was controlled by the background color (RGB 150, 150, 150, as in Kuchinke et al., 2007), which was used on every screen.

Participants were first familiarized with the experiment and the eye tracker. After the chin rest was adjusted, such that eye position was maintained centrally on the horizontal axis and a slightly above-centrally on the vertical axis, participants were calibrated on the eye tracker. The calibration routine randomly presents nine fixation points (indicated by the movement of a blue dot) over the range of the display and participants “follow the dot” as it moves to each location. If the software or the researcher identified any missed fixations, the calibration routine was repeated. All participants were successfully calibrated within two attempts.

To familiarize participants with the voices and faces used in this experiment, I used a modified version of the familiarization task used by Dodson, Holland, and Shimamura (1998). For each face-name pair, participants were shown a full-screen image of a face and its name while listening to a question spoken by that speaker; all four face-name pairs were shown sequentially, speaking the same question[2]. After the fourth face, participants selected the answer to the question from on-screen multiple-choice options. After each block of 12 questions (four blocks total), participants completed a familiarization test, in which they were presented with 20 single words, five spoken by each of the four speakers. After hearing each word, participants were shown the four face-name pairs at the bottom of the screen, in random order on every trial. They selected the face-name pair that they believed matched the voice. All participants completed the voice-learning phase with at least 70% accuracy (chance was 25% in HS and 50% in LS).

---

[2] This departs from the method used by Dodson et al. (1998), and was meant to encourage participants to notice the differences among the voices. Pilot testing verified that the current method produced better voice learning than the original method.

The main experiment began with general instructions on the computer screen, indicating to participants that their gaze must be focused on the computer screen throughout the entire experiment in order to keep the program moving along at an acceptable pace. Participants then received a demonstration of a study trial (described below), and were encouraged to ask any clarification questions before beginning the experiment.

The study phase consisted of 80 trials, all proceeding as outlined in Figure 5, with 20 trials per speaker. On each trial, participants first focused on a central fixation cross for 1500 ms. Following the offset of the fixation cross, two vertical and two horizontal lines divided the screen into a 9-box grid comprised of equally-sized cells (see Figure 5). After an inter-stimulus interval (ISI) of 250 ms, a word in 18-pt bold Arial font was presented in one of seven cells on the screen (items were never presented centrally or in the lower-middle box[3]). After 1000 ms, the word disappeared and a star was presented in another cell for 750 ms. After the star disappeared, a face appeared in a third location on the screen and the spoken version of the word was played over the participants' headphones in the voice paired to that face. Following the offset of the spoken word, participants viewed a blank screen for 2000 ms prior to the start of the next trial.

The placement of items throughout the study trials was determined pseudo-randomly, such that the locations were chosen from one of 5 possible "patterns," constrained such that the face, the printed word, and the star were never presented

---

[3] This constraint was added because participants tended to look at the central boxes in the middle and bottom rows during the test trials (the middle center box is the location of the fixation cross and the bottom center box is where participants look in anticipation of the source monitoring judgment).

in the same locations or in the same row (e.g., the word and the face could not both appear in the middle row during a single trial). Each pattern was used with each face-name pair equally often.

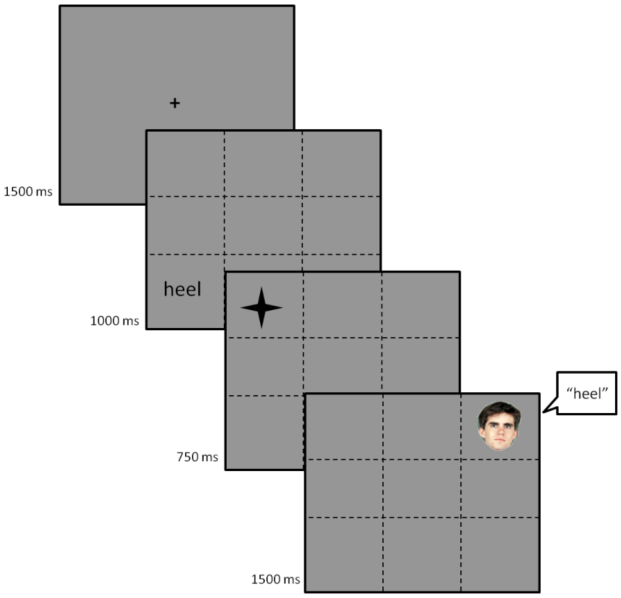


Figure 5. Schematic study trial.

Experiment 1

Experiment 1 assessed the extent to which retrieval cues influence the degree of retrieval success, and whether eye movements are reinstated across learning and retrieval. During encoding, participants studied words spoken by four distinct speakers as they completed a series of predetermined eye movement patterns. Fixations and pupil diameters were monitored continuously throughout the study trials, ensuring that the patterns were being followed. In a subsequent item recognition and source monitoring task, participants were tested with old and new items spoken by two new speakers, each bearing different levels of similarity to the original voices. During retrieval, test items were presented auditorily while only the 9-box grid was visible. That was meant to encourage eye movements. Further, by recording continuous measures of pupil diameter, I constructed bias-free pupillary ROCs to examine the nature of recognition memory, broadly-defined.

Participants

Eighty-six students were recruited from the undergraduate population at Arizona State University to participate in exchange for partial course credit or cash payment ($15). All participants had normal, or corrected-to-normal, vision and were native English speakers. Twelve participants were excluded prior to data analysis (three did not complete the voice familiarization phase, five were missing more than 7% of fixations[4], and four had eye-tracking failures[5]), leaving 74 participants in the remaining analyses. Thirty-eight participants were randomly assigned to HS and 36 were randomly assigned to LS.

Test Procedures

Following a 3-minute break, participants were familiarized with the test instructions, which were similar to those given in standard source monitoring experiments. During each test trial, participants heard a word spoken over their headphones by one of the two test-speakers (assigned to test words randomly within the constraint that each test voice be used equally often with words spoken by the original speakers). After a 1000-ms ISI, participants heard a low-pitched

---

[4] Seven percent was chosen as a cutoff because higher values typically yield extensive segments of missing data. With large segments of missing data, entire trials need to be dropped from analysis.

[5] Eye-tracking failures were equipment malfunctions, which resulted in large portions of missing data.

tone, which indicated that they should give their old/new recognition response[6] (*item recognition*). Regardless of the participants' responses, they were always asked to indicate the name of the person who spoke the word during study (*source identification*) by pressing one of four response keys corresponding to the location of the speaker (locations were changed on every trial). Participants were instructed to guess the source if they could not remember, or if they judged the item to be new. This was intended to reveal source information in the absence of item memory (as in Kurilla & Westerman, 2010), as well as any response biases. No feedback was given.

Results and Discussion

Across all three experiments reported here, pupil results were conducted on each participant's "better eye" (defined as the eye with fewer missing observations). Missing observations were filled in by linear interpolation (averaging the pupil diameter for 50-ms before and after the missing point). Response time (RT) measures were trimmed prior to analysis, by filtering outliers, defined as RTs falling 3 or more standard deviations above or below the mean. Alpha for all significance tests was set at .05, and multiple comparisons were Bonferroni-corrected.

Item Recognition Analyses

Because item hit rates did not differ across the subordinate conditions ($p >$ .05), they were collapsed into superordinate groups, HS and LS, for signal detection analyses. Participants' sensitivity (d') and bias (c) were analyzed in

---

[6] Note that the brief wait period was meant to encourage eye movements during memory retrieval.

separate 2-way ANOVAs. Although participants in the LS groups tended to have higher *d'* (*M* = .68, *SE* = .11), relative to the participants in the HS groups (*M* = .46, *SE* = .08), this difference was not reliable (*p* = .12). The groups also did not reliably differ in their response biases, *p* = .87. Neither the bias estimate for the LS group (*M* = .11, *SE* = .08), nor the HS group (*M* = .09, *SE* = .07), differed reliably from zero, both *p*s > .05.

To examine the influence of test voice on item recognition, hit rates were analyzed in a 4 (Condition: HS1/HS2/LS1/LS2) x 2 (Test Gender: Different/Same) mixed model, repeated measures (RM) ANOVA[7]. Although there was a main effect of Test Gender, *F*(1, 70) = 15.92, *p* < .05, η^2_p = .19, this effect was qualified by an interaction with Condition, *F*(3, 70) = 4.45, *p* = .01, η^2_p = .16. Pairwise comparisons revealed that the locus of the effect was in the HS groups. When the gender of the test voice matched the gender of the studied voices, participants' hit rates were reliably higher, relative to when they were changed. In HS1, participants studied words by male speakers; when the test speaker was also a male, hit rates were higher (*M* = .69, *SE* = .05), relative to when the test speaker was a female (*M* = .44, *SE* = .05). The reverse pattern was observed for HS2 (M_{same} = .62, SE_{same} = .05, $M_{different}$ = .44, $SE_{different}$ = .05), suggesting that participants were biased to say "old" when the test gender matched the study gender.

---

[7] In all analyses examining the influence of voice, HS and LS groups were analyzed by their separate subordinate conditions, because study voice gender differed across HS1 and HS2.

Complementary analyses were performed on correct rejection rates, to determine whether test voice gender also biased "new" responses. Once again, there was an interaction between Test Gender and Condition, $F(3, 70) = 5.79$, $p = .001$, $\eta^2_p = .20$. In the HS groups, participants were more accurate when rejecting items spoken in the opposite gender of the studied words. In HS1, participants correctly rejected words spoken by a female ($M = .67$, $SE = .05$) more frequently than words spoken by a male ($M = .48$, $SE = .06$). The same pattern held for participants in HS2; "new" responses were more accurate when the test speaker was a male ($M = .73$, $SE = .06$), relative to a female ($M = .53$, $SE = .05$). In short, hit and correct rejection rates indicated that test voices had a clearly biasing effect on participants' responses. When the test voice matched the superficial characteristics of the studied voice, participants were biased to respond "old;" the reverse held when the test voice did not match the characteristics of the studied voice.

Source-Monitoring Analyses

To determine whether the test voice influenced source recognition performance, I analyzed source hit rates in a 4 (Condition) x 2 (Test Gender) mixed model, RM ANOVA. Once again, the main effect of Test Gender, $F(1, 70) = 62.85$, $p < .05$, $\eta^2_p = .47$, was qualified by an interaction with Condition, $F(3, 70) = 19.11$, $p < .05$, $\eta^2_p = .45$. As can be seen in Figure 6, participants' source decisions were only influenced by the gender of the test speaker in the LS conditions. Regardless of the gender of the study speaker, when the gender of the test speaker matched, participants' source decisions were more accurate, relative to when the gender of the test speaker did not match. For example, in LS1, participants recalled the correct source more often (*M* = .48, *SE* = .02) when the study and test genders were the same, relative to when the study and test genders differed (*M* = .19, *SE* = .04). This is another bias effect, as was evident in the source hit rates. When participants were able to use gender to narrow down the source response options, they were more likely to select the correct speaker.

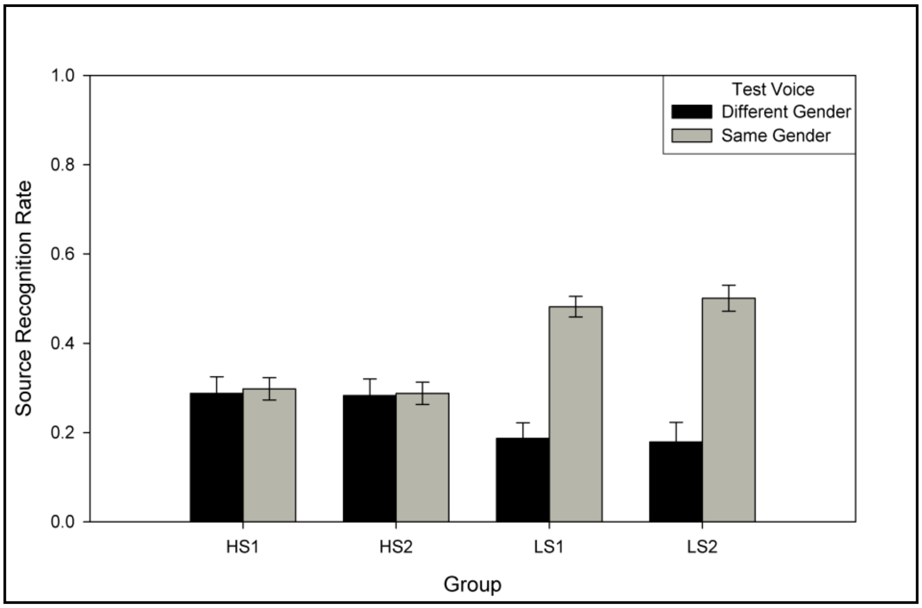


Figure 6. Source recognition rates by Group, as a function of test speaker. Error bars represent standard error of the mean.

To examine source-monitoring performance more precisely, participants' behavioral responses were sorted into a 5 x 5 table, with rows corresponding to the true source and the columns corresponding to the chosen source. Appendix B contains the 5 x 5 frequency tables summarized across participants in each of the four conditions, and Appendix C contains complementary tables displayed in percentages. For each participant, I calculated a specific-source identification (SSI) score and a partial-source identification (PSI) score (see Appendix D for the formulae, which were derived from those provided by Dodson et al., 1998). The SSI represents the probability of recalling the specific source (e.g., recognizing that Chloe was the original speaker). The PSI score represents the probability of recognizing the correct gender of the original speaker, in the absence of specific source recognition (e.g., selecting Anne, rather than Erik or Steve, when the original speaker was Chloe). Both scores are centered at zero, representing chance level decisions. Scores above zero reflect more accurate responses (e.g., a PSI score above zero reflects greater likelihood of selecting a same-gender, yet technically incorrect, source).

To examine specific source monitoring performance, conditions were collapsed into groups, HS and LS, because source monitoring performance did not differ across like-conditions (e.g., LS1 versus LS2). SSI scores were analyzed in a 2 (Group: HS/LS) x 2 (Item Recognition: Hit/Miss) mixed model, RM ANOVA. Data from three participants (two in HS and one in LS) were excluded for response frequencies of zero. There was a marginal effect of group ($p = .059$), reflecting slightly more accurate performance by the LS group ($M = .32$, $SE = .01$), relative to the HS group ($M = .28$, $SE = .01$). Item recognition performance had a reliable effect on SSI scores, $F(1, 69) = 16.98$, $p < .05$, $\eta^2_p = .20$. When participants correctly recognized the item, they were also more likely to retrieve specific source information ($M = .33$, $SE = .02$), relative to when they did not recognize an old item ($M = .26$, $SE = .01$). Although SSI was larger when participants were able to recognize items, it should be noted that one-sample *t*-tests, comparing the item correct and item incorrect SSI scores to zero, indicated that both values were reliably above zero, both *p*s < .05. Regardless of item memory, participants were able to retrieve specific source details.

Because participants in the HS groups could not, by design, have partial source information based on speaker gender, PSI scores were only analyzed in the LS groups (one participant was excluded for missing data) in a 2-way ANOVA comparing scores by whether or not the item was recognized. Item Recognition affected PSI scores, $F(1, 35) = 20.18$, $p < .05$, $\eta^2_p = .37$, but in the opposite direction of the SSI scores. Failing to recognize an old item yielded above-chance partial source recognition ($M = .52$, $SE = .06$). When participants correctly recognized items, they were less likely to recall partial source information ($M = .16$, $SE = .04$)[8]. Note that, although the PSI was larger during item incorrect trials, relative to item correct trials, one-sample *t*-tests comparing the values to zero indicate that both conditions are associated with PSI scores reliably above zero, both *p*s < .05. This finding is consistent with previously-reported findings of partial source memory (Dodson et al., 1998) and source memory in the absence of item memory (Kurilla & Westerman, 2010). Because participants' eye movements were monitored during encoding, it is unlikely that they were not attending to the events as presented, suggesting that dual-process explanations of partial source memory reflecting incomplete encoding are not applicable. Rather,

---

[8] This is a reflection of the prior analysis, which indicated that participants were more likely to recall source-specific information following correct item recognition.

recollection seems to follow a continuous retrieval process, capable of eliciting partial and specific source memories.

Response Time Analyses

To determine whether the specificity of participants' memories influenced their speed of responding, RTs during correct item and source recognitions were analyzed in separate ANOVAs. Item recognition RTs were analyzed in a 2 (Group) x 2 (Source Recognition) mixed model, RM ANOVA. No reliable main effects or interactions emerged, all *p*s > .05. Source recognition RTs were analyzed in a 2 (Group) x 2 (Item Recognition) mixed model, RM ANOVA. Participants in the LS group responded more quickly (*M* = 1411 ms, *SE* = 98) than participants in the HS group (*M* = 1832 ms, *SE* = 95), *F*(1, 72) = 9.6, *p* = .003, η^2_p = .12. This main effect was qualified by an interaction with Item Recognition, *F*(1, 72) = 5.09, *p* = .03, η^2_p = .07. Pairwise comparisons revealed that participants in the LS group responded to source questions more quickly following *incorrect* item recognitions (*M* = 1335 ms, *SE* = 105), relative to correct item recognitions (*M* = 1486 ms, *SE* = 102). Although it is possible that participants responded more quickly following incorrect item responses because they assumed that they were guessing, the effect size is relatively small and does not replicate across the remaining experiments. As such, this effect will not receive further attention.

Source RTs were also analyzed in the LS group by the specificity of the source memories in a 3-way (Source Memory: Full/Partial/None) ANOVA. No reliable effects emerged, *p* = .65. Overall, the RT analyses from Experiment 1 indicate that participants responded more quickly to the source discrimination when that discrimination is objectively easier, in the LS groups.

Pupillary ROC Analyses

For each participant, the mean and range of pupil diameters were calculated during test trials, allowing the range to be segmented by the standard deviation of the distribution. Because pilot data indicated that higher confidence estimates were associated with enlarged pupils, the relationship between pupil size and "confidence" was positive for old test trials and negative for new test trials. To create confidence "bins," the highest confidence estimate ("6") was associated with pupil diameters 1.5 or more standard deviations above the mean. A "5" was associated with pupil diameters at least 1 standard deviation above the mean, and so forth, in one-half standard deviation increments. As mentioned, this pattern was reversed for "new" trials. Response frequencies, cumulative response proportions, and z-scores for the ROC data can be found in Appendix E. Note that no subjects from LS2 were included in any ROC analysis, due to a programming error, which resulted in loss of eye-tracking data collection during "new" trials.

ROCs and *z*-ROCs are presented in Figure 7, with separate lines for the HS and LS groups. The points within each graph were fit with quadratic equations (the summary statistics are presented in Table 2), but subject-level statistics were used to analyze the characteristics of the curves. None of the ROC statistics (e.g., quadratic constants, slope, intercept) differed between the HS and LS groups, as indicated by one-way ANOVAs, all *p*s > .05. The quadratic constants of the lines fit to the ROC data were reliably different from zero, $t(58) = -14.49$, $p < .05$. The variance accounted for by a quadratic fit (97%) was reliably greater than the variance accounted for by a linear fit (94%), suggesting that the ROCs demonstrated significant curvature, *t*(58) = -15.31, *p* < .05, Cohen's *d* = 1.20.

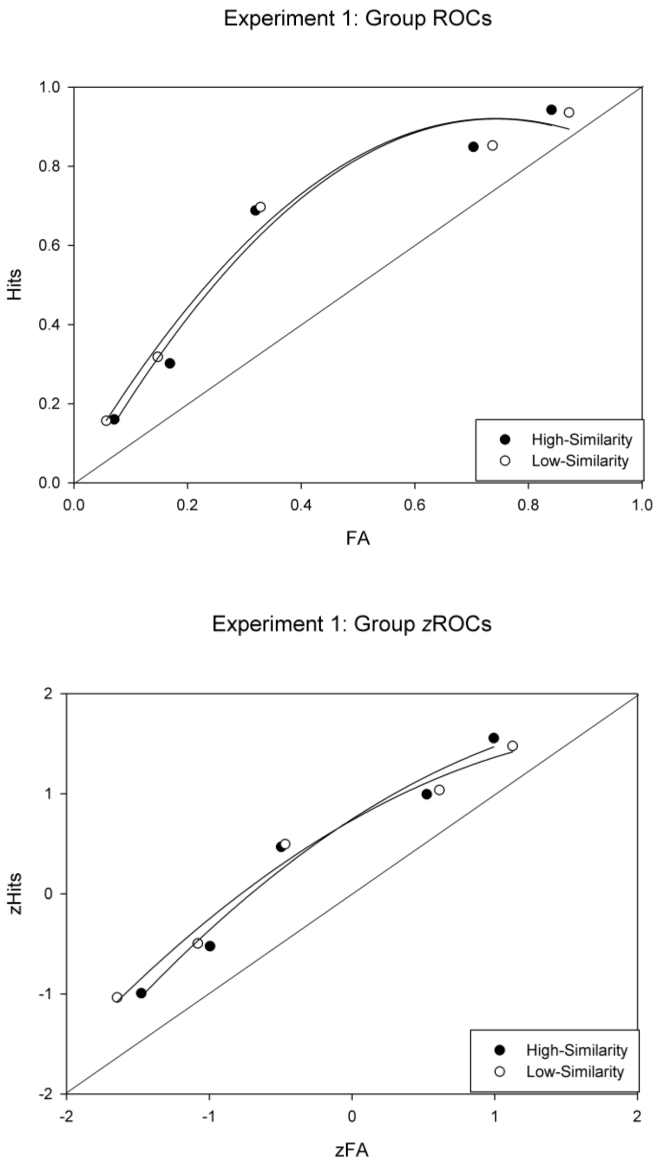


Figure 7. Pupillary group ROCs (top panel) and *z*-ROCs (bottom panel) from Experiment 1. Both graphs are shown with the best-fitting quadratic lines.

Table 2

Statistics for pupillary ROCs and z-ROCs in Experiment 1.

	ROCs		***z*-ROCs**	
Group	**Quadratic Constant**	**Slope**	**Quadratic Constant**	**Slope**
LS Group	-1.03 (.07)	1.90 (.08)	-0.03 (.01)	.83 (.02)
HS Group	-1.01 (.10)	1.94 (.11)	-0.06 (.01)	.86 (.01)

In the *z*-ROCs, quadratic and linear lines were again fit to the points. None of the points on the HS and LS lines differed in one-way ANOVAs. Although the quadratic constants for the *z*-ROCs were different from zero, $t(58) = -6.55$, $p < .05$, and therefore violated Hilford et al.'s (2002) second regularity of item recognition ROCs (i.e., that *z*-ROCs should be rectilinear), it is worth noting that these differences were relatively small. In fact, the variance accounted for by a linear solution (98%) increased marginally (but statistically reliably) when a quadratic fit was adopted (98.1%), $t(58) = -2.26$, $p = .03$, Cohen's $d = 0.28$. This pattern, and the small effect size, suggest that the functions are largely linear, consistent with the predictions of UVSD. Further support for UVSD comes from the slopes of the *z*-ROCs; the slopes for both HF and LF words were reliably below 1.0, t(58) = -19.26, $p < .05$, suggesting that the target distribution had a greater standard deviation than the lure distribution.

Eye Movement Analyses

Research has revealed that eye movements precede conscious recollection, and are preferentially guided to old locations or items within 500-ms of stimulus onset (Moscovitch, 2008; Hannula & Ranganath, 2009). As such, eye movements were analyzed during the test phase only, during predefined "interest periods" consisting of the time between the initiation of the test word and the completion of the old/new item response. Fixations were defined as moments when the eyes remained in a 100-pixel area for at least 100-ms (as in Goldinger, He, & Papesh, 2009). Once a fixation was established, it was labeled according to what was on-screen *during the corresponding study trial.* For example, if the study pattern involved cells 1, 6, and 9, those cells were labeled "regions of interest" (ROIs) during the test trial for that word (cells 2, 3, 4, and 7 would therefore be "non-regions of interest").

Eye movements were first analyzed by examining raw fixation rates during "old" test trials that were directed to ROIs or non-ROIs in a 2 (Fixation location: ROI/non-ROI) x 2 (Group) mixed model ANOVA. There was a marginal main effect of Group, $F(1, 72) = 3.73$, $p = .057$, $\eta^2_p = .05$, revealing that participants in the HS group made more fixations ($M = 114$, $SE = 10$) than participants in the LS group ($M = 87$, $SE = 10$). There was also a main effect of Fixation Location, $F(1, 72) = 6.21$, $p = .01$, $\eta^2_p = .08$. Although ROIs comprised only three cells, and non-ROIs comprised four, 107 ($SE = 8$) total fixations were directed to ROIs, whereas 94 ($SE = 7$) were directed to non-ROIs.

This finding was complemented by an analysis on the proportions of fixations (per trial) directed to ROIs versus non-ROIs by Group. On average, 55% of fixations during "old" trials were directed to ROIs, whereas 45% were directed to non-ROIs, $F(1, 72) = 16.21$, $p < .05$, $\eta^2_p = .18$. By chance, participants would be expected to fixate in ROIs 43% of the time. In reality, they fixated within those areas 55% of the time, which was reliably greater than chance, *t*(73) = 9.69, *p* < .01. This finding is consistent with previously-documented accounts of spontaneously reinstated eye movements during memory retrieval (e.g., Laeng & Todorescu, 2007; Spivey & Geng, 2000).

Fixation durations were analyzed in a 2 (Group) x 2 (Fixation Location: ROI/non-ROI) mixed model, RM ANOVA. There was no reliable effect of Group on fixation duration, *p* = .99. There were, however, reliable effects of Fixation Location on the duration of participants' fixations. When fixating within an ROI, participants' fixations were longer (*M* = 498 ms, *SE* = 18), relative to when fixating within a non-ROI (*M* = 453, *SE* = 16), *F*(1, 72) = 5.42, *p* = .02, η^2_p = .07. This suggests that participants not only completed more fixations to ROIs, but that they spent significantly more time within those areas prior to making a response. Research on eye movement rates (to nothing) during memory retrieval find similar effects (see Hannula et al., 2011). Fixation durations within ROIs and non-ROIs were also compared to average fixations during "new" trials in separate paired-samples *t*-tests, collapsing across group[9]. No reliable differences were observed, both *p*s > .05.

To determine whether reinstated fixations were associated with more accurate item memory, the proportion of fixations to ROIs was analyzed in a paired-samples *t*-test, comparing hits to misses. No reliable difference emerged, *t*(72) = -.59, *p* = .56. A separate paired samples *t*-test was conducted on refixations by source accuracy (correct versus incorrect). When participants

---

[9] New trials were not included in the initial analysis to avoid missing data. As mentioned earlier, LS2 participants' eye-tracking data during new trials was not recorded.

correctly identified the original speaker, they were more likely to engage in refixations (*M* = .43, *SE* = .02), relative to when they were unable to recall the original speaker (*M* = .39, *SE* = .01), *t*(72) = -2.61. *p* = .01. A one-way ANOVA revealed that this effect was not sensitive to the detail of source memories (e.g., full versus partial) in the LS group, *p* = .25. Although some theories (e.g., O'Regan, 1992) suggest that NVGPs are epiphenomenal, and reflect the fact that there are no internal memory representations, the finding that accurate source memory was reliably associated with increased fixation rates to ROIs provides initial evidence to the contrary. When participants retrieve detailed memories, they tend to fixate within previously-viewed locations.

It is also possible that reinstated eye movements reflect the speed with which memory processes resolve. As such, participants' proportion of fixations to ROIs on each trial were quartiled, and labeled as high (>.75), mid-high (>.50), mid-low (>.25), and low (<=.25). RTs during item recognition responses and source judgments were analyzed in separate 2 (Group) x 4 (Fixation Proportion) mixed model, RM ANOVAs. Because both item and source responses were prompted, no differences were expected to emerge. For item RTs, no differences were observed by Group, *p* = .22, but there was a reliable main effect of Fixation Proportion, *F*(1, 68) = 17.22, *p* < .05, η^2_p = .20. As shown in the left panel of Figure 8, participants responded more quickly during high and low fixation proportion trials. The same effect emerged in analyses on source RTs, *F*(1, 68) = 49.63, *p* < .05, η^2_p = .42. As shown in the right panel of Figure 8, participants' source judgments were faster following both high and low fixation proportions.

The RT results reveal an interesting pattern. Specifically, RTs are faster during item and source judgments when the rate of reinstated eye movements is very low or very high. RTs during the intermediate levels are relatively slow. This pattern suggests, potentially, two memory processes at work, one fast, automatic process, and one slow, deliberate process. Although it is tempting to equate these processes to familiarity and recollection, respectively, the rates of fixations within ROIs suggest that accurate source memory is associated with higher rates of reinstated eye movements. If source memory is primarily characterized by recollection, and those memories are also associated with higher rates of reinstated eye movements, then those memories are also retrieved more quickly, which is inconsistent with dual-process conceptualizations of recollection.

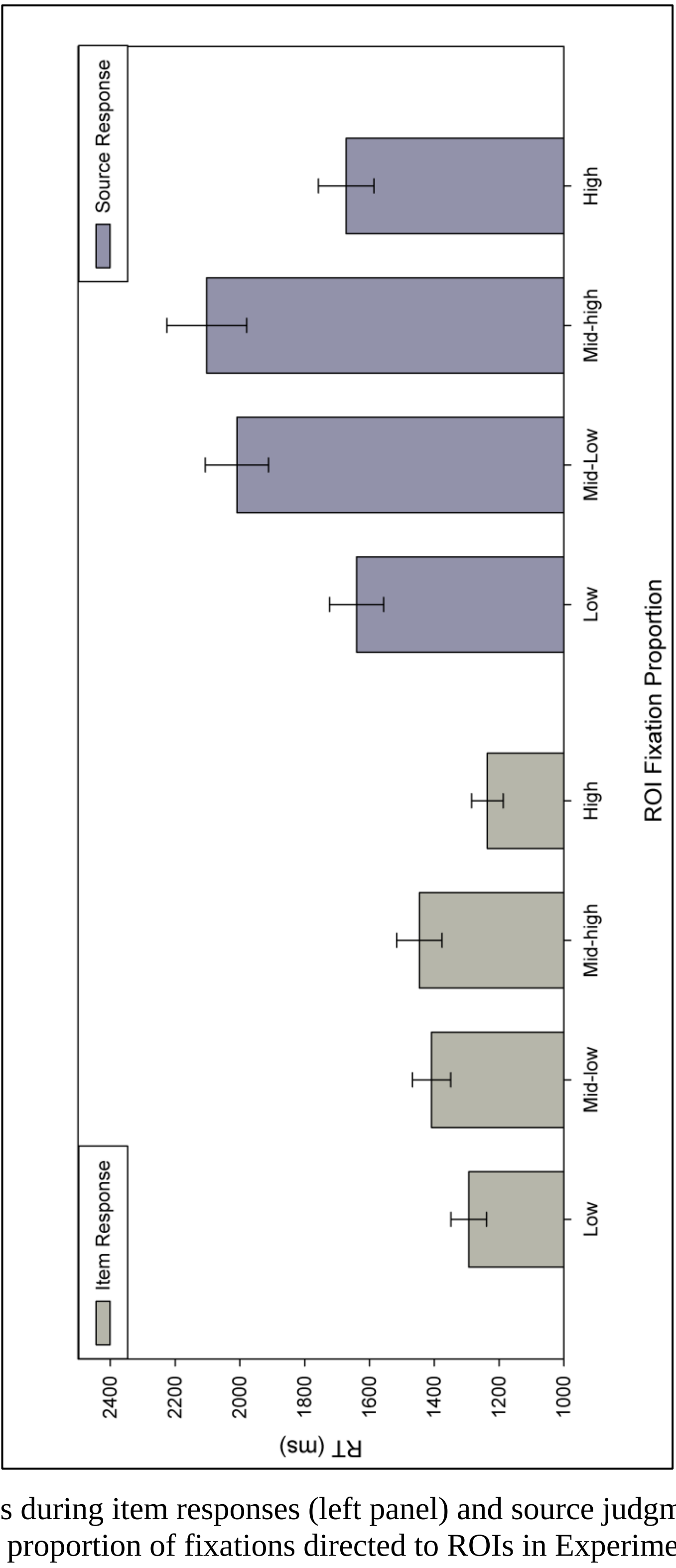


Figure 8. RTs during item responses (left panel) and source judgments (right panel) by the proportion of fixations directed to ROIs in Experiment 1. Error bars represent standard error.

Experiment 2

Given the finding in eye movements from Experiment 1 (namely that participants spontaneously reinstated encoding fixations, which were associated with more accurate source memories), Experiment 2 aimed to assess whether these refixations could be externally-cued to influence memory processes. By requiring participants to fixate in regions of interest (or non-regions of interest) prior to hearing a retrieval cue, I tested whether item and source memory are aided or disrupted by non-spontaneous eye movements.

Participants

Fifty-six students were recruited from the undergraduate population at Arizona State University to participate in exchange for partial course credit or cash payment ($15). All participants had normal, or corrected-to-normal, vision and were native English speakers. Eight participants were excluded prior to data analysis (three did not finish within a 90-minute time-frame, two did not complete the voice familiarization phase, one never responded “new” during the test trials, one never responded “old,” and one had greater than 7% missing fixations), leaving 48 participants in the remaining analyses. Twenty-three participants were randomly assigned to the HS conditions and 25 were assigned to the LS conditions.

Test Procedures

Following the 3-minute break, participants received verbal test instructions. Unlike Experiment 1, participants completed a single fixation for 1000-ms prior to hearing the test word. Fifty percent of their fixations preceding “old” trials were directed to one of two ROIs, the original location of the word (*item location*) or the original location of the face (*source location*). The remaining 50% of fixations were directed to one of the unused locations; all fixations preceding “new” trials were random. All other response characteristics were the same as Experiment 1.

Results and Discussion

Item Recognition Analyses

Participants’ *d'* and *c* were analyzed in separate 2-way ANOVAs. Although participants in the LS groups had higher *d'* (*M* = .59, *SE* = .11), relative to the participants in the HS groups (*M* = .38, *SE* = .12), this difference was not reliable, *p* = .19. The groups also did not reliably differ in their response biases, *p* = .87. Response biases in the LS group (*M* = .10, *SE* = .10) did not differ from zero, *p* = .25, but those in the HS group (*M* = .16, *SE* = .06) were reliably above zero, *p* = .02, indicating that HS participants had a conservative response bias.

To examine the influence of test voice on item recognition, hit rates were analyzed in a 4 (Condition: HS1/HS2/LS1/LS2) x 2 (Test Gender: Different/Same) mixed model, RM ANOVA. The interaction between Condition and Test Gender was marginally significant, *F*(3, 44) = 2.65, *p* = .06, $\eta^2_p = .15$. Pairwise comparisons revealed that participants in HS1 were reliably affected by the test voice gender, such that male voices yielded more hits (*M* = .65, *SE* = .06), relative to female voices (*M* = .45, *SE* = .06). Because participants in HS1 studied words spoken by all male speakers, hearing a male voice at test biased them to call the item “old.”

The complementary finding in correct rejection rates was observed in a 4 (Condition) x 2 (Test Gender) mixed model, RM ANOVA. Test Gender and Condition interacted, $F(3, 44) = 3.52$, $p = .02$, $\eta^2_p = .19$, such that participants in the HS conditions more accurately rejected new items when the test speaker was opposite in gender from the study speakers. In HS1, participants more correctly rejected words spoken by a female ($M = .68$, $SE = .07$), relative to a male ($M = .47$, $SE = .07$). The same pattern held for participants in HS2; “new” responses were more accurate when the test speaker was a male ($M = .70$, $SE = .08$), relative to a female ($M = .59$, $SE = .08$).

Source-Monitoring Analyses

To determine whether the test voice influenced source recognition performance, I analyzed source hit rates in a 4 (Condition) x 2 (Test Gender) mixed model, RM ANOVA. Once again, the main effect of Test Gender, $F(1, 44) = 44.59$, $p < .05$, $\eta^2_p = .50$, was qualified by an interaction with Condition, $F(3, 44) = 19.5$, $p < .05$, $\eta^2_p = .57$. As can be seen in Figure 9, participants’ source decisions were only influenced by the gender of the test speaker in the LS conditions. As in the LS conditions in Experiment 1, when the gender of the test speaker matched that of the study speaker, participants’ source decisions were more accurate, relative to when the two genders did not match. This effect occurred across both male and female study speakers, suggesting that participants’ memories contained some degree of episodic voice detail.

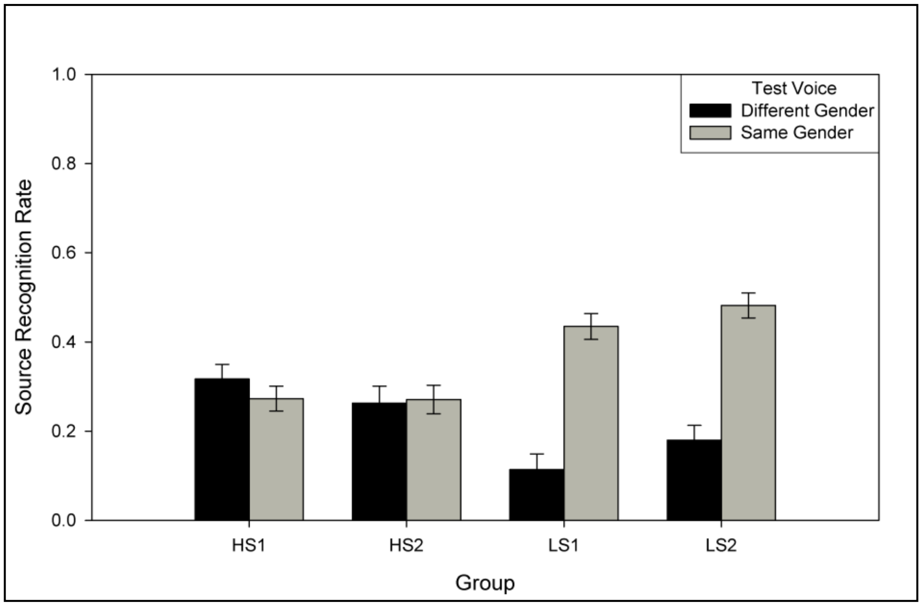


Figure 9. Source recognition rates by Group, as a function of test speaker. Error bars represent standard error of the mean.

To determine whether the test voice influenced source recognition performance, I analyzed source hit rates in a 4 (Condition) x 2 (Test Gender) mixed model, RM ANOVA. Once again, the main effect of Test Gender, $F(1, 44) = 44.59$, $p < .05$, $\eta^2_p = .50$, was qualified by an interaction with Condition, $F(3, 44) = 19.5$, $p < .05$, $\eta^2_p = .57$. As can be seen in Figure 9, participants' source decisions were only influenced by the gender of the test speaker in the LS conditions. As in the LS conditions in Experiment 1, when the gender of the test speaker matched that of the study speaker, participants' source decisions were more accurate, relative to when the two genders did not match. This effect occurred across both male and female study speakers, suggesting that participants' memories contained some degree of episodic voice detail.

As in Experiment 1, participants' behavioral responses were sorted into a 5 x 5 table, with rows corresponding to the true source and the columns corresponding to the chosen source (see Appendix F for frequency tables, and Appendix G for percentages). To examine specific source monitoring performance, participants' SSI scores were analyzed in a 2 (Group: HS/LS) x 2 (Item Recognition: Hit/Miss) mixed model, RM ANOVA. Data from one HS participant was excluded for at least one response frequency of zero. As in Experiment 1, item recognition performance had a reliable effect on SSI scores, $F(1, 45) = 11.83$, $p = .001$, $\eta^2_p = .21$. When participants correctly recognized items, they were also more likely to retrieve specific source information ($M = .33$, $SE = .02$), relative to when they did not recognize old items ($M = .24$, $SE = .01$).

LS participants' PSI scores were analyzed in a 2-way (Item Recognition) ANOVA. Participants' item recognition performance affected PSI scores, $F(1, 24) = 33.06$, $p < .05$, $\eta^2_p = .58$, replicating the finding from Experiment 1. Failure to recognize old items yielded partial source recollection ($M = .69$, $SE = .07$); participants recalled the gender of the speaker without recalling the specific name. The PSI score was reliably lower when participants correctly recognized old items ($M = .18$, $SE = .04$).

Response Time Analyses

As in Experiment 1, item recognition RTs were unaffected by Group or subsequent source recognition, as confirmed by a 2 (Group) x 2 (Source Recognition) mixed model, RM ANOVA, all *p*s > .05. RTs during correct source recognitions were also analyzed in a 2 (Group) x 2 (Item Recognition) mixed model, RM ANOVA. Participants in the LS group respond more quickly (*M* = 1404 ms, *SE* = 93) than participants in the HS group (*M* = 1788 ms, *SE* = 99), $F(1, 45) = 7.92$, $p = .007$, $\eta^2_p = .15$. No other main effects or interactions were reliable. Additional analyses on source RTs in the LS group revealed no effect of source memory detail (full/partial/none) in a 3-way ANOVA, $p = .96$.

The RT results replicated those from Experiment 1. Specifically, participants' source retrieval time was affected by the global difficulty of the task, such that those in the LS group responded to the source discrimination prompt more quickly than those in the HS group.

Pupillary ROC Analyses

Pupillary ROCs were constructed in the same manner as Experiment 1, and detailed statistics are presented in Appendix H. In this experiment, however, all participants in all conditions were included in the analyses.

ROCs and *z*-ROCs are presented in Figure 10, with separate trend lines for the HS and LS groups. The points within each graph were fit with quadratic equations (see Table 3 for summary statistics), but subject-level statistics were used to analyze the characteristics of the curves. None of the ROC statistics (e.g., quadratic constants, slope, intercept) differed between the HS and LS groups, as indicated by one-way ANOVAs, all *p*s > .05. The quadratic constants of the lines fit to the ROC data were reliably different from zero, $t(44) = -12.28$, $p < .05$. Further, the variance accounted for by a quadratic fit (98%) was reliably greater than the variance accounted for by a linear fit (95%), suggesting that the ROCs were curvilinear, $t(44) = -9.95$, $p < .05$, Cohen's $d = 1.28$. This finding is consistent with single-process, signal-detection views of episodic memory, which predict curved ROCs (Wixted, 2007).

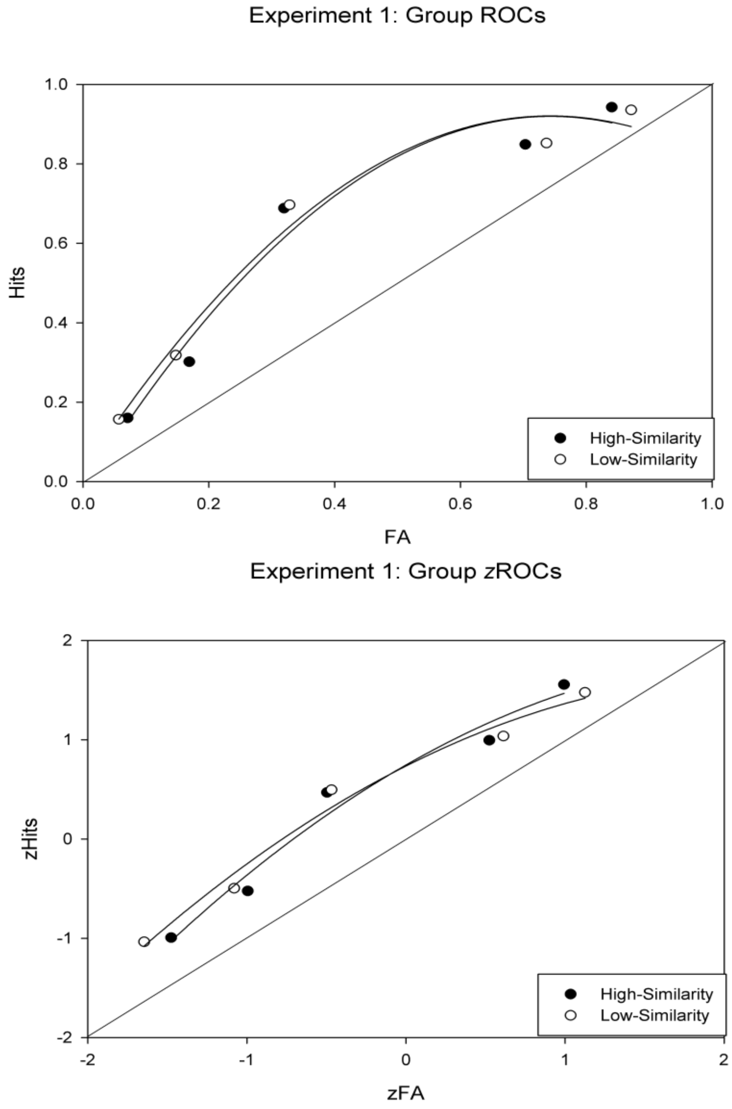


Figure 10. Pupillary group ROCs (left panel) and *z*-ROCs (right panel) from Experiment 2. Both graphs are shown with the best-fitting quadratic lines.

Table 3

Statistics for pupillary ROCs and z-ROCs in Experiment 2.

	ROCs		***z*-ROCs**	
Group	**Quadratic Constant**	**Slope**	**Quadratic Constant**	**Slope**
LS Group	-1.51 (.17)	2.38 (.11)	-0.14 (.07)	.95 (.04)
HS Group	-1.22 (.13)	2.06 (.17)	-0.05 (.03)	.97 (.05)

To examine the *z*-ROCs, quadratic and linear lines were fit to the points. Between the HS and LS groups, none of the equation parameters differed in one-way ANOVAs, *p*s > .05. Although the quadratic constants for the *z*-ROCs were different from zero, *t*(44) = -2.44, *p* = .02, violating the regularity that *z*-ROCs are rectilinear (Hilford et al., 2002), these differences were relatively small. The variance accounted for by a linear solution (96.2%) increased marginally (but statistically reliably) when a quadratic fit was adopted (97.6%), *t*(44) = -5.06, *p* < .05, Cohen's *d* = 0.64. This pattern suggests that the functions are predominantly linear, yet more curved than Experiment 1. Also inconsistent with UVSD, the slope of the line through the *z*-ROCs (.96) does not reliably differ from 1.0, *t*(44) = -1.21, *p* = .23, suggesting that the strength distribution of targets and lures is consistent with equal-variance models.

Eye Movement Analyses

To determine whether reinstated single fixations influence item recognition performance, participants' hit rates were analyzed in a 2 (Group) x 3 (Fixation Location: Item/New/Source) mixed model, RM ANOVA. No reliable effects emerged, all *p*s > .05.

In Experiment 1, refixations were associated with accurate source recognition performance; as such, source hit rates were analyzed in a 2 (Group) x 2 (Item Recognition) x 3 (Fixation Location) mixed model, RM ANOVA. Only the main effect of Item Recognition was reliable, $F(1, 45) = 16.25$, $p < .05$, $\eta^2_p = .27$. Consistent with earlier analyses, correct item recognition was associated with correct source recognition.

To examine whether reinstated fixations affected the specificity of source memories, partial source recognition performance (defined as choosing the correct source *gender* in the absence of correct specific source recall in LS groups) was analyzed in a 2 (Item Recognition) x 3 (Fixation Location) RM ANOVA. No reliable effects or interactions were observed.

In Experiment 1, participants' item and source RTs were facilitated by very frequent and very infrequent rates of fixations to ROIs. This was not the case in Experiment 2. Participants RTs during item and source judgments were analyzed in separate 2 (Group) x 3 (Fixation Location) mixed model, RM ANOVAs. No reliable effects emerged, all *p*s > .05

Experiment 3

Experiment 1 revealed that participants spontaneously refixate upon previously-viewed locations during a memory retrieval task with no visual demands. Experiment 2, however, revealed that completing externally-cued fixations to relevant or irrelevant screen locations did not influence item or source memory. Part of the discrepancy between these two findings could reflect the fact that participants completed multiple spontaneous refixations in Experiment 1, whereas they completed only a single fixation in Experiment 2. It is possible that memory traces for encoding experiences include the motoric component of the eye movement series, and that cueing a single piece of that component is insufficient to elicit a change in memorial experience. As such, Experiment 3 investigated the role of full motor processes, not just fixations, in the formation and retrieval of source memories. In this experiment, participants were required to complete a series of eye movements prior to the presentation of the test stimulus. By manipulating the oculomotor program during retrieval, I assessed the functional role of eye movement sequences during retrieval from memory.

Participants

Sixty-two students were recruited from the undergraduate population at Arizona State University to participate in exchange for partial course credit or cash payment ($15). All participants had normal, or corrected-to-normal, vision and were native English speakers. Five participants were excluded prior to data analysis (three were missing more than 7% fixations, one never responded “new” during the test trials, and one never responded “old”), leaving 57 participants in the remaining analyses. Twenty-five participants were randomly assigned to HS, and 32 were randomly assigned to LS.

Test Procedures

During the test phase of Experiment 3, participants reinstated entire fixation patterns. Prior to hearing each test word, participants completed three fixations, guided by a star-shaped figure, for 750-ms each. Fixation patterns during old trials could fall into one of four categories: *Full reinstatement* (participants completed the same series of fixations during retrieval as during encoding), *item reinstatement* (participants completed two random fixations and one directed to the original location of the printed word), *source reinstatement* (participants completed two random fixations and one directed to the original location of the face), or *new pattern* (participants completed a new series of fixations). Each pattern type was used equally often with old and new items (although, by design, all patterns are “new” during new test trials). All other response characteristics were the same as Experiment 1.

Results and Discussion

Item Recognition Analyses

Participants’ *d'* and *c* were analyzed in separate 2-way ANOVAs. Participants in the LS groups had higher *d'* (M = .59, SE = .11), relative to the participants in the HS groups (*M* = .22, *SE* = .06), *F*(1, 55) = 7.09, *p* = .01. The groups did not reliably differ in their response biases, *p* = .49. Bias in the LS group (*M* = .13, *SE* = .06) was marginally different from zero, *p* = .05, indicating that those participants were somewhat conservatively biased. Bias in the HS group (*M* = .07, *SE* = .08) did not differ from zero, *p* > .05.

To examine the influence of test voice on item recognition, hit rates were analyzed in a 4 (Condition: HS1/HS2/LS1/LS2) x 2 (Test Gender: Different/Same) mixed model, RM ANOVA. Unlike the previous experiments, the gender of the test voice did not influence responding, *F*(1, 53) = 2.97, *p* = .09, η^2_p = .05. No effects or interactions were reliable. Similarly, the gender of the test voice did not influence participants’ ability to correctly reject new items. No effects or interactions in a 4 (Condition) x 2 (Test Gender) RM ANOVA on correct rejection rates were reliable, all *p*s > .05.

Source-Monitoring Analyses

To determine whether the test voice influenced source recognition performance, I analyzed source hit rates in a 4 (Condition) x 2 (Test Gender) mixed model, RM ANOVA. The main effect of Test Gender, $F(1, 53) = 33.54$, $p < .05$, $\eta^2_p = .39$, was qualified by an interaction with Condition, $F(3, 53) = 11.82$, $p < .05$, $\eta^2_p = .40$. As can be seen in Figure 11, participants' source decisions were only influenced by the gender of the test speaker in the LS conditions. When the gender of the test voice matched the gender of the study voice, participants' source judgments were more accurate. Although participants had no reason to assume that the gender of the test speaker was diagnostic (in LS conditions, two study speakers were male and two were female), when that gender matched the study gender, participants were better able to correctly identify the original speaker. As in Experiments 1 and 2, this is interpreted as a bias effect: Participants used the test gender to narrow the field of response options from four to two. As can be seen in Figure 11, LS participants' performance when the test speaker was the same-gender as the studied speaker did not differ reliably from chance (50%, assuming that they were biased to exclude the opposite-gender response options).

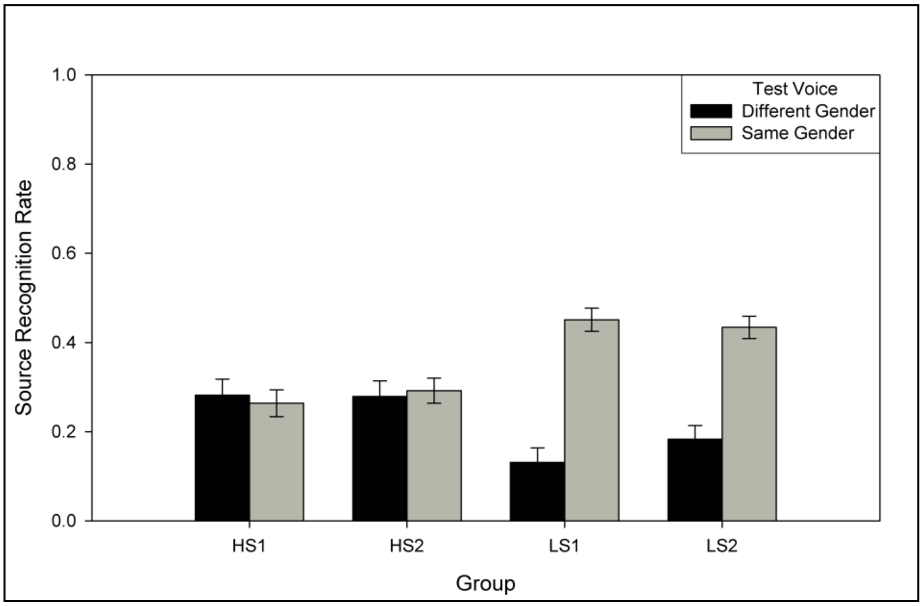


Figure 11. Source recognition rates by Group, as a function of test speaker. Error bars represent standard error of the mean.

Behavioral responses were sorted into a 5 x 5 frequency table, with rows corresponding to the true source and the columns corresponding to the chosen source (see Appendix I for the frequency tables, and Appendix J for the tables expressed as percentages). To examine specific source monitoring performance, participants' SSI scores were analyzed in a 2 (Group: HS/LS) x 2 (Item Recognition: Hit/Miss) mixed model, RM ANOVA. As in Experiments 1 and 2, item recognition performance had a reliable effect on SSI scores, $F(1, 55) = 13.47$, $p = .001$, $\eta^2_p = .20$. When participants correctly recognized items, they were also more likely to retrieve specific source information ($M = .32$, $SE = .01$), relative to when they did not recognize old items ($M = .26$, $SE = .01$). As confirmed by one-sample *t*-tests, both values were reliably above zero, both *ps* < .05.

LS participants' PSI scores were analyzed in a 2-way (Item Recognition) ANOVA, which revealed that Item Recognition reliably affected PSI scores, $F(1, 31) = 41.91$, $p < .05$, $\eta^2_p = .58$. When participants were unable to correctly recognize old items, they nevertheless demonstrated partial source recollection ($M = .69$, $SE = .07$). When they correctly recognized items, however, they were less likely to recall partial source information ($M = .21$, $SE = .04$). Both PSI estimates are reliably above zero, as confirmed by one-sample *t*-tests, both *p*s < .05. This is consistent with findings from Experiments 1 and 2. Overall, this establishes that partial source memory in the absence of accurate item memory is a reliable and replicable effect, and exists contrary to the predictions of many dual-process accounts of memory. Whereas dual-process theories assume that recollection reflects high-confidence, all-or-none recollection, these data indicate that recollection can occur even with relatively low item strength (otherwise, participants would have correctly recognized the items). This finding cannot be accommodated by existing dual-process accounts, but is easily predicted and explained by single-process, signal-detection theories (e.g., Starns et al., 2008).

Response Time Analyses

A 2 (Group) x 2 (Source Recognition) analysis on correct item RTs revealed no reliable interactions or main effects, all *p*s > .05, consistent with the previous two experiments. As in Experiments 1 and 2, a 2 (Group) x 2 (Item Recognition) mixed model, RM ANOVA on correct source RTs revealed that participants in the LS group responded more quickly (*M* = 1465 ms, *SE* = 77) than participants in the HS group (*M* = 1802 ms, *SE* = 87), *F*(1, 55) = 8.44, *p* = .005, η^2_p = .13. A main effect of Item Recognition, *F*(1, 55) = 5.22, *p* = .03, η^2_p = .09, revealed that participants responded to the source judgment more quickly when they recognized the item (*M* = 1517 ms, *SE* = 82), relative to when they did not recognize the item (*M* = 1750 ms, *SE* = 72). The interaction was not reliable. Additional analyses on source RTs in the LS group revealed no effect of source memory specificity (full/partial/none) in a 3-way ANOVA, *p* = .35.

The RT results from Experiment 3 replicate those from Experiments 1 and 2. Specifically, judgment times were clearly affected by the overall difficulty of the task. When discriminating between four highly-similar sources, participants' decision times were longer, relative to when discriminating between four less-similar sources. In short, task difficulty influenced the accumulation of evidence in favor of one source over the others, as operationalized via RTs.

Pupillary ROC Analyses

Pupillary ROCs were constructed in the same manner as Experiment 1, and detailed statistics are presented in Appendix K. All participants in all conditions were included in the analyses.

ROCs and *z*-ROCs are presented in Figure 12, with separate lines for the HS and LS groups. The points within each graph were fit with quadratic equations (see Table 4 for summary statistics), but subject-level statistics were used to analyze the characteristics of the curves. None of the ROC statistics (e.g., quadratic constants, slope, intercept) differed between the HS and LS groups, as indicated by one-way ANOVAs, all *p*s > .05. The quadratic constants of the lines fit to the ROC data were reliably different from zero, $t(56) = -16.6$, $p < .05$, suggesting curved lines. Further, the variance accounted for by a quadratic fit (97%) was reliably greater than the variance accounted for by a linear fit (94%), confirming that the ROCs were curved, $t(56) = -12.45$, $p < .05$, Cohen's $d = 1.39$. As in Experiments 1 and 2, this finding is consistent with single-process, signal-detection views of recognition memory, and inconsistent with the predictions of dual-process theory. What the curved ROCs suggest is that memory retrieval can be accurate across degrees of memory strength (or confidence, or response bias). Dual-process theories predict that item and source ROCs should be linear, reflecting the function of a threshold-based recollection process, which is typically associated with the highest level of confidence or strength. This was not observed across any of the experiments reported here.

To examine the *z*-ROCs, quadratic and linear lines were fit to the points. None of the equation parameters for the HS and LS lines differed in one-way ANOVAs, $p > .05$. The quadratic constants for the *z*-ROCs were different from zero, $t(56) = -3.38$, $p = .001$, suggesting that the lines were curvilinear. Moreover, the variance explained by a linear solution (96.2%) increased marginally (but statistically reliably) when a quadratic fit was adopted (97.7%), $t(56) = -5.03$, $p < .05$, Cohen's $d = 0.70$. This pattern suggests that the functions are somewhat curvilinear, in contrast with UVSD. To further examine the *z*-ROC predictions of UVSD, I analyzed the slope of the line through the functions ($M = .94$, $SE = .02$) in a one-sample *t*-test, comparing the values to 1.00. The analysis revealed that the slopes were statistically reliably below 1.0, which is consistent with unequal variance and targets and lures, *t*(56) = -2.67, *p* = .01. Although this is consistent with UVSD predictions, the slightly curved functions are somewhat inconsistent. The discrepancy between the ROC and *z*-ROC results will be discussed further in the General Discussion, but across the three experiments reported here, the results are predominantly consistent with strength-based theories of recognition memory.

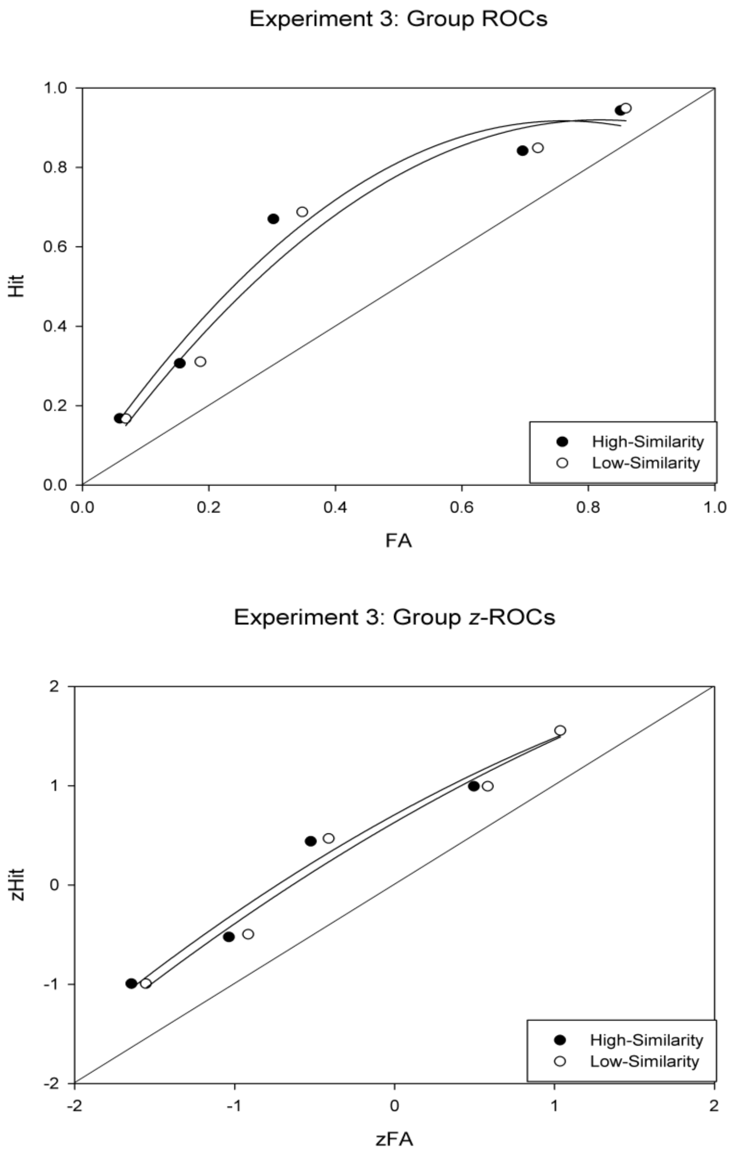


Figure 12. Pupillary group ROCs (top panel) and *z*-ROCs (bottom panel) from Experiment 3. Both graphs are shown with the best-fitting quadratic line.

Table 4

Statistics for pupillary ROCs and z-ROCs in Experiment 3.

	ROCs		***z*-ROCs**	
Group	**Quadratic Constant**	**Slope**	**Quadratic Constant**	**Slope**
LS Group	-1.31 (.09)	2.18 (.08)	-0.10 (.03)	.96 (.03)
HS Group	-1.48 (.15)	2.28 (.11)	-0.14 (.06)	.92 (.03)

Eye Movement Analyses

To determine whether reinstated fixation patterns influenced item recognition performance, participants' hit rates were analyzed in a 2 (Group) x 4 (Fixation Pattern: Full/Item/New/Source) mixed model, RM ANOVA. No reliable effects emerged, all *p*s > .05.

Source hit rates were also analyzed as a function of reinstated fixation patterns, in a 2 (Group) x 2 (Item Recognition) x 4 (Fixation Pattern) mixed model, RM ANOVA. Only the main effect of Item Recognition was reliable, $F(1, 53) = 10.44$, $p = .002$, $\eta^2_p = .17$. Consistent with earlier analyses, correct item recognition was associated with correct source recognition. No effect of Fixation Pattern was observed, $p = .40$.

To examine whether reinstated fixation patterns affected the specificity of source memories, partial source recognition performance (i.e., as choosing the correct source gender in the absence of correct specific source recall in LS groups) was analyzed in a 2 (Item Recognition) x 4 (Fixation Pattern) RM ANOVA. No reliable effects or interactions were observed.

In Experiment 1, when participants fixated in ROIs very frequently (or very infrequently), their item and source judgments were faster, relatively to mid-level fixation proportions. This finding did not occur in Experiment 2, when participants engaged in single fixations during retrieval, and it did not replicate here, in Experiment 3, when participants engaged in full fixation patterns at retrieval. Participants RTs during item and source judgments were analyzed in separate 2 (Group) x 3 (Fixation Location) mixed model, RM ANOVAs. No reliable effects emerged, all *p*s > .05.

General Discussion

The present experiments were designed to address four general research aims, all of which were generally focused on elucidating the nature of recognition memory. In each experiment, participants completed a series of eye movements while encoding words spoken by one of four speakers. The speakers were selected to have a predetermined similarity relationship, both to each other, and with subsequent test speakers. Through this manipulation, I examined the influence of similarity on item and source memory (Research Aim 1). Further, participants in each experiment provided source judgments (i.e., they named the original speaker) even when they failed to provide accurate old/new item recognition responses. This method allowed me to examine degrees of specificity in source memories, irrespective of item memory (Research Aim 2). Third, by monitoring pupil diameter during retrieval, I constructed pupillary ROCs and *z*-ROCs, allowing me to make inferences about the broader conceptualization of recognition memory (Research Aim 3). Lastly, across experiments, participants’ eye movements at retrieval were manipulated, such that they were consistent or inconsistent with the fixation patterns followed during encoding. This allowed me to determine whether reinstated eye movements are functional or necessary during memory retrieval (Research Aim 4).

Research Aim 1: The Influence of Similarity

Across all three experiments, overwhelming evidence suggested that voice similarity influenced memory responses. Although participants in the LS groups generally had higher *d'* estimates, relative to participants in the HS groups, this finding was only reliable in Experiment 3. In Experiments 1 and 2, *d'* did not reliably differ across the HS and LS groups. Despite this, the trend indicates that the global task difficulty, as defined by the similarity of the source voices, had some effect on participants' ability to judge items as studied or not, even though voice information was not useful in determining whether items were old or new. In fact, results suggest that global difficulty affected discriminability, and not necessarily the rate of information accumulation. Evidence for this is found in the analyses of item RTs wherein no differences were observed between the LS and HS groups. This conclusion should be interpreted cautiously, however, as the trends were not consistently reliable, and part of the evidence is drawn from a null effect.

Clearer differences were observed by examining the influence of voice more closely, although most of the effects apparently reflected strategic responding or response biases. For example, participants in the HS groups, who studied words by either all males or all females, were consistently biased by the gender of the test speaker, both during old and new trials. When the test speaker's gender matched the studied genders, participants' hit rates were higher, relative to when the test speaker's gender was the opposite of the studied genders. This pattern extended to new trials; higher correct rejection rates were associated with opposite gender test voices. In short, although participants were instructed to ignore the test voice, and told that it held no discriminative value for old/new status of the item, the voice detail was apparently too salient to ignore. Similarly, participants in the LS groups, who studied words by two males and two females, used voice detail to respond strategically during the source judgment. When presented with a male test voice, for example, participants seemed to narrow the response set down to only the two males, yielding a chance response rate of 50%, instead of 25%. The apparent "boost" to source memory when the gender detail matched across study and test was simply another bias effect: Source judgments in the LS groups did not cross chance levels of performance, using 50% to estimate chance.

Although this research was intended to illuminate the role of similarity in item and source memory, with a working hypothesis that predicted increases in performance for the LS groups, the most reliable findings were bias effects. Although the effects were not as strong as intended, there were some non-bias effects in the current studies. For example, *d'* was generally higher when similarity was low, and participants in the LS groups responded to the source judgment more quickly than participants in the HS groups. The *d'* difference is most theoretically interesting. Whereas the difference in source RTs could be an artifact of the response bias based on voice (i.e., LS participants had a reliable voice cue on every trial by which to narrow down the response options; those in HS only had that voice cue on half of the trials), the *d'* difference suggests that global difficulty plays a role in item memory. This is similar to an effect reported by Dodson et al. (1998) using a similar paradigm. Participants in their high-similarity conditions performed less well on old/new item recognition, relative to participants in their low-similarity conditions. Although this effect is in direct conflict with some studies, in which no effect of source similarity is observed on item memory (e.g., Bayen et al., 1996; Ferguson et al., 1992; Lindsay & Johnson, 1991), it is consistent with other findings, in which item similarity influences old/new recognition performance (e.g., Nelson, Brooks, & Wheeler, 1975; Runquist, 1978). For example, when the conceptual similarity of items is high (e.g., all items are four-footed animals), old/new recognition performance suffers, relative to when those items are mixed with items from another conceptual category (Schmidt, 1985).

Dodson et al. (1998) interpreted their observed differences in item memory performance across the HS and LS conditions, which was generally replicated here, as one of response bias: When participants are tasked with making more difficult source judgments, they are biased to respond "new." That interpretation does not necessarily hold for the present findings. Whereas Dodson et al., only requested source judgments following "old" responses, I required source judgments for all items, with the goal of eliminating a "new" response bias. The *d'* result reported here seems to reflect general task difficulty, rather than strategic responding. Unfortunately, *d'* was so low in the current experiments (and, in fact, never reached 1.0) that any differences in memory were likely diminished, and reflected in the inconsistency of the finding. Future work will aim to increase overall *d'*, which should increase the "space" within which to observe an effect.

Research Aim 2: The Specificity of Source Memory

More theoretically meaningful data were observed by examining the specificity of participants' source memories. The current experiments replicated and extended the findings of Dodson et al. (1998). Specifically, Dodson and colleagues observed partial source memories in the absence of specific source memories. Although they observed evidence that source recollection was best characterized as a continuous process, they interpreted their data within a modified multinomial model, which typically assumes threshold-like recollection. In their model, source memory occurs following one of three routes; it can be specific, partial and guessed, or absent and guessed. Although their multinomial model nicely described their data, it would have difficulty describing the current data. Specifically, multinomial models predict discrete cognitive states, each of which is dependent upon the preceding state (see Bayen, Murnane, & Erdfelder, 1996; Dodson & Shimamura, 2000; Meiser & Bröder, 2002). For example, in Dodson et al.'s multinomial model, source memory, either specific or partial, can only occur following intact item memory. This is a product of their task: Source memory was only queried following "old" item responses. In the present research, I probed source memory following all item responses, which served two functions: 1) It eliminated the bias to respond “new” to relatively weak memories, which would allow participants to skip the source judgment, and 2) it allowed an estimate of specific and partial source memory to be obtained in the absence of item memory. According to some authors (e.g., Kurilla & Westerman, 2010), this is a plausible outcome, as it relies on a continuous recollection process. Given the implicit assumption of some researchers who utilize discrete-state multinomial models (e.g., Batchelder & Riefer, 1990), however, this should not be possible, as source memory is contingent upon item memory.

In the current studies, specific and partial source identification rates were operationalized with specific source identification (SSI) scores and partial source identification (PSI) scores (see Dodson et al., 1998). As discussed earlier, both scores are centered at zero; values reliably above zero reflect more accurate responding (e.g., a PSI score above zero reflects above-chance partial-source recollection). Consistent with every known theory of memory, SSI scores were highest when item memory was intact. Inconsistent with several theories of memory, however, SSI scores were also above-chance when item memory failed. In other words, when participants were unable to recognize items, a fundamental step in discrete, threshold models of memory, they still demonstrated above-chance ability to identify the specific source of the learned (but unrecognized) information.

Further evidence for a continuous view of the recollection process was obtained in PSI scores. Regardless of whether participants were able to recognize the item, PSI scores were reliably above zero, revealing above-chance partial source recollections. For example, participants may have failed to recall that the word *pickle* was originally spoken by Chloe, but, with above-chance frequency, they were able to respond to the source judgment with a within-gender error (e.g., responding “Anne”). In fact, PSI scores were higher when the participants were *unable* to recognize the item. Although memory strength was not sufficient to cross a hypothetical item recognition criterion, participants nonetheless demonstrated intact partial source memories. This finding is difficult to reconcile with dual-process theories and multinomial modeling approaches to source memory, both of which assume that recollection is either/or and that memories follow discrete stage-like processing. Recent multinomial source models, based on the assumption of continuous, rather than discrete, recollection processes seem capable of both predicting and explaining this finding.

The multinomial model depicted in Figure 13 was originally described by Batchelder and Riefer (1990), but was adapted to the four-source case, with a partial source memory component, by Dodson et al. (1998). Visual inspection clearly demonstrates an assumption inherent to all standard multinomial models: Source recollection is only possible following “old” item judgments. In this model, old items are recognized with probably D_1. Failing to recognize an old item occurs with probably $1 - D_1$, but participants can guess that an item is old with probably b. The outcomes of the various discrete memory processes can be observed by following the “branches” of the tree. For example, a participant can recognize an item (D_1), yet fail to recall its original source ($1 - d_1$). With probability P_a, the participant can recall partial source information, leading to either a correct guess (e_1) of true source, Erik, or an incorrect guess ($1 - e_1$) of the same-gender source, Art. The critical assumption of this sort of model, which differentiates it from signal-detection models, is that participants can only "guess" source information when they fail to retrieve the item (the [$1 - D$] pathways, and any pathway terminating in a 'g' parameter). Figure 14 shows an attempt to modify this model to suit the current paradigm, in which participants were assumed to have partial source memory (not guessing) in the absence of item memory. As can seen by examining the terminal responses and backtracking through the branches of this model, participants only guess when they fail to retrieve source information. In other words, this model assumes that participants' memory responses are usually "informed;" responses rarely reflect guessing. This assumption is supported by the PSI and SSI scores in the current experiments. Participants' responses indicated that their source judgments were usually at least partial recollections of the studied event. As can also be seen by examining this model, it is far too complex to be useful. In fact, attempts to simulate the current data with this model failed to converge upon a solution, either because behavioral performance was generally too poor or because the model had too many parameters.

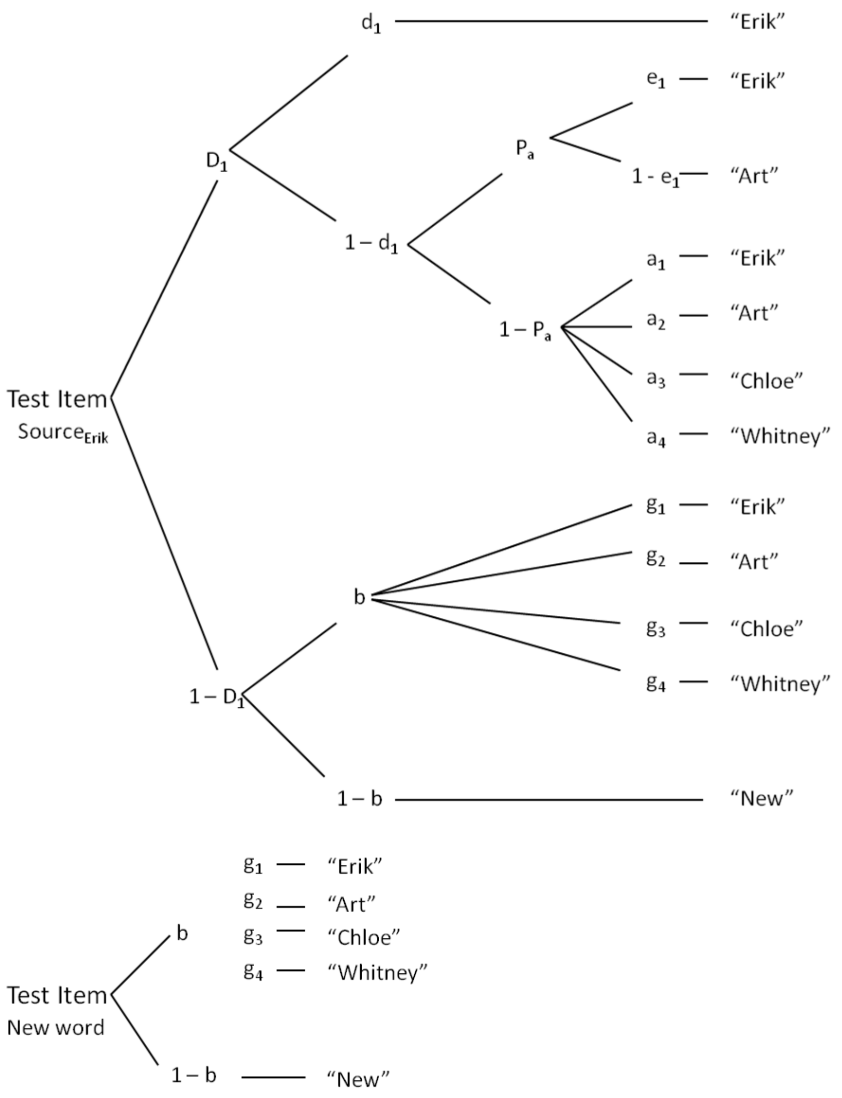


Figure 13. Multinomial processing tree model of specific and partial source memory during old (upper tree) and new (lower tree) recognition trials. D_1 = probability of detection Erik items as old; d_1 = probability of identifying the source of Erik items; a_i = probability of guessing that a detected item is from source i; b = probability of guessing an item is old; g_i = probability of guessing that an undetected item is from source i. Adapted from Dodson et al. (1998).

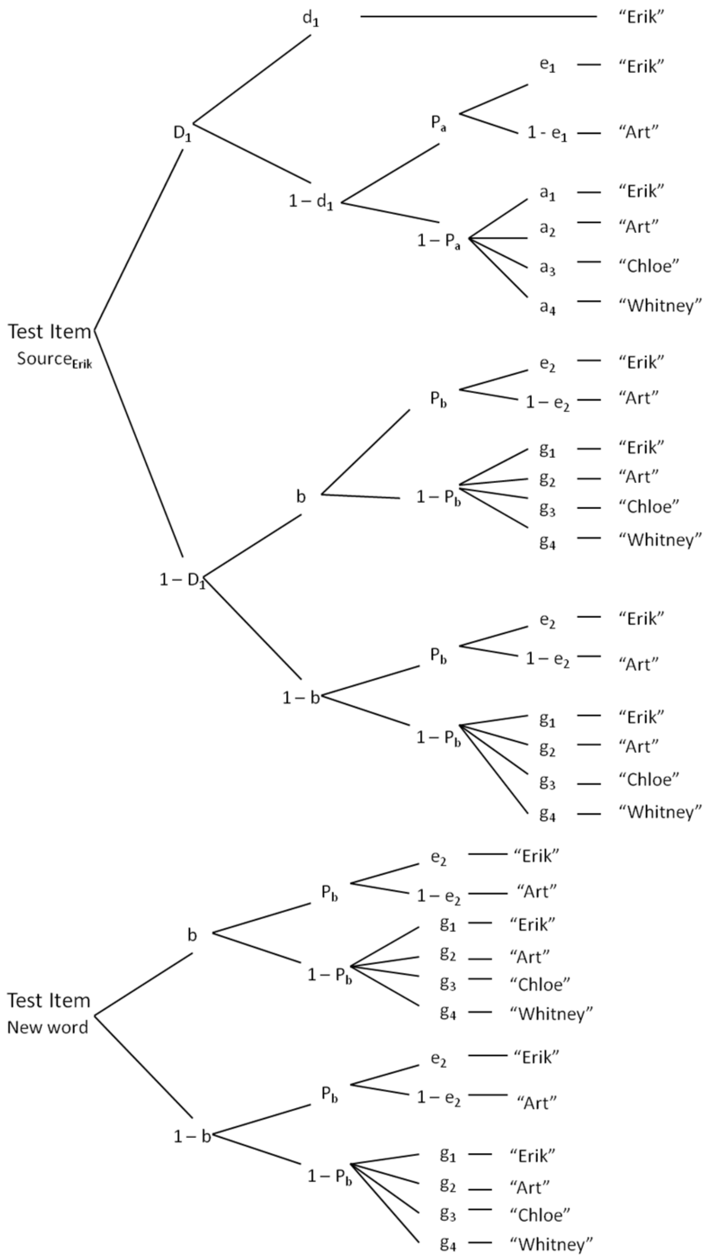


Figure 14. Modified multinomial processing tree model of specific and partial source memory during old (upper tree) and new (lower tree) recognition trials. D_1 = probability of detection Erik items as old; d_1 = probability of identifying the source of Erik items; P_1 = probability of identifying the source of Erik items; e_1 = probability of guessing the correct source of Erik items after identifying the gender; a_i = probability of guessing that a detected item is from source i; b = probability of guessing an item is old; g_i = probability of guessing that an undetected item is from source i. Adapted from Dodson et al. (1998).

Another possible reason for the failure of the model depicted in Figure 14 is that it is an attempt to model a continuous recollection process with a model designed to handle discrete, threshold-like recollection. Recent models of source memory have begun to incorporate the principles of signal-detection theory (e.g., Banks, 2000; DeCarlo, 2003; Glanzer et al., 2004; Hilford et al., 2002; Qin et al., 2001; Slotnick & Dodson, 2005; Slotnick et al., 2000; Starns et al., 2008), eliminating the need for the threshold assumption inherent in standard multinomial models. Wickens (2002) described a signal-detection model of source retrieval in which each source was represented by an equal-variance Gaussian distribution, as in equal-variance signal-detection models described earlier. According to this model, source discriminability is described by the overlap of the two distributions; the less overlap, the more discriminable the sources. Based on where the evidence falls, relative to a decision criterion, participants choose one source over the other. This model, however, is only one-dimensional, and must be expanded into a multivariate model in order to accurately describe data from combined item and source recognition tasks.

The multivariate, signal-detection model depicted in Figure 15 is an idealized model of several recent attempts to describe item and source recognition tasks, when there are two sources of information to be discriminated (e.g., Slotnick & Dodson, 2005; Starns et al., 2008). The model includes three bivariate normal distributions, one for Source A items, one for Source B items, and one for new items. The x-axis represents item strength, with strength increasing from left to right. As can be seen in the model, new items are further left on this dimension, relative to Source A and B items, which do not differ from one another in the standard case. Source strength lies along the y-axis. In this model, source evidence favors either Source B (higher on the axis) or Source A (lower on the axis). New items are equidistant between Sources A and B, reflecting a lack of association to either source. As in signal-detection models, the evidence strength within each distribution is described by a Gaussian distribution, and would be plotted in z-space (not depicted here, see Slotnick and Dodson, 2005, for a 3D treatment).

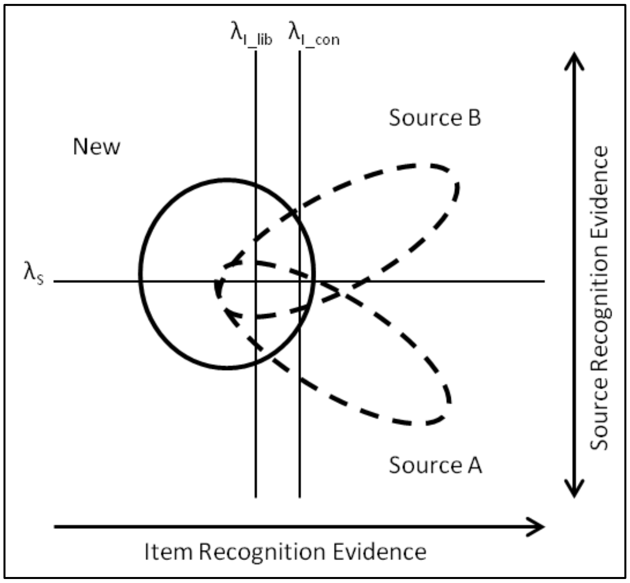


Figure 15. The multivariate signal-detection model for item and source recognition (adapted from Starns et al., 2008). The ovals are cross-sections through hypothetical 3D distributions, where ‘strength’ lies on the z-axis. The solid horizontal and vertical lines represent possible response criteria for item (vertical) and source (horizontal) judgments.

As described by Starns and colleagues (2008), each of the ovals in Figure 15 is centered at the mean of the distributions for item and source strength, and can be thought of as a horizontal slice through the 3D distribution. The shapes of the distributions are contingent upon the correlation between values on the item and source dimensions. DeCarlo (2003) fit several data sets with this model, and observed no correlations for new items (as expected), but strong negative correlations for Source A items. That strong negative correlation indicated that as item strength increased, source evidence was lower on the source dimension (i.e., it favored Source A). Positive correlations were observed for Source B items. This relationship is why the contours for the Source A and B items are skewed down and up, respectively, and is consistent with the finding that accurate item recognition is associated with accurate source recognition (Glanzer et al., 2004).

Decisions are modeled within this framework by assuming multiple decision criteria, at least one for item recognition decisions and one for source judgments. The item recognition criterion is placed somewhere on the x-axis to optimize performance; as in signal-detection theory, as the criterion is shifted leftward, responding becomes more liberal. As in Starns et al. (2008), this is hypothetically represented within the figure by λ_{I_lib} and λ_{I_con}, for liberal and conservative criteria, respectively. Another decision criterion is adopted for the source judgment. This criterion (λ_S) is placed along the y-axis; values above it are judged Source B and below it are judged Source A. The method by which the model predicts source memory in absence of item memory is depicted by examining the regions of space that fall to the left of the item recognition criteria. Assuming a liberal response criterion, only a small portion of the Source A and B strength falls to the left of criterion, yielding “new” item responses and near-chance source responses. When the item recognition criterion is more conservative, however, the items are still called new, but the source dimensions are more differentiated, which should yield more accurate source recognition performance. Starns and colleagues (2008) manipulated participants' response criteria by changing test instructions to indicate that either 25% or 75% of the items were old. Doing this, they observed data entirely consistent with their model. In fact, item response bias tended to be more conservative than liberal in the current experiments, which is consistent with the multivariate signal-detection explanation for accurate source memory in the absence of item memory.

To adapt this two-source model to fit the four-source data from the current experiments, additional dimensions would need to be described. In addition to the item recognition (x-axis) and strength (z-axis) dimensions, third and fourth dimensions (e.g., y_1-axis for male sources and y_2-axis for female sources, in the LS groups) would need to be added. Visually, this proves difficult, but conceptually, it is possible. By positing an extra dimension along which two other sources lie, the multivariate signal-detection model would provide a concise, but visually unappealing, model of the current data. No existing multinomial models are capable of describing source memory in the absence of item memory without incorporating fundamental changes. For example, Starns et al. (2008) discussed how multinomial models could be modified to predict source memory in the absence of item memory. Specifically, extant multinomial models would need to predict a third memory state. In addition to "detected" and "undetected" states, such models would need to incorporate an intermediate state of knowledge. This modification, however, still subscribes to the notion of discrete states; it simply posits more of them. Models without discrete, threshold states more elegantly describe partial source recollection, and source recollection in the absence of item memory, two effects that clearly emerged in the present studies, and which are difficult to accommodate in standard dual-process models.[10]

Research Aim 3: The Nature of Recognition Memory

Klauer and Kellen (2010) recently called into question the utility of ROC analyses for adjudicating between models of memory. They suggested, for instance, that threshold models can predict curved ROCs if a plausible mapping function relates latent cognitive states to the rating scale. Further, when binary old/new decisions are used, discrete state models predict linear ROCs, but when confidence estimates are used, high-threshold models are able to predict curved ROCs (Malmberg, 2002). Although Klauer and Kellen (2010) recently compared the performance of several models in explaining ROCs from item and source memory judgments, none of the models that they considered were capable of explaining partial source recollection, or source recollection in the absence of item memory. (The models included a hybrid signal-detection model proposed by Hautus, Macmillan, and Rotello (2008), a variable-recollection dual-process model proposed by Onyper et al. (2010) and a new discrete-state model.) Because none of those models are capable of describing the current source monitoring

---

[10] Note that Schutz and Broder (2011) recently provided evidence that high-threshold models (e.g., Bayen et al., 1996) and multivariate signal-detection models (e.g., DeCarlo, 2003) both adequately describe the data when model fits are considered without rating-based ROC analyses. The added value of multivariate signal-detection models over high-threshold models, however, is that signal-detection based models predict partial source memory, and source memory in the absence of item memory. High-threshold models are unable to account for this finding.

data, they will not be considered as candidates for explaining the current ROC analyses. Rather, the ROC analyses from the current experiments will be explained by their conformity to the overarching theories of memory discussed earlier, single- and dual-process theories. ROC analyses are still one of the predominant methods by which to compare and contrast different models, and those in the present study collapsed across item and source recognition, consistent with the notion that they are based on the same underlying processes (as recommended by Klauer & Kellen, 2010)

Overall, the pupillary ROC analyses provided evidence that was predominantly consistent with continuous models, and inconsistent with threshold-based models. Across all three experiments, the ROCs were consistently curvilinear, and the *z*-ROCs were predominantly linear. The shape-based conclusions were backed up by subject-based ROC analyses. Whereas the quadratic constant for ROCs was always negative, and statistically different from zero, the same term in the z-ROC analyses was typically near zero (but reliably different from zero, given the number of data points fit), consistent with a linear function. Although the *z*-ROC was not as linear as a signal-detection model would predict, it was also not as curved as a dual-process account would predict. On the whole, the results are broadly consistent with single-process, signal-detection based views, which predict curved ROCs and linear *z*-ROCs. As summarized by Parks, Murray, Elfman, and Yonelinas (2011), dual-process theories predict strongly curvilinear *z*-ROCs during tasks that rely on recollection (e.g., source judgments); *z*-ROCs, they argue, should become more linear as familiarity plays a greater role in source retrieval. Because the sources in the present experiment were equated for familiarity (e.g., they were used equally often), there was no reason why one source should have been more familiar than any other source.

Although Yonelinas (1999) observed increases in the linearity of *z*-ROCs as familiarity increased, many studies have observed linear *z*-ROCs in the absence of manipulations intended to enhance the use of familiarity (e.g., Glanzer et al., 2004; Qin et al., 2001; Slotnick et al., 2001). For example, Qin and colleagues (2001) presented participants with spoken statements for a source memory test and observed clearly linear *z*-ROCs. They argued that recollection must be behind accurate source memories, and that a continuous recollection process is responsible for the linearity. Further, they suggested that the apparent continuous nature of recollection was the direct result of the complexity of the material; previous studies, they argued, in which U-shaped *z*-ROCs were observed, used relatively simple materials that do not permit *levels* of recollected detail.

Parks et al. (2011) recently examined this complexity explanation by using materials that varied in their level of complexity from relatively simple (auditory words) to complex (audiovisual sentences). By plotting source ROCs and *z*-ROCs, Parks et al. observed that increases in the complexity of stimulus materials resulted in a curving of the source ROCs (and a concomitant flattening of the *z*-ROCs), which proves to be difficult to explain under standard dual-process accounts. By making the assumption, however, that recollection is a threshold process, with strength characterized by a Gaussian distribution, Parks and colleagues were able to rectify curvilinear source ROCs and linear *z*-ROCs with dual-process theory. This modified model, the variable recollection dual-process (VRDP) model (Sherman, Atri, Hasselmo, Stern, & Howard, 2003; for similar models, see DeCarlo, 2003; Kelley & Wixted, 2001; Onyper, Zhang, & Howard, 2010) assumes that recollection is characterized by gradations in strength that fall above the threshold for simple materials and around the threshold for complicated materials. With simple materials, high-confidence responses reflect threshold recollection processes, but with complex materials, *relatively* high-confidence responses (e.g., 5 or 6) can be associated with recollection. This does not mean that recollection is a graded decision process; the decision process is still categorical, with familiarity queried as a backup (Parks et al., 2008; Yonelinas, 1994). In essence, the effect of increases in complexity causes the recollection strength to be weaker and more variable.

This argument is similar to that provided for the effects of feature overlap (e.g., stimuli consisting of all suburban houses, which share many features). Elfman, Parks, and Yonelinas (2008) appealed to the complementary learning systems model (CLS; Norman & O'Reilly, 2003) to explain the finding that stimuli with high feature overlap yield curved ROCs (and linear *z*-ROCs). CLS is a biologically plausible model of the functions of medial temporal lobe (MTL) systems in encoding and retrieval. According to this model, the hippocampus encodes items using a fast pattern separation process; it is also the sole supporter of recollection. Other areas within the MTL (e.g., parahippocampal cortices) support the backup process, familiarity. When an item is presented for retrieval, the hippocampus first responds by initiating a pattern completion process, filling in the item with stored traces. A recollection decision is then made by determining whether the pattern completion process surpassed threshold. If not, then familiarity processes are consulted. According to Elfman et al. (2008), this model can explain the effect of feature overlap by assuming that high feature overlap yields a slower, more graded pattern completion process during retrieval. The same type of explanation was invoked by Parks et al. (2011) to explain the effects of complexity: Recollection, in the model, was still a threshold process, but it was characterized by a wider Gaussian distribution, which resembles a continuous process.

Although these models are capable of predicting the ROC results described here, they are incapable of fully explaining the observed pattern of data. Specifically, VRDP, DPSD, and CLS do not incorporate any method by which source information, whether specific or partial, can be retrieved in the absence of item memory. Each of those models considers recollection to be a threshold process, associated with the highest level(s) of confidence. The data presented here argue firmly against that prediction. Even in situations known to produce low levels of confidence (i.e., item misses), participants were still able to retrieve enough source information to support specific or partial source recollection. Further, these data were observed with clearly curvilinear ROCs and linear *z*-ROCs. Taken together, the behavioral and ROC data do not suggest that recollection is a threshold process. Rather, they suggest a graded, continuous recollection process.

The last method by which dual-process and threshold theories attempt to explain the pattern of ROC results observed here is by positing the existence of “unitized familiarity” and suggesting that this supports source retrieval. As described earlier, unitization reflects the combination of item and source material into a single unit, such that the source becomes a feature of the item, rather than an episodic detail associated with the item. Accurate recollection of details, therefore, can theoretically be based on a graded unitized familiarity process, and not necessarily on recollection, which is typically assumed to be a threshold process. Although unitized familiarity is typically used to explain ROC data in associative recognition tasks, source memory tasks are equally recollection-based, and the predictions of unitized familiarity have been extended to the source memory literature (see Diana, Yonelinas, & Ranganath, 2008, 2010). Diana and colleagues discuss the reality of unitized familiarity largely on the basis of neuroimaging results showing that the hippocampus and parahippocampus are active during successful encoding and retrieval of source memories (Davachi, Mitchell, & Wagner, 2003; Kensinger & Schacter, 2006; Ranganath et al., 2003), whereas the perirhinal cortex is active for familiarity-based memories, but not with source memories (see Davachi et al., 2003; Kensinger & Schacter, 2006; Uncapher, Otten, & Rugg, 2006; Weis et al., 2004). Yonelinas, Kroll, Dobbins, and Soltani (1999) had earlier described encoding procedures that promote unitization, such as asking participants to make color-word associations by imagining the object described by the word in the color in which it was written (the most common example is the word ‘elephant’ written in red font, yielding a red elephant).

One of the most highly-cited neuroimaging results relevant to the concept of unitized familiarity is that of Staresina and Davachi (2006). Participants in their study encoded words printed on one of four background colors; their task was to imagine that item in the arbitrary color of the background, a procedure known to promote unitization (particularly when the color is plausible for the item). Their results revealed that, contrary to other source memory findings, the perirhinal cortex was active during source (background color) retrieval. They concluded that the perirhinal cortex supported associative processing of intra-item details. If item and source details become unitized into a single episodic trace, then it is possible that source retrieval tasks can yield curvilinear ROCs based on unitized familiarity, without detriment to the threshold recollection assumption inherent to dual-process models.

Recent evidence suggests that item and source information can be processed in such a way as to promote unitization, yielding linear *z*-ROCs that are still consistent with dual-process theories (e.g., Diana, Yonelinas, & Ranganath, 2008; Quamme. Yonelinas, & Norman, 2007; Yonelinas et al., 1999). For example, participants in Diana et al.’s study encoded words and colors under conditions that either promoted unitization (i.e., rating the plausibility of the item in the background color) or did not (i.e., making pleasantness rating or size judgments depending on the background color). As predicted, they observed that increasing the contribution of familiarity to the source judgment task (e.g., via unitization) increased the curvilinearity of the source ROCs, and also increased familiarity estimates from DPSD. From this, the authors concluded that discrepancies in neural and ROC data from source memory tasks may be explainable by appealing to unitized familiarity, and not by abandoning the threshold recollection assumption in many dual-process theories. They also concluded that source recognition tasks are not process-pure measures of recollection, and on this latter point, there is wide agreement.

Although unitized familiarity provides a plausible account for the finding of curvilinear ROCs, the present studies included no manipulations or encoding procedures to promote unitization. In fact, the assignment of items to sources to retrieval cues was entirely random. If unitized familiarity were to play a role in explaining the ROC data reported here, one would have to assume that participants spontaneously unitized the voice and word information during encoding, which is a bit of a logical stretch. Further, Mickes, Johnson, and Wixted (2010) recently demonstrated that the concept of unitized familiarity is highly associated with recollective processes, including remember/know and unexpected cued recall. A more reasonable dual-process account for the current data would seem to appeal to the complexity of the stimulus materials. As reviewed by Parks et al. (2011), highly complex material is more likely to yield curvilinear ROCs.

But how complex is complex enough to yield curved ROCs that are still compatible with dual-process theories? Parks and colleagues (2011) suggested that single words, presented auditorily, yield the standard dual-process prediction of linear ROCs (and curved *z*-ROCs). It is only when the material was audiovisual, and expanded to full sentences, that the recollection process “broke down” and became more graded, yielding curved ROCs (and linear *z*-ROCs). It is arguable that the present stimuli fell somewhere in between these two extremes, and were instead “somewhat complex.” Because of the many methods proposed to explain discrepant findings within dual-process theory (e.g., complexity, unitized familiarity, variable recollection), a more parsimonious explanation is simply that participants are able to recollect to varying degrees, consistent with the predictions of signal-detection models (e.g., UVSD or CDP). Whereas dual-process theories propose different mechanisms based on the quality of the stimuli or task demands, signal-detection views offer a simple, straightforward account of episodic memory: Memory strength is always present, but is often not strong enough to support detailed recollection, instead yielding less detailed memories. Critically, this means that recognition can occur along various points of confidence, or strength, which would reveal ROCs exactly as observed in the present study. Future work will focus on the utility of pupillary ROCs, above and beyond those based on confidence estimates, by using manipulations known to influence memory strength and the shape of the curve.

Research Aim 4: The Functional Role of Eye Movements

Lastly, one of the major goals of the present research was to determine whether, and to what extent, eye movements are reinstated across encoding and retrieval. Earlier work established a clear link between non-visually guided patterns (NVGPs) of eye movements and long-term memory retrieval (see Ehrlichman & Micic, 2012, for a review). For example, when retrieving more difficult information from long-term memory, spontaneous saccadic eye movements are more frequent, relative to when retrieving easier information, or information from working memory (Ehrlichman et al., 2007; Micic et al., 2010). Research on NVGPs, however, has been silent on whether or not these eye movements were functional for memory, or whether they reflected reinstatements of eye movements from encoding.

Research on saccadic reinstatement has documented many cases in which eye movement patterns were reinstated across encoding and retrieval (e.g., Ballard, Hayhoe, & Pelz, 1995; Holm & Mäntylä, 2006; Richardson & Spivey, 2000; Ryan, Hannula, & Cohen, 2007; Spivey & Geng, 2000; Yarbus, 1967, as cited in Tatler et al., 2010). In most of these studies, eye movement patterns were reinstated, but unrelated to improvements in memory retrieval. For example, Johannson, Holsanova, Dewhurst, and Holmqvist (2011) restricted participants' eye movements during the encoding of auditory and visual information. Although this impaired memory relative to a free-viewing encoding condition, participants did not reinstate the single-fixations when allowed to move their eyes at retrieval. From this, they concluded that eye movements at retrieval are not reinstatements of those from encoding. This is inconsistent with their earlier findings, in which they observed that participants spontaneously move their eyes to previously-viewed locations (now blank) during memory retrieval (Johansson, Holsanova, & Holmqvist, 2006) and with other findings in which participants spontaneously reinstated single fixations from restricted-viewing encoding conditions (Laeng & Teodorescu, 2002). Johansson et al. (2011) did, however observe that restricting eye movements at retrieval had a detrimental effect on performance. When participants were not allowed to freely move their eyes during the recall of a previously-viewed scene, their recall performance became less detailed, both qualitatively and quantitatively. This finding, they suggested, demonstrated that eye movements are not purely epiphenomenal, and that spontaneously reinstated eye movements have a functional role for memory retrieval (see also Holm & Mäntylä, 2006; Mäntylä & Holm, 2007).

In the current studies, eye movements were manipulated across encoding and retrieval in order to determine whether reinstated (or disrupted) eye movements have a facilitative (or detrimental) role in memory retrieval. Experiment 1 produced the most reliable results. Specifically, participants spontaneously reinstated eye movements to former regions of interest during memory retrieval (when the screen contained no visual information other than an empty 9-box grid) above chance-levels of fixations. Further, spontaneously reinstated eye movements were associated with more accurate source memories and faster item and source judgments. Externally-cueing participants to reinstate either single fixations (Experiment 2) or full eye-movement patterns (Experiment 3) did not facilitate retrieval. In fact, there was a trend for item *d’* to decrease across the three experiments (see Figure 16). Although this trend was not reliable, $F(2, 176) = 1.02$, $p = .36$, it clearly depicts that *d’*, despite already being quite low, decreased as the participants’ eye movements were more manipulated.

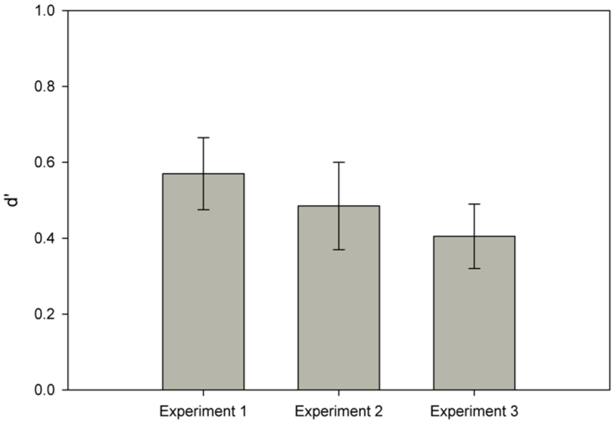


Figure 16. Average *d'* across Experiments 1, 2, and 3. Error bars represent standard error of the mean.

Upon first consideration, the eye movement results are consistent with those of Holm and Mäntylä (2006; Mäntylä & Holm, 2007), who observed that reinstated eye movements were associated with increases in remember responses in the RK procedure. Although there are many differences between the two tasks, researchers have often likened source retrieval to R responses (e.g., Conway & Dewhurst, 1995). The rationale is that source retrieval typically involves recollection (but see the discussion of unitized familiarity, above), and that participants are often able to report the subjective quality of their memories, using "remember" to index some degree of recollection. Further, despite the dual-process assumption that recollection is slow and effortful, R responses are typically fast and accurate, a qualitative pattern replicated here during accurate source memories that were also characterized by high rates of reinstated eye movements. Inconsistent with this, however, was the relationship between low rates of reinstated eye movements and the speed with which participants made item and source judgments. Specifically, the RT patterns for very low reinstatement rates mirrored that for very high reinstatement rates; they were both very fast, relative to mid-level reinstatement rates. Although the directionality of this relationship is unclear (i.e., based on the pattern of results, it is impossible to determine whether the effect is facilitative or detrimental), it certainly suggests some benefit for memory retrieval based on oculomotor reinstatement. Curiously, the pattern suggests that the memory processes in trials with very low and very high rates of refixations are the same (or at least similar), and that another process or strategy was engaged during the mid-level rates. Although it is tempting to relate these processes to familiarity (typically assumed to be fast and automatic) and recollection (typically assumed to be slow and effortful), that conclusion may be a bit premature. For instance, very high rates of refixations were associated with both faster RTs and more accurate source judgments. As described earlier, typical conceptualizations of recollection cannot explain a fast recollection process; rather, dual-process theorists would need to appeal to a fast unitized familiarity process to explain this result. As discussed above, the current experiments did not encourage unitization processes, making this interpretation speculative, at best.

So how do very low and very high rates of refixations result in the same, relatively speeded, RTs? I suggest that the results reflect one basic memory process, characterized by its strength of evidence (similar to the multivariate signal-detection models discussed above, see Starns et al., 2008), but differentiated based on its spread of activation to associated details. As Ferreira et al. (2008) suggested, eye movements are part of a rich, detailed encoding trace, and activating one part of this trace will typically activate other parts of the trace. For example, when presented with a test item, participants often spontaneously engage in the oculomotor behaviors associated with that item. Strong memories, as discussed above, typically result in fast decisions. The data reported here suggest that these fast decisions do not always contain the same level of detail. When presented with a test item, the resulting item memory strength could be very strong, yet undifferentiated, and still yield a fast decision. On the other hand, the memory could be very strong, and differentiated, and also yield a fast decision. The difference between the two types of memory lies in the amount of details they contain. Whereas one memory is fast, and lacks associated detail (yielding low rates of refixations and low source accuracy), the other memory is also fast, but contains rich details (yielding high rates of refixations and higher source accuracy). The data from Experiment 1 were consistent with this interpretation: When participants were able to retrieve source details, they were also more likely to reinstate their encoding fixations.

A more thorough account of the data, however, should attempt to explain the reliable findings from Experiment 1 within the context of the null results in Experiments 2 and 3. Although it is ill-advised to draw conclusions from null results, the spectacular lack of effects cannot be ignored (see below for future directions that might provide more incisive tests about the role of eye movements in memory retrieval). Recall that the present work was motivated, in part, by the anecdotal distinction between remembering (clearly recognizing a coworker at the mall) and knowing (knowing that you know that coworker, in the absence of any qualitative details), Mandler's (1980) 'butcher on the bus' phenomenon. The former type of memory is what Tulving (1983) had in mind when he coined the term "episodic memory;" episodic memories are memories that involve some degree of episodic detail, such as recognizing spatiotemporal context. These memories are distinct from *implicit* memories, which can be observed when prior exposure influences future behavior in the absence of overt awareness of the prior exposure (Graf & Schacter, 1985). In fact, Schacter and Tulving (1994) suggested that implicit memory phenomena reflect the function of knowledge (perceptual or semantic), whereas explicit memory phenomena reflect episodic retrieval. Evidence for the separability of implicit and explicit memory systems is often observed in empirical dissociations (see Roediger & McDermott, 1993) and neuroanatomical dissociations (e.g., Graf & Schacter, 1985; Warrington & Weiskrantz, 1968). In priming studies, amnesics, for instance, and patients with hippocampal lesions typically show intact implicit memory, but faulty explicit memory.

Priming studies are truly the paradigm case for the existence of implicit memory; if implicit and explicit memory are truly separate memory systems, then the reinstatement or alteration of episodic details should have no effect on priming. Empirically, however, several researchers have observed large episodic effects in implicit memory. For example, in perceptual implicit priming studies, priming is reduced when the typeface of words changes across the prime and the target (see Tenpenny, 1995, for a review). Unless some episodic detail is retained in the implicit memory trace for the primed item, there should be no reason for typeface to influence performance. McKone and French (2001) discussed multiple meanings of the term *episodic*, and how episodic-like effects in implicit memory may be explained by appealing to different definitions. They described three meanings, the first of which is that episodic memories make autobiographical reference to one's past (something amnesic patients, discussed above, cannot do). The second is that episodes code the "intrinsic context" of an item, including the properties that are unavoidably processed while performing the task. For example, during visual word priming, the intrinsic context contains visual processes, orthographic identification, phonological processing, etc. Using this definition of "episodic" explains why priming studies demonstrate episodic effects. The third definition is that episodes code "extrinsic context," which consists of information processed during the task, but which is not vital to completing the task (e.g., time of day, environmental context, mood, etc.). Changes in extrinsic context underlie classic demonstrations of context-based memory retrieval (e.g., Godden & Baddeley, 1975, see Smith & Vela, 2001, for a review), but have been inconsistently linked to differences in implicit memory performance (see Jacoby, 1983; McKone & French, 2001; Mulligan, 2011; Parker, Dagnall, & Coyle, 2007; Parker, Gellatly, & Waterman, 1999; Smith, Heath, & Vela, 1990).

Based on these definitions, the eye movement manipulations in the present research were intended to foster intrinsic context-based memory effects, consistent with McKone and French's (2001) second definition. To date, little research has examined the influence of intrinsic context reinstatement on episodic or implicit memory. In the present study, the external environment remained the same throughout testing and any mood-based changes over the brief course of the experiment were expected to be minimal (i.e., extrinsic context was constant). Rather, in Experiment 1, participants' own eye movements at retrieval were expected to reveal intrinsic context reinstatement. To some degree, they did, as reinstated eye movements were associated with better source memory and faster evidence accumulation. Experiments 2 and 3 involved external-cues to prior intrinsically-relevant (or irrelevant) states, with little success. Cueing previously-viewed locations or entire sequences of prior eye movements had no effect on participants' episodic memory performance (and, in fact, created a negative trend).

One potential explanation for the finding that externally-cued intrinsic contexts failed to influence episodic memory could be that intrinsic context effects act upon implicit memory. In the literature on eye movements in memory, one key finding supports this possibility: Amnesics, who have explicit memory deficits, but intact implicit memory, demonstrate reliable eye movement-based memory effects in studies of face and scene processing (e.g., Ryan et al., 2000; but see Smith & Squire, 2008). Laeng and colleagues (2007) observed that amnesic patients, despite being unable to retrieve new semantic facts from episodic memory, nevertheless reinstated the eye fixations that were executed during the encoding of those facts. By contrast, eye movement studies of amnesic patients' relational memory (which relies upon hippocampal processes, like recollection-, or detail-, based memories) demonstrate no such eye-movement-based effects. Whereas neurologically intact controls fixate on the locations of changes in visual scenes, even in the absence of awareness of the change, amnesic patients show no such effect (Hannula et al., 2007; Ryan et al., 2007; Ryan et al., 2000). Although eye movement-based memory effects are predicted by hippocampal activity during relational memory tasks (Hannula & Ranganath, 2009), their existence in patients with hippocampal amnesia suggests that their locus resides in the implicit memory system (or, alternatively, that the hippocampus selectively supports relational processing, as their studies were designed specifically to reveal that). Because the present study was episodic in nature, it is likely that effects of implicit memory were overshadowed by the episodic demands of the task. Had the task involved an implicit measure of memory, such as stem-completion, intrinsic context effects would have been more likely to emerge. Future work will adopt this approach, in order to determine whether intrinsic context, as operationalized via eye movements, influence implicit or explicit memory processes.

Finally, the present results add to the growing, and somewhat inconsistent, body of research on the putatively functional role of eye movements across learning and retrieval. Abundant evidence suggests that eye movements play no functional role in memory processes, and are better described by appeals to *spatial indexing*, whereby spatiotemporal information is linked to internal representations in order to offload from working memory (e.g., Richardson & Kirkham, 2004; Richardson & Spivey, 2000; Ballard et al., 1997). Such an explanation falls directly in line with O'Regan's (1990) concept of the world as an external memory store. Although variants of this theory do not necessarily subscribe to the notion that an external memory store precludes the existence of an internal memory store, proponents tend to remain agnostic as to whether such an internal store is necessary (e.g., Richardson et al., 2009).

On the other hand, several researchers have found evidence to suggest that eye movements, while not necessarily reinstatements of those from encoding (e.g., Johansson et al., 2011), are functional for memory retrieval (Ferreira et al., 2008; Laeng & Teodorescu, 2002). These findings suggest that integrated memory representations are derived from perceptual and linguistic experience, and that such memory representations are reactivated upon retrieval. If eye movements are one of the components of this rich memory trace, then a retrieval cue that sufficiently activates the trace should also activate the eye movement pattern. This account is consistent with Kent and Lamberts' (2008) notion of memory retrieval as mental simulation, by which they argue that encoding-retrieval interactions (e.g., transfer-appropriate processing, encoding specificity, etc.) reflect the ability of the cognitive system to manipulate internal representations while relying upon the perceptual processes that were engaged during encoding (note that this viewpoint is consistent with those of Kolers and Roediger, 1984, or Barsalou, 1999). Several recent memory theories make similar claims regarding the necessity of contextual and perceptual information from the environment (e.g., Barsalou, 1999; Ferreira et al., 2008; Glenberg, 1997; Rubin, 2006). In fact, Barsalou's (1999, 2008) perceptual symbol systems theory directly assumes that when participants are asked about information in memory, they engage in a ‘mental simulation’ on the relevant symbols.

Kent and Lamberts (2008) described two types of mental simulations capable of supporting cognitive processes relevant to memory retrieval. The first type are ‘explicit simulations,’ or simulations that are based on retrieval from episodic memory. During explicit simulations, mental operations work upon the perceptual symbols that were present during encoding, and give rise to feelings of recollection (as in autobiographical memory, e.g., Conway, Pleydell-Pearce, Whitecross, & Sharpe, 2002). The second type are implicit, or unconscious, simulations, which are activated during implicit memory tasks and do not rely upon retrieval of episodic detail. The major qualitative difference between the two types of simulation is that they result in two different experiences of memory. This is perhaps because they rely upon different mechanisms that act upon the same perceptual symbols (Barsalou, 2008), or because the “format” of the information differs across the two forms of information (Hegarty, 2004). Given the dissociations between implicit and explicit memory discussed above, it seems more likely that, if one assumes that mental simulations support cognitive processes, the content of those simulations should differ across episodic and implicit memory. Because implicit memories are generally impervious to changes in external context, it is sensible to assume that the mental simulations underlying those memories do not contain the relevant extrinsic detail.

To support their theory of mental simulations, Kent and Lamberts (2008) described myriad neuropsychological demonstrations of encoding-retrieval ‘matches’ in brain activity (e.g., Goldberg, Perfetti, & Schneider, 2006; Otten, 2007; Simmons et al., 2007), suggesting that the processes that support successful encoding are also recruited during retrieval. They also appealed to many of the eye movement studies discussed above. Specifically, they cited evidence from Laeng and Teodorescu (2002), Mäntylä and Holm (2006; Holm & Mäntylä, 2008), and Johannson et al. (2006) suggesting that eye movements play a functional role in memory processing. For example, Tremblay et al. (2006) observed improved recall ability in participants who spontaneously engaged in rehearsal-like eye movements following encoding of visual-spatial material. The results from the present study, and the demonstrations from Richardson and colleagues (Hoover & Richardson, 2008; Richardson & Spivey, 2000; Spivey & Geng, 2002) offer conflicting views. Specifically, eye movements, and reinstated eye movements, are not consistently associated with improvements in memory performance. This finding is in stark contrast to Kent and Lamberts’ (2008, p.95) prediction that “sensory-motor encoding procedures…can aid retrieval when they are re-enacted.” The current results demonstrate that externally cueing the re-enactment of encoding procedures does not facilitate memory, and Richardson and Spivey (2000) observed no effect of reinstated eye movements on retrieval success. Although it can be argued that the lack of relationship between eye movement patterns and retrieval success in the present and above-mentioned studies reflects the fact that none of the studies directly tapped visuospatial memory information, if memory traces are mental simulations of the encoding event, some defined relationship would be expected to emerge. In fact, Godijn and Theeuwes (2012) observed disruptive effects of cued eye movements on retrieval from working memory, similar to the pattern observed in *d'* across the three experiments reported here. This does not, however, rule out the possibility that maintaining unnatural fixations is a cognitively demanding task in and of itself, one which draws resources away from the primary task. Future research will focus on the role of reinstated eye movements using a three-phase memory procedure (e.g., Jacoby et al., 2005). Using this procedure, one can document the degree to which participants naturally reinstate eye movements at Test 1 and use it to predict memory performance on Test 2. Clearly, there is much work to be done to determine the role of eye movements in memory retrieval.

General Conclusions

The present research revealed new insights in episodic memory, and the role of eye movements in retrieval from episodic memory. In contrast to dual-process theories, or theories that posit a threshold recollection process, episodic memory appears to be characterized by a graded continuum of memory strength. The finding of partial source recollection, and of source retrieval in the absence of item memory, strongly suggests that people do not rely upon categorical either/or memory processes, even when those memories should be based on recollective processes. Consistent evidence was observed via pupillary ROC and *z*-ROC analyses. Whereas dual-process theories predict curvilinear ROCs and linear *z*-ROCs, signal-detection views predict the reverse, which was the pattern consistently observed here. Lastly, eye movements appear to have a role in retrieval processes, as they are spontaneously reinstated during retrieval, but it is possible (and likely) that those eye movements better describe the function of implicit, relative to episodic, memory. Taken together, the results support a multivariate signal-detection view of recognition memory, with rich, detailed memory traces capable of integrating rich details consisting of visual, spatial, temporal, and motoric properties.

References

Aggleton, J. P., Vann, S. D., Denby, C., Dix, S., Mayes, A. R., Roberts, N., & Yonelinas, A. P. (2005). Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia, 12,* 1810-1823. doi: 10.1016/j.neuropsychologia.2005.01.019

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419-439. doi: 10.1006/jmla.1997.2558

Althoff, R. R. & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 997-1010. doi: 10.1037/0278-7393.25.4.997

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The “blank screen paradigm.” *Cognition,* 93, 79-87. doi: 10.1016/j.cognition.2004.02.005

Altmann, G. T. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247-264. doi: 10.1016/S0010-0277(99)00059-1

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7,* 66-80. doi: 10.1162/jocn.1995.7.1.66

Ballard, , D. H., Hayhoe, M. M., Pook, P. K., Rao, R. P. N. (1997). Pointing the way. *Behavioral and Brain Sciences, 20,* 758-763. doi: 10.1017/S0140525X97431618

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459. doi:10.3758/BF03193014

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science, 11,* 267-273. doi: 10.1111/1467-9280.00254

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577-660.

Barsalou , L. W. (2008). Cognitive and neural contributions to understanding the conceptual system. *Current Directions in Psychological Science, 17*, 91-95. doi: 10.1111/j.1467-8721.2008.00555.x

Batchelder, W. H. & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97,* 548-564. doi: 10.1037//0033-295X.97.4.548

Bayen, U. J., Murnane,, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 197-215. doi: 10.1037//0278-7393.22.1.19

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91,* 276-292. doi: 10.1037/0033-2909.91.2.276

Beatty, J. & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science 5,* 371–372.

Bergstrom, K. J. & Hiscock, M. (1988). Factors influencing ocular motility during the performance of cognitive tasks. *Canadian Journal of Psychology, 42*, 1-23. doi: 10.1037/h0084174

Bink, M. L., Marsh, R. L., & Hicks, J. L. (1999). An alternative conceptualization to memory “strength” in reality monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 804-809. doi: 10.1037/0278-7393.25.3.804

Brown, R. & McNeill, D. (1966). Tip of the tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325-337. doi: 10.1016/S0022-5371(66)80040-3

Buchanan, T. W., Etzel J. A., Adolphs, R., & Tranel, D. (2006). The influence of autonomic arousal and semantic relatedness on memory for emotional words. *International Journal of Psychophysiology, 61*, 26-33. doi: 10.1016/j.ijpsycho.2005.10.022

Cansino, S. & Trejo-Morales, P. (2008). Neurophysiology of successful encoding and retrieval of source memory. *Cognitive, Affective, and Behavioral Neuroscience, 8,* 85-98. doi: 10.3758/CABN.8.1.85

Clark, K. B., Naritoku, D. K., Smith, D. C., Browning, R. A., & Jensen, R. A. (1999). Enhanced recognition memory following vagus nerve stimulation in human subjects. *Nature Neuroscience, 2*(1), 94-98. doi: 10.1038/4600

Cohen, G. & Faulkner, D. (1986). Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology, 4,* 187-197.

Cook, G. I., Marsh, R., L. & Hicks, J. L. (2006). Source memory in the absence of successful cued recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32*, 828-835. doi: 10.1037/0278-7393.32.4.828

Conway, M. A. & Dewhurst, S. A. (1995). Remembering, familiarity, and source monitoring. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology,* 48, 125-140.

Conway, M. A., Pleydell-Pearce, C. W., Whitecross, S., Sharpe, H. (2002). Brain imaging autobiographical memory. *Psychology of Learning and Motivation: Advances in Research and Theory, 41*, 22-263. doi: 10.1016/S0079-7421(02)80008-1

Conway, M. A., Dewhurst, S. A., Pearson, N., & Sapute, A. (2001). The self and recollection reconsidered: How a 'failure to replicate' failed and why trace strength accounts of recollection are untenable. *Applied Cognitive Psychology, 15*, 673-686. doi: 10.1002/acp.740

Croiset, G., Nijsen, M. J.M.A., & Kamphuis, P.J.G.H. (2000). Role of corticotropin-releasing factor, vasopressin and the autonomic nervous system in learning and memory. *European Journal of Pharmacology, 405*, 225-234.

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition*, 28, 923-938.

Curran, T. & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 531-547. doi: 10.1037/0278-7393.21.3.531

Curran, T. & Hintzman, D. L. (1997). Comparing retrieval dynamics in recognition memory and lexical decision. *Journal of Experimental Psychology: General, 126*, 228-247. doi: 10.1037/0096-3445.126.3.228.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317-367. doi: 10.1006/cogp.2001.0750

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories.

Proceedings of the National Academy of Sciences, 100, 2157-2162. doi: 10.1073/pnas.0337195100

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology, 47*, 292-303. doi: 10.1016/S0022-2496(03)00005-1

Dewhurst, S. A. & Hitch, G. J. (1999). Cognitive effort and recollective experience in recognition memory. *Memory, 7*, 126-146. doi: 10.1080/741944067

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2008). The effects of unitization on familiarity-based source memory: Testing a behavioral prediction derived from neuroimaging data. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34*, 730-740. doi: 10.1037/0278-7393.34.4.730

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2010). Medial temporal lobe activity during source retrieval reflects information type, not memory strength. *Journal of Cognitive Neuroscience, 8*, 1808-1818. doi: 10.1162/jocn.2009.21335

Dietrich, D. E., Emrich, H. M., Waller, C., Wieringa, B. M., Johannes, S. Münte, T. F. (2000). Emotion/cognition-coupling in word recognition memory of depressive patients: An event-related potential study. *Psychiatry Research, 96*, 15-29. doi: 10.1016/S0165-1781(00)00187-6

Dodson, C. S., Holland, P. W. & Shimamura, A. P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 24, 1121-1136. doi: 10.1037/0278-7393.24.5.1121

Dodson, C. S., Bawa, S., & Slotnick, S, D. (2007). Aging, source memory, and misrecollections. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33*, 169-181. doi: 10.1037/0278-7393.33.1.169

Dodson, C. S. & Johnson, M. K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology, 106*, 541-557. doi: 10.2307/1422968

Dodson, C. S. & Shimamura, A. P. (2000). Differential effects of cue dependency on item and source memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 1023-1044. doi: 10.1037//0278-7393.26.4.1023

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition,* 24, 4, 523-533.

Donaldson, W., MacKenzie, T. M., & Underhill, C. F. (1996). A comparison of recollective memory and source monitoring. *Psychonomic Bulletin & Review, 3*, 486-490.

Duarte, A., Ranganath, C., Winward, L., Hayward, D. & Knight, R. T. (2004). Dissociable neural correlates for familiarity and recollection during the encoding and retrieval of pictures. *Cognitive Brain Research,* 18, 255-272. doi: 10.1016/j .cogbrainres.2003 .10.010.

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, 111, 2, 524-542. doi: 10.1037/0033-295X .111.2.524.

Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review, 115,* 426-446. doi: 10.1037/0033-295X.115.2.426

Eberhard, K., Spivey-Knowlton, S., Sedivy, J. & Tanenhaus, M. (1995). Eye movements as a window into real-time spoken language processing in natural contexts. *Journal of Psycholinguistic Research*, 24, 409-436. doi: 10.1007/BF02143160

Edelman, J. A. & Goldberg, M. W. (2001). Dependence of saccade-related activity in the primary superior colliculus on visual target presence. *Journal of Neurophysiology, 86*, 676-691.

Ehrlichman, H. & Barrett, J. (1983). Random saccadic eye-movements during verbal-linguistic and visual-imaginal tasks. *Acta Psychologica, 53*, 9-26. doi: 10.1016/0001-6918(83)90013-6

Ehrlichman, H. & Micic, D. (2012). Why do people move their eyes when they think? *Current Directions in Psychological Science, 21*, 96-100. doi: 10.1177/0963721412436810

Ehrlichman, H., Micic, D., Sousa, A., & Zhu, J. (2007). Looking for answers: Eye movements in non-visual cognitive tasks. *Brain and Cognition, 64*, 7-20. doi: 10.1016/j.bandc.2006.10.001

Ekman, P., & Matsumoto, D. (1993). *Japanese and Caucasian neutral faces (JACNeuF*) [Photographs on CD–Rom]. San Francisco, CA.

Eldridge, L. L., Engel, S. A., Zeineh, M. M., Bookheimer, S. Y., & Knowlton, B. J. (2000). A dissociation of encoding and retrieval processes in the human hippocampus. *Journal of Neuroscience, 25*, 3280-3286. doi: 10.1523/JNEUROSCI.3420-04.2005

Elfman, K. W., Parks, C. M., & Yonelinas, A. P. (2008). Testing a neurocomputational model of recollection, familiarity, and source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 34*, 752-768. doi: 10.1037/0278-7393.34.4.752

Engelkamp, J. & Dehn, D. M. (1997). Strategy and consciousness in remembering subject-performed actions. *Sprache & Kognition, 16*, 94-109.

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition*, *22*, 34-39.

Eschenko, O. & Sara, S. J. (2008). Learning-dependent, transient increase of activity in noradrenergic neurons of locus coeruleus during slow wave sleep in the rat: Brain stem-cortex interplay for memory consolidation? *Cerebral Cortex, 18,* 2596-2603. doi: 10.1093/cercor/bhn020

Fendrich, D. W. (1998). Recognition benefits from the reinstatement of a spatial representation of motoric processing. *Psychological Research, 61*, 125-134.

Fendrich, D. W., Healy, A. F., & Bourne, L. E. (1991). Long-term repetition effects for motoric and perceptual procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 137-151. doi: 10.1037/0278-7393.17.1.137

Ferguson, S. A., Hashtroudi, S., & Johnson, M. K. (1992). Age-differences in using source-relevant cues. *Psychology and Aging, 7,* 443-452. doi: 10.1037//0882-7974.7.3.443

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*, 405-410. doi: 10.1016/j.tics.2008.07.007

Fontana, F. (1765). *Dei moti dell'iride*. Lucca: J. Giusti.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition,* 16, 309-313.

Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian*

Journal of Experimental Psychology, 50, 114-122. doi: 10.1037/1196-1961.50.1.114

Gardiner, J.M. & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition, 19,* 617-623. doi: 10.3758/BF03197157

Gardiner, J. M. & Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition, 7,* 285-288. doi: 10.1006/ccog.1997.0321

Gardiner, J. M. & Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory, 10,* 83-98. doi: 10.1080/0965210143000281

Gardiner, J. M. & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18,* 579-583.

Gianaros, P. J., Van Der Veen, F. M., & Jennings, J. R. (2004). Regional cerebral blood flow correlates with heart period and high-frequency heart period variability during working-memory tasks: Implications for the cortical and subcortical regulation of cardiac autonomic activity. *Psychophysiology,* 41, 4, 521-530 10.1111/1469-8986 .2004.00179.x

Gibson, J. J. (1950). The perception of visual surfaces. *American Journal of Psychology, 63,* 367-384. doi: 10.2307/1418003

Gibson, J. J. (1966). Problem of temporal order in stimulation and perception. *Journal of Psychology, 62,* 141-149.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience, 10,* 252-269. doi: 10.3758/CABN.10.2.252

Glanzer, M., Hilford, A., & Kim, K. (2004). Six regularities of source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1176-1195. doi: 10.1037/0278-7393.30.6.1176

Glanzer, M., Kim, K., Hilford, A. & Adams, J. K. (1999). Slope of the receiver operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 25, 500-513. Doi: 10.1037/0278-7393.25.2.500

Glenberg, A. (1997). What memory is for. *Behavioral Brain Sciences, 20,* 1-55.

Gnadt, J. W., Bracewell, R. M., & Andersen, R. A. (1991). Sensorimotor transformation during eye movements to remembered visual targets. *Vision Research, 31*, 693-715. doi: 10.1016/0042-6989(91)90010-3

Godden, D. & Baddeley, A. (1975). When does context influence recognition memory? *British Journal of Psychology, 71*, 99-104.

Godijn, R. & Theeuwes, J. (2012). Overt is no better than covert when rehearsing visuo-spatial information in working memory. *Memory & Cognition, 40*, 52-61. doi: 10.3758/s13421-011-0132-x

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain areas. *Journal of Neuroscience, 26*, 4917-4921. doi: 10.1523/JNEUROSCI.5389-05.2006

Goldinger, S. D., He, Y., & Papesh, M. H. (2009) Deficits in cross-race face learning: Insights from eye-movements and pupillometry. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*, 1105-1122. doi: 10.1037/a0016548

Goldinger, S. D. & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science, 21*, 90-95. doi: 10.1177/0963721412436811

Goldwater, B. C. (1972). Psychological significance of pupillary movements. *Psychological Bulletin, 77*, 340-355. doi: 10.1037/h0032456

Gordon, R. D. & Irwin, D. E. (1996). What is an object file? Evidence from priming studies. *Perception & Psychophysics, 58*, 1260-1277. doi: 10.3758/BF03207558

Guo, C., Duan, L., Li, W., & Paller, K. A. (2006). Distinguising source memory and item memory: Brain potentials at encoding and retrieval. *Brain Research, 1118*, 142-154. doi: 10.1016/j.brainres.2006.08.034.

Graf, P. & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 11*, 501-518. doi: 10.1037/0278-7393.11.3.501

Granholm, E., Asarnow, R. F., Sarkin, A. J. & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology,* 33, 457-461. doi: 10.1111/j.1469-8986.1996.tb01071.x

Granholm, E. & Steinhauer, S. R. (2004). Introduction: Pupillometric measures ofcognitive and emotional processing. *International Journal of Psychophysiology,* 52, 1-6. doi: 10.1016/j.ijpsycho .2003.12.001

Granholm, E. & Verney, S. P. (2004). Pupillary responses and attentional allocation problems on the backward masking task in schizophrenia. *International Journal of Psychophysiology*, 52, 37-52 10.1016/j.ijpsycho .2003.12.004

Gronlund, S. D. & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 846-858. doi: 10.1037/0278-7393.15.5.846

Gruppuso, V., Lindsay, D. S. & Kelley, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 23, 259-278. doi: 10.1037/0278-7393.23.2.259

Guttentag, R. E. & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory and Language, 37*, 502-516. doi: 10.1006/jmla.1997.2532

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth a glance: Using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience, 4,* 1-16. doi: 10.3389/fnhum.2010.00166

Hannula, D. E. Baym, C. L., Warren, D. E., & Cohen, N. J. (2012). The eyes know: Eye movements as a veridical index of memory. *Psychological Science, 23, 1-10*. doi: 10.1177/0956797611429799.

Hannula, D. E. & Ranganath, C. (2009). The eyes have it: Hippocampal activity predicts expression of memory in eye movements. *Neuron, 63*, 592-599. doi: 10.1016/j.neuron.2009.08.025

Hannula, D. E., Ryan, J. D., Tranel, D., & Cohen, N. J. (2007). Rapid onset relational memory effects are evidence in eye movements behavior, but not in hippocampal amnesia. *Journal of Cognitive Neuroscience, 19,* 1690-1705. doi: 10.1162/jocn.2007.19.10.1690

Hautus, M. J., Macmillan, N. A., & Rotello, C. M. (2008). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review, 15*, 889-905. doi: 10.3758/PBR.15.5.889

Heathcote, A., Raymond, F., & Dunn, J. (2006). Recollection and familiarity in recognition memory: Evidence from ROC curves. *Journal of Memory and Lamguage, 55*, 495-514. doi: 10.1016/j.jml.2006.07.001

Heaver, B. & Hutton, S. B. (2011). Keeping an eye on the truth? Pupil size changes associated with recognition memory. *Memory, 19,* 398-405. doi: 10.1080/09658211.2011.575788

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*, 280-285. doi: 10.1016/j.tics.2004.04.001

Henderson, J. M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General, 123,* 410-426. doi: 10.1037/0096-3445.123.4.410

Henderson, J. M., Weeks Jr., P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complexscene viewing. *Journal of Experimental Psychology: Human Perception & Performance,* 25, 210-228. doi: 10.1037//0096-1523 .25.1.210

Henderson, J. M., Williams, C. C., Castelhano, M. S., & Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception & Psychophysics, 65*, 725-734.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition, 33*, 98-106.

Hess, E. H. (1965). Attitude and pupil size. *Scientific American,* 212, 46-54.

Hess, E.H., & Polt, J.M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*, 1190–1192.

Hess, E.H., Seltzer, A.L., & Schlien, J.M. (1965). Pupil response of hetero- and homosexual males to pictures of men and women: A pilot study. *Journal of Abnormal Psychology, 70*, 165-168.

Hicks, J. L. & Marsh, R. L. (2002). On predicting the future states of awareness for recognition of unrecallable items. *Memory & Cognition, 30*, 60-66.

Hicks, J. L., Marsh, R. L. & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 503-508. doi: 10.1037//0278-7383.28.3.503

Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L. T. (2002). Regularities of source recognition: ROC analysis. *Journal of Experimental Psychology: General*, 131, 494-510 10.1037//0096-3445 .131.4.494

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review, 93*, 411-428. doi: : 10.1037//0033-295X.93.4.411

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528-551. doi: 10.1037/0033-295X.95.4.528

Hintzman, D. L. & Caulton, D. A. (1997). Recognition memory and modality judgments: A comparison of retrieval dynamics. *Journal of Memory and Language*, *37*, 1-23. doi: 10.1006/jmla.1997.2511

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition*, *26*, 449-462.

Hintzman, D. L. & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33*, 1-18. doi: 10.1006/jmla.1994.1001

Hirshman, E. (1998). On the logic of testing the independence assumption in the process-dissociation procedure. *Memory & Cognition, 26*, 857-859. doi: 10.3758/BF03201168

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-paradigm. *Memory & Cognition*, *25*, 345-351. doi: 10.3758/BF03211290

Holdstock, J. S., Mayes A. R,, Gong, Q. Y., Roberts, N., & Kapur, N. (2005). Item recognition is less impaired than recall and associative recognition in a patient with selective hippocampal damage. *Hippocampus*, *15*, 203-215. doi: 10.1002/hipo.20046

Hollingworth, A. & Henderson, J. M. (1998). Does consistent scene context facilitate object perception. *Journal of Experimental Psychology: General,* 127, 398-415. doi: 10.1037/0096-3445 .127.4.398

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review,* 8, 761-768.

Holm, L. & Mäntylä, T. (2007). Memory for scenes: Refixations reflect retrieval. *Memory & Cognition,* 35, 1664-1674.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences, 8,* 494-500. doi: 10.1016/j.tics.2004.08.007

Hoover, M. A. & Richardson, D. C. (2008). When facts go down the hole: Contrasting features and objecthood as indexes to memory. *Cognition, 108*, 533-542. doi: 10.1016/j.cognition.2008.02.011

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 9¸*21-38. doi: 10.1037//0278-7393.9.1.21

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. doi: 10.3758/BF03196776

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language,* 30, 513-541. doi: 10.1016/0749-596X(91)90025-F

Jacoby, L. L. (1998). Invariance in automatic influences of memory: Toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 3-26. doi: 10.1037/0278-7393.24.1.3

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25* 3-22. doi: 10.1037/0278-7393.25.1.3

Jacoby, L. L., Begg, I. M., & Toth, J. P. (1997). In defense of functional independence: Violations of assumptions underlying the process-dissociation procedure? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 484-495. doi: 10.1037/0278-7393.23.2.484

Jacoby, L. L. & Rhodes, M. G. (2006). False remembering in the aged. *Current Directions in Psychological Science, 15,* 49-53. doi: 10.1111/j.0963-7214.2006.00405.x

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler, (Eds.), Scientific approaches to consciousness. Carnegie Mellon Symposia on cognition. (pp. 13-47). Hillsdale, NJ: Erlbaum. Janisse, M. P. (1974). Pupil size, affect and exposure frequency. *Social Behavior and Personality,* 2, 125-146. doi: 10.2224/sbp.1974.2 .2.125

Johansson, M., Mecklinger, A. & Treese, A. (2004). Recognition memory for emotional and neutral faces: An event-related potential study. *Journal of Cognitive Neuroscience,* 16, 1840-1853. doi: 10.1162 /0898929042947883

Johansson, R., Holsanova, J., Dewhurts, R., & Holmqvist, K. (2011). Eye movemnts during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance, VOL,* 1-26. doi: 10.1037/a0026585.

Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and complete darkness. *Cognitive Science: A Multidisciplinary Journal, 30,* 1053-1079. doi: 10.1207/s15516709cog0000_86

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

Johnson, M. K. & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67-85. doi: 10.1037/0033-295X.88.1.67

Joordens, S. & Merikle, P. M. (1993). Independence or redundancy? Two models of conscious and unconscious influences. *Journal of Experimental Psychology: General, 122,* 462-467. doi: 10.1037/0096-3445.122.4.462

Kafkas, A. & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology, 64,* 1971-1989. doi:10.1080/17470218 .2011.588335

Kahneman, D. (1973). *Attention and effort*. New York: Prentice Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*, 1583-1585. doi:10.1126/science.154.3756.1583

Kahneman, D., Beatty, J., & Pollack. I. (1967). Perceptual deficit during a mental task. *Science, 157,* 218-219. doi: 10.1126/science.157.3785.218

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175-219. doi: 10.1016/0010-0285(92)90007-O

Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology, 41*, 175-185. doi: 10.1111/j.1469-8986.2004.00147.x

Kelley, R. & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 701-722. doi: 10.1037//0278-7393.27.3.701

Kensinger, E. A. & Schacter, D. L. (2006). Neural processes underlying memory attribution on a reality-monitoring task. *Cerebral Cortex, 16*, 1126-1133. doi: 0.1093/cercor/bhj054

Kent, C. & Lamberts, K. (2008). The encoding-retrieval relationship: Retrieval as mental simulation. *Trends in Cognitive Sciences, 12,* 92-98. doi: 10.1016/j.tics.2007.12.004

Kirwan, C. B., Wixted, J. T., & Squire, L. R. (2008). Activity in the medial temporal lobe predicts memory strength, whereas activity in the prefrontal cortex predicts recollection. *Journal of Neuroscience, 28*, 10541-10548. doi: 10.1523/JNEUROSCI.3456-08.2008

Klauer, K. C. & Kellen, D. (2010). Toward a complete decision model of item and source recognition: A discrete-state approach. *Psychonomic Bulletin & Review, 17,* 465-478. doi: 10.3758/PBR.17.4.465

Klimesch, W., Doppelmayr, M., Yonelinas, A., Kroll, N. E. A., Lazzara, M., Röhm, D., & Gruber, W. (2001). Theta synchronization during episodic retrieval: Neural correlates of conscious awareness. *Cognitive Brain Research, 12*, 33-38. doi: 10.1016/S0926-6410(01)00024-6

Knowlton, B. J. & Squire, L. R. (1995). Remembering and knowing: Two difference expression of declarative memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 699-710. doi: 10.1037//0278-7393.21.3.699

Kolers, P. A. & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior, 23*, 425-449. doi: 10.1016/S0022-5371(84)90282-2

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review,* 100, 609-639 10.1037/0033-295X .100.4.609

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*, 311-333. doi: 10.1037/0096-3445.124.3.311

Koriat, A., Levy-Sadot, R., Edry, E. & de Marcas, S. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 29, 1095-1105. doi: 10.1037/0278-7393.29.6.1095

Koriat, A. & Lieblich, I. (1974). What does a person in a TOT state know that a person in a don't know state doesn't know. *Memory & Cognition, 2,* 647-655. doi: 10.3758/BF03198134

Koss, M. (1986). Pupillary dilation as an index of central nervous system α2-adrenoceptor activation. *Journal of Pharmacology Methods, 15*, 1-19.

Kuçera, H. & Francis, W. N. (1967). *A computational analysis of present-day American English.* Providence, RI: Brown University Press.

Kuchinke, L., Võ, M. L-H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology, 65,* 132-140.

Kurilla, B. P. & Westerman, D. L. (2010). Source memory for unidentified stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 398-410. doi: 10.1037/a0018279

Laeng, B. Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science, 7,* 18-27. doi: 10.1177/1745691611427305

Laeng, B. & Teodorescu, D. S. (2002). Eye scanpaths during visual imagery re-enact those of perception of the same visual scene. *Cognitive Science*, 26, 207-231. doi: 10.1016/S0364-0213 (01)00065-9

Laeng, B., Waterloo, K., Johnsen, S. H., Bakke, S. J., Låg, T., Simonsen, S. S. & Høgsæt, J. (2007). The eyes remember it: oculography and pupillometry during recollection in three amnesic patients. *Journal of Cognitive Neuroscience, 19,* 1888-1904. doi: 10.1162/jocn.2007 .19.11.1888

Lindsay, D. S. & Johnson, M. K. (1991). Recognition memory and source monitoring. *Bulletin of the Psychonomic Society, 29*, 203-205.

Loewenfeld, I.E. (1958). Mechanisms of reflex dilation of the pupil: Historical review and experimental analysis. *Documenta Ophthalmologica, 12*, 185-448.

Lyle, K. B. & Jacobs, N. E. (2010). Is saccade-induced retrieval enhancement a potential means of improving eyewitness evidence? *Memory*, *18*, 581-594. doi: 10.1080/09658211.2010.493891

Lyle, K. B., Logan, J. M., & Roediger, H. L. (2008). Eye movements enhance memory for individuals who are strongly right-handed and harm it for individuals who are not. *Psychonomic Bulletin & Review, 15*, 515-520. doi: 10.3758/PBR.15.3.515

Lyle, K. B., Roediger, H. L., & McCabe, D. P. (2008). Handedness is related to memory via hemispheric interaction: Evidence from paired associate recall and source memory tasks. *Neuropsychology, 22,* 523-530. doi: 10.1037/0894-4105.22.4.523

Macmillan, N. A. & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 380-387. doi: 10.1037//0278-7393.28.2.380

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review,* 87, 252-271. doi: 10.1037/0033-295X .87.3.252

Mäntylä, T. (1997). Recollection of faces: Remembering differences and knowing similarities. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23, 1203-1216. doi: 10.1037/0278-7393 .23.5.1203

Mäntylä, T. & Holm, L. (2006). Gaze control and recollective experience in face recognition. *Visual Cognition,* 14, 365-386. doi: 10.1080 /13506280500347992

Maril, A., Simons, J. S., Weaver, J. J., & Schacter, D. L. (2005). Graded recall success: An event-related fMRI comparison of tip of the tongue afid feeling of knowing. *NeuroImage, 24*, 1130-1138. doi: 10.1016/j.neuroimage.2004.10.024

Marsh, R. L. & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 1137-1151. doi: 10.1037//0278-7393.24.5.1137

McElree, B., Dolan, P. O. & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 563-582. doi: 10.1037//0278-7393 .25.3.563

McKone, E. & French, B. (2001). In what sense is implicit memory "episodic"? The effect of reinstating environmental context. *Psychonomic Bulletin & Review, 8*, 806-811. doi: 10.3758/BF03196221

Meiser, T. & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28,* 116-137. doi: 10.1037/0278-7393.28.1.116

Melcher, D. & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision Research,* 39, 2929-2946. doi: 10.1016/S0042-6989 (99)00029-2

Micic, D., Ehrlichman, H., & Chen, R. (2010). Why do we move our eyes while trying to remember? The relationship between non-visual gaze patterns and memory. *Brain and Cognition*, *74,* 210-224. doi: 10.1016/j.bandc.2010.07.014

Mickes, L., Johnson, E. M., & Wixted, J. T. (2010). Continuous recollection versus unitized familiarity in associative recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 36*, 843-863. doi: 10.1037/a0019755

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implication for dual-process theories of recognition memory. *Psychological Science, 20*, 509-515. doi: 10.1111/j.1467-9280.2009.02324.x

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533. doi: 10.1016/S0022-5371(77)80016-9

Moscovitch, M. (2008). The hippocampus as a "stupid," domain-specific module: Implications for theories of recent and remote memory, and of imagination. *Canadian Journal of Experimental Psychology, 62*, 62-79. doi: 10.1037/1196-1961.62.1.62

Moscovitch, M. & McAndrews, M. P. (2002). Material-specific deficits in "remembering" in patients with unilateral temporal lobe epilepsy and excisions. *Neuropsychologia, 40*, 1335-1342. doi: PII S0028-3932(01)00213-5

Mulligan, N. W. (2011). Conceptual implicit memory and environmental context. *Consciousness and Cognition, 20*, 737-744. doi: 10.1016/j.concog.2010.11.008

Mulligan, N. W. & Hirshman, E. (1997). Measure the bases of recognition memory: An investigation of the process-dissociation framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 280-304. doi: 10.1037/0278-7393.23.2.280

Naveh-Benjamin, M. (1988). Recognition memory of spatial location information: Another failure to support automaticity. *Memory & Cognition, 16*, 437-445.

Nelson, D. L., Brooks, D. H., & Wheeler, J. W. (1975). Sensory and meaning features in stimulus recognition and associative retrieval. *Journal of Experimental Psychology: Human Learning and Memory, 1*, 711-719. doi: 10.1037//0278-7393.1.6.711

Norman, K. A. & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning systems approach. *Psychological Review, 110*, 611-646. doi: 10.1037/0033-295X.110.4.611

Noton, D. & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science, 171*, 308-311. doi: 10.1126/science.171.3968.308

Onyper, S. V., Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General, 139*, 341-364. doi: 10.1037/a0018926

O'Regan, J. K. (1992). Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461-488.

Otero, S.C., Weekes, B.S., & Hutton, S.B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48,* 1346-1353. doi: 10.1111/j.1469-8986.2011.01217.x

Otten, L. J. (2007). Fragments of a larger whole: Retrieval cues constrain observed neural correlates of memory encoding. *Cerebral Cortex, 17,* 2030-2038. doi: 10.1093/cercor/bhl111

Papesh, M. H. (in prep). Pupil diameter reveals unconscious "confidence": Empirical ROCs derived from pupillometry.

Papesh, M. H., & Goldinger, S. D. (2011). Your effort is showing! Pupil dilation reveals memory heuristics. In P. Higham & J. Leboe (Eds.), *Constructions of Remembering and Metacognition* (pp. 215-224). Palgrave Macmillan.

Papesh, M. H. & Goldinger, S. D. (2012) Pupil-blah-metry: Cognitive effort in speech planning reflected by pupil dilation. *Attention, Perception, & Psychophysics, 74,* 754-765. doi: 10.3758/s13414-011-0263-y

Papesh, M. H., Goldinger, S.D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, *83*, 56-64. doi:10.1016/j.ijpsycho.2011.10.002

Parker, R. E. (1978). Picture processing during recognition. *Journal of Experimental Psychology: Human Perception & Performance,* 4, 284-293. doi: 10.1037//0096-1523 .4.2.284

Parker, A., Dagnall, N., & Coyle, A. (2007). Environmental context effects in conceptual explicit and implicit memory. *Memory, 15,* 423-434. doi: 10.1080/09658210701309834

Parker, A., Gellatly, A., & Waterman, M. (1999). The effect of environmental context manipulation on memory: Dissociation between perceptual and conceptual implicit tests. *European Journal of Cognitive Psychology, 11,* 555-570. doi: 10.1080/095414499382291

Parks, C. M., Murray, L. J., Yonelinas, A. P., Elfman, K., & Yonelinas, A. P. (2011). Variations in recollection: The effects of complexity on source recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *37*, 861-873. doi: 10.1037/a0022798

Parks, C. M., Murray, L. J., Elfman, K., & Yonelinas, A. P. (2008). Variations in recollection: The effects of complexity on source recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *37*, 861-873. doi: 10.1037/a0022798

Parks, C. M. & Yonelinas, A. P. (2007). Moving beyond pure signal detection models: A reply to Wixted (2007). *Psychological Review*, 114, 188-202. doi: 10.1037/0033-295X .114.1.188

Partala, T. & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59*, 185-198. doi: 10.1016/S1071-5819(03)00017-X

Perfect, T. J., Mayes, A. R., Downes, J. J. & Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 49, 3, 797-813. doi: 10.1080/027249896392603

Pezdek, K., Roman, Z., & Sobolik, K. G. (1986). Spatial memory for objects and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 530-537. doi: 10.1037/0278-7393.12.4.530

Pollatsek, A., Rayner, K., & Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 199-210. doi: 10.1037/0096-1523.16.1.199

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly Journal of Experimental Psychology, 60*, 211-229.

Porter, G. & Troscianko, T. (2003). Pupillary response to grating stimuli. *Perception, 32*, 156.

Psychology Software Tools. (2006). E-Prime (Version 1.2). http://www.pstnet.com

Quamme, J. R., Frederick, C., Kroll, N. E. A., Yonelinas, A. P., & Dobbins, I. G. (2002). Recognition memory for source and occurrence: The importance of recollection. *Memory & Cognition, 30*, 893-907. doi: 10.3758/BF03195775

Quamme, J. R., Yonelinas, A. P., & Norman, K. A. (2007). Effect of unitization on associative recognition in amnesia. *Hippocampus, 17*, 192-200. doi: 10.1002/hipo.20257

Qin, J., Raye, C. L., Johnson, M. K. & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of*

Experimental Psychology: Learning, Memory, & Cognition, 27, 1110-1115. doi: 10.1037/0278-7393.27.4.1110

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89-102.

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 22, 365-377. doi: 10.1037/0278-7393 .22.2.365

Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. *Society of Neuroscience Abstracts*, *19*, 974.

Rajkowski, J., Majczynski, H., Clayton, E., & Aston-Jones, G. (2004). Activation of monkey locus coeruleus neurons varies with difficulty and performance in a target detection task*. Journal of Neurophysiology, 92*, 361–371.

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M. & D'Esposito, M. D. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia,* 42, 2-13. doi: 10.1016/j .neuropsychologia .2003.07.006

Ratcliff, R., Sheu, C., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99,* 518-535. doi: 10.1037/0033-295X.99.3.518

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin,* 124, 372-422. doi: 10.1037/0033-2909 .124.3.372

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16,* 1-21. doi: 10.3758/PBR.16.1.1

Richardson, D. C., Altmann, G. T. M., Spivey, M. J.. & Hoover, M. A. (2009). Much ado about eye movements to nothing: A response to Ferreira et al.: Taking a new look at looking at nothing. *Trends in Cognitive Sciences, 13,* 235-236. doi: 10.1016/j.tics.2009.02.006

Richardson, D. C., Dale, R., & Spivey, M. J. (2007). Eye movements in language and cognition: A brief introduction. In M. Gonzalez- Marquez, I.

Mittelberg, S. Coulson, & M. J. Spivey (Eds.), *Methods in cognitive linguistics* (pp. 323–344). Amsterdam, Netherlands: John Benjamins.

Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science, 33,* 1468-1482. doi: 10.1111/j.1551-6709.2009.01057.x

Richardson, D. C. & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General, 133,* 46-62. doi: 10.1037/0096-3445.133.1.46

Richardson, D. C. & Spivey, M. J. (2000). Representation, space, and Hollywood Squares: Looking at things that aren't there anymore. *Cognition, 76,* 269-295. doi: 10.1016/S0010-0277(00)00084-6

Richardson-Klavehn, A., Gardiner, J. M., & Java, R. I. (1996). Memory: Task dissociations, process dissociations and dissociationsof consciousness. In G. Underwood (Ed.), Implicit cognition (pp.85-158). Oxford, England: Oxford University Press.

Ross, J. & Ma-Wyatt, A. (2003). Opposite effects of saccades on short-term and long-term visual memory. *Perception, 32,* 6-6.

Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review, 12,* 865-873. doi: 10.3758/BF03196778

Rotello, C. M. & Zeng, M. (2008). Analysis of RT distributions in the remember-know paradigm. *Psychonomic Bulletin & Review, 15,* 825-832. doi: 10.3758/PBR.15.4.825

Rubin, D. C. (1975). Within word structure in tip-of-the-tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior, 14,* 392-397. doi: 10.1016/S0022-5371(75)80018-1

Rubin, D. C. (2006). The basic-systems model of episodic memory. *Perspectives on Psychological Science, 1,* 277-311. doi: 10.1111/j.1745-6916.2006.00017.

Rugg, M. D., Schloerscheidt, A. M. & Mark, R. E. (1998). An electrophysiological comparison of two indices of recollection. *Journal of Memory & Language,* 39, 47-69. doi: 10.1006/jmla.1997.2555

Runquist, W. N. (1978). Physical similarity and stimulus recognition: Interference could still be associative. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 32-36. doi: 10.1037//0278-7393.4.1.32

Ryan, J. D. & Cohen, N. J. (2004a). Processing and short-term retention of relational information in amnesia. *Neuropsychologia,* 42, 497-511 10.1016/j .neuropsychologia .2003.08.011

Ryan, J. D. & Cohen, N. J. (2004b). The nature of change detection and on-line representations of scenes. *Journal of Experimental Psychology: Human Perception and Performance,* 30, 5, 988-1015 10.1037/0096-1523 .30.5.988

Ryan, J. D., Hannula, D. E., & Cohen, N. J. (2007). The obligatory effects of memory on eye movements. *Memory, 15,* 508-525. doi: 10.1080/09658210701391022

Sauvage, M. M., Fortin, N. J., Owens, C. B., Yonelinas, A. P., & Eichenbaum, H. (2008). Recognition memory: Opposite effects of hippocampal damage on recollection and familiarity. *Nature Neuroscience, 11*, 16-18. doi: 10.1038/nn2016

Schacter, D. L. & Tulving, E. (1994). What are the memory systems of 1994? *Memory Systems 1994*, 1-38.

Schmidt, S. R. (1985). Encoding and retrieval processes in the memory for conceptual distinctive events. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 11,* 565-578. doi: 10.1037/0278-7393.11.3.565

Schutz, J. & Bröder, A. (2011). Signal detection and threshold models of source memory. *Experimental Psychology, 58*, 293-311. doi: 10.1027/1618-3169/a000097

Schwartz, (2002). The strategic control of retrieval during tip-of-the-tongue states. *Korean Journal of Thinking and Problem-solving, 12*, 27-38.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition,* 71, 109-147. doi: 10.1016/S0010-0277 (99)00025-6

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia. 45,* 2802-2810. doi: 10.1016/j.neuropsychologia.2007.05.002

Slotnick, S. D. & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition, 33*, 151-170.

Slotnick, S. D., Klein, S. A., Dodson, C. S., & Shimamura, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 1499-1517. doi: 10.1037//0278-7393.26.6.1499

Smith, S. M., Heath, F. R., & Vela, E. (1990). Environmental context-dependent homophone spelling. *American Journal of Psychology, 103,* 229-242. doi: 10.2307/1423144

Smith, C. N. & Squire, L. R. (2008). Experience-dependent eye movements reflect hippocampus-dependent (aware) memory. *The Journal of Neuroscience, 28,* 12825-12833. doi: 10.1523/jneurosci.4542-08.2008

Spivey, M. J. & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research,* 65, 235-241. doi: 10.1007 /s004260100059

Spivey, M. J. & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10,* 281-284. doi: 10.1111/1467-9280.00151

Stanners, R. F., Coulter, M., Sweet, A. W. & Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion*, 3, 319-340. doi: 10.1007/BF00994048

Staresina, B. P. & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *Journal of Neuroscience, 26*, 9162-9172. doi: 10.1523/JNEUROSCI.2877-06.2006

Starns, J. J., Hicks, J. L., Brown, N. L., & Martin, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition, 36,* 1-8. doi: 10.3758/MC.36.1.1

Steinhauer, S.R., & Hakerem, G. (1992). The pupillary response in cognitive psychophysiology and schizophrenia. *Annals of the NY Academy of Sciences, 658*, 182–204.

Steinhauer, S. R., Boller, F., Zubin, J. & Pearlman, S. (1983). Pupillary dilation to emotional visual stimuli revisited. *Psychophysiology,* 20.

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervations of pupillary dilation during sustained processing. *International Journal of Psychophysiology, 52*, 77-86. doi: 10.1016/j.ijpsycho.2003.12.005

Sterpenich, V., D'Argembeau, A., Desseilles, M., Balteau, E., Albouy, G., Vandewalle, G., Degueldre, C., Luxen, A., Collette, F., & Maquet, P. (2006). The locus ceruleus is involved in the successful retrieval of emotional memories in humans. *Journal of Neuroscience, 26*, 7416-7423. doi: 10.1523/JNEUROSCI.1001-06.2006

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science,* 268, 5217, 1632-1634. doi: 10.1126/science.7777863

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception, 1*, 7-27. doi: 10.1068/i0382

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 2,* 339-363. doi: 10.3758/BF03210972

Tremblay, S., Saint-Aubin, J., & Jalbert, A. (2006). Rehearsal in serial memory for visual-spatial information: Evidence from eye movements. *Psychonomic Bulletin & Review, 13*, 452-457. doi: 10.3758/BF03193869

Tulving, E. (1983). *Elements of episodic memory.* London: Oxford University Press.

Turvey, M. T. (1977). Contrasting orientations to theory of visual information processing. *Psychological Review, 84,* 67-88. doi: 10.1037//0033-295X.84.1.67

Uncaper, M. R., Otten, L. J., & Rugg, M. D. (2006). Episodic encoding is more than the sum of its parts: An fMRI investigation of multifeatural contextual encoding. *Neuron, 52*, 547-556. doi: 10.1016/j.neuron.2006.08.011

Uncapher, M. R. & Rugg, M. D. (2005). Effects of divided attention on fMRI correlates of memory encoding. *Journal of Cognitive Neuroscience, 17*, 1923-1935. doi: 10.1162/089892905775008616

van den Berg, D. P. G. & van der Gaag, M. (2012). Treating trauma in psychosis with EMDR: A pilot stuy. *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 664-671. doi: 10.1016/j.jbtep.2011.09.011

Van Gerven, P. W. M., Paas, F., Van Merriënboer, J.J.G., Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*, 167-174. doi: 10.1111/j.1469-8986.2003.00148.x

Verfaellie, M., Rajaram, S., Fossum, K., & Williams, L. (2008). Not all repetition is alike: Different benefits of repetition in amnesia and normal memory. *Journal of the International Neuropsychological Society, 14*, 365-372. doi: 10.1017/S1355617708080612

Võ, M. L-H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology, 45*, 130-140. doi: 10.1111/j.1469-8986.2008.00745.x

Wais, P. E. (2008). fMRI signals associated with memory strength in the medial temporal lobes: A meta-analysis. *Neuropsychologia, 46*, 3185-3196. doi: 10.1016/j.neuropsychologia.2008.08.025

Wais, P. E., Wixted, J.T., Hopkins, R.O., Squire, L.R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron, 49*, 459−466. doi: 10.1016/j.neuron.2005.12.020

Wais, P.E., Squire, L.R., & Wixted, J.T. (2009). In search of recollection and familiarity signals in the hippocampus. *Journal of Cognitive Neuroscience, 22*, 109-123. doi: 10.1162/jocn.2009.21190

Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature, 217*, 972-974. doi: 10.1038/217972a0

Winograd, E. & Church, V. E. (1988). Role of spatial location in learning face-name associations. *Memory & Cognition, 16*, 1-7.

Wickens, T. D. (2002). *Elementary signal detection theory.* New York: Oxford University Press.

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

Wixted, J. T. (2009). Remember/know judgments in cognitive neuroscience: An illustration of the underrepresented point of view. *Learning & Memory, 16,* 406-412. doi: 10.1101/lm.1312809

Wixted, J. T. & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117,* 1025-1054. doi: 10.1037/a0030874

Wixted, J. T. & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11,* 616-641.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidencefor a dual-process model. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 20, 1341-1354. doi: 10.1037//0278-7393 .20.6.1341

Yonelinas, A. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an ROC analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 1415-1434. doi: 10.1037/0278-7393.25.6.1415

Yonelinas, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. Philosophical Transactions of the Royal Society of London. *Series B, Biological Sciences,* 356, 1363-1374.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language,* 46, 441-517. doi: 10.1006/jmla.2002.2864

Yonelinas, A. P. & Jacoby, L. L. (1996). Response bias and the process-dissociation procedure. *Journal of Experimental Psychology: General, 125,* 422-434. doi: 10.1037/0096-3445.125.4.422

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G., & Soltani, M. (1999). Recognition memory for faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin & Review, 6,* 654-661. doi: 10.3758/BF03212975

Yonelinas, A. P., Kroll, N. E. A., Quamme, J. R., Lazzara, M. M., Sauve, M. J., Widaman, K. F. & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience, 5*, 1236-1241. doi: 10.1038/nn961

Yonelinas, A. P. & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin,* 133, 800-832. doi: 10.1037/0033-2909 .133.5.800

Yonelinas, A. P., Regehr, G., & Jacoby, L. L. (1995). Incorporating response bias in a dual-process theory of memory. *Journal of Memory and Language, 34*, 821-835.

Zimmer, H. D. (1998). Spatial information with pictures and words in visual short-term memory. *Psychological Research*, *61*, 277-284. doi: 10.1007/s004260050032

APPENDIX A

STIMULI

Low-Frequency

acid
apple
arm
beak
bear
belt
bib
boar
boot
branch
brood
bug
cane
cat
cord
dangle
deaf
dime
dish
dud
egg
fad
fame
fence
folder
fox
glean
hag
harp
haste
hat
haunt
hive
hoop
hose
leaf
lick
locker
minor
mute
pan
panda
pat
peas
pot
raid
rake
rose
rye
saddle
scale
scope
sew
shape
shark
sheep
sink
skill
skirt
skunk
slot
spider
sprang
stew
toffee
wink
witch
wood

High-Frequency

act
also
bad
bed
beer
big
blood
boat
book
boss
box
boy
card
care
church
coffee
day
doctor
dog
door
evil
eye
face
fat
father
feel
fish
force
found
gun
hair
hand
handle
hate
head
heart
help
hope
house
key
leave
like
made
man
middle
mind
nose
paper
pay
phone
plane
ring
safe
saw
show
simple
sleep
start
stay
stick
still
strong
table
truck
watch
week
wife
woman

APPENDIX B

FREQUENCY TABLES, EXPERIMENT 1

Low-Similarity Condition 1

True	Response				
Source	"Art"	"Chloe"	"Erik"	"Whitney"	"New"
Art	**109**	70	75	64	222
Chloe	39	**109**	91	81	220
Erik	49	65	**137**	64	225
Whitney	59	65	76	**128**	212
New	61	67	89	75	**588**

Low-Similarity Condition 2

True	Response				
Source	"Anne"	"Don"	"Emma"	"Steve"	"New"
Anne	**100**	44	36	45	155
Don	49	**76**	36	41	178
Emma	41	57	**74**	34	174
Steve	34	60	38	**75**	173
New	54	63	54	89	**351**

High-Similarity Condition 1

True	Response				
Source	"Art"	"Don"	"Erik"	"Steve"	"New"
Art	**83**	66	81	64	226
Don	70	**92**	76	66	216
Erik	59	69	**106**	52	234
Steve	58	71	95	**75**	221
New	70	74	115	64	**437**

High-Similarity Condition 2

True	Response				
Source	"Anne"	"Chloe"	"Emma"	"Whitney"	"New"
Anne	**79**	49	57	70	225
Chloe	59	**61**	53	64	243
Emma	83	50	**73**	52	222
Whitney	47	55	64	**77**	237
New	63	57	72	80	**448**

APPENDIX C

FREQUENCY TABLE PERCENTAGES, EXPERIMENT 1

Low-Similarity Condition 1

True Source	**Response**				
	"Art"	"Chloe"	"Erik"	"Whitney"	"New"
Art	**20.18**	12.96	13.88	11.85	40.79
Chloe	7.22	**20.19**	16.85	15.00	57.89
Erik	9.07	12.04	**25.37**	11.85	59.21
Whitney	10.93	12.04	14.07	**23.70**	55.79
New	6.93	7.61	10.11	8.50	**66.82**

Low-Similarity Condition 2

True Source	**Response**				
	"Anne"	"Don"	"Emma"	"Steve"	"New"
Anne	**26.32**	11.58	9.47	11.84	40.79
Don	12.89	**20.00**	9.47	10.79	46.84
Emma	10.79	15.00	**19.47**	8.95	45.79
Steve	8.95	15.79	10.00	**19.74**	45.53
New	8.83	10.31	8.83	14.56	**57.44**

High-Similarity Condition 1

True Source	**Response**				
	"Art"	"Don"	"Erik"	"Steve"	"New"
Art	**15.96**	12.69	15.58	12.31	43.46
Don	13.46	**17.69**	14.62	12.69	41.54
Erik	11.35	13.27	**20.38**	10.00	45.00
Steve	11.35	13.65	18.27	**14.42**	42.50
New	9.21	9.74	15.13	8.42	**57.50**

High-Similarity Condition 2

True Source	**Response**				
	"Anne"	"Chloe"	"Emma"	"Whitney"	"New"
Anne	**15.19**	9.42	10.96	13.46	43.27
Chloe	11.35	**11.73**	10.19	12.31	46.73
Emma	15.96	9.62	**14.04**	10.00	42.69
Whitney	9.04	10.58	12.31	**14.81**	42.69
New	8.75	7.92	10.00	11.11	**62.22**

APPENDIX D

SSI AND PSI SCORE FORMULAS

Source-specific identification score (item correct):

$$\frac{(P(\text{“A”}|A) + P(\text{“D”}|D) + P(\text{“C”}|C) + P(\text{“W”}|W))}{(((1 - P(\text{“N”}|A)) + ((1 - P(\text{“N”}|D)) + ((1 - P(\text{“N”}|C)) + ((1 - P(\text{“N”}|W)))}$$

Partial-source identification score (item correct):

$$\frac{(P(\text{“D”}|A) + P(\text{“A”}|D) + P(\text{“W”}|C) + P(\text{“C”}|W)) - (.5 * (P(\text{“C”}|A) + P(\text{“W”}|A) + P(\text{“C”}|D) + P(\text{“W”}|D) + P(\text{“A”}|C) + P(\text{“D”}|C) + P(\text{“A”}|W) + P(\text{“D”}|W)))}{((1 - P(\text{“A”}|A) - P(\text{“N”}|A)) + ((1 - P(\text{“D”}|D) - P(\text{“N”}|D)) + ((1 - P(\text{“C”}|C) - P(\text{“N”}|C)) + ((1 - P(\text{“W”}|W) - P(\text{“N”}|W))}$$

Source-specific identification score (item incorrect):

$$\frac{(P(\text{“A”}|A) + P(\text{“D”}|D) + P(\text{“C”}|C) + P(\text{“W”}|W))}{(P(\text{“D”}|A) + P(\text{“A”}|D) + P(\text{“W”}|C) + P(\text{“C”}|W)) + P(\text{“C”}|A) + P(\text{“W”}|A) + P(\text{“C”}|D) + P(\text{“W”}|D) + P(\text{“A”}|C) + P(\text{“D”}|C) + P(\text{“A”}|W) + P(\text{“D”}|W)))}$$

Partial-source identification score (item incorrect):

$$\frac{(P(\text{“D”}|A) + P(\text{“A”}|D) + P(\text{“W”}|C) + P(\text{“C”}|W)) - (.5 * (P(\text{“C”}|A) + P(\text{“W”}|A) + P(\text{“C”}|D) + P(\text{“W”}|D) + P(\text{“A”}|C) + P(\text{“D”}|C) + P(\text{“A”}|W) + P(\text{“D”}|W)))}{((1 - P(\text{“A”}|A)) + ((1 - P(\text{“D”}|D)) + ((1 - P(\text{“C”}|C)) + ((1 - P(\text{“W”}|W))}$$

APPENDIX E

ROC DATA, EXPERIMENT 1

Response Frequencies: HS Groups

	Very sure new				*Very sure old*		
	1	**2**	**3**	**4**	**5**	**6**	**∑**
Old	175	285	488	1175	430	487	3040
New	242	209	583	229	149	108	1520

Response Frequencies: LS Groups

	Very sure new				*Very sure old*		
	1	**2**	**3**	**4**	**5**	**6**	**∑**
Old	108	140	261	635	273	263	1680
New	108	113	343	152	76	48	840

Cumulative Response Proportions: HS Groups

	Very sure new				*Very sure old*	
	1	**2**	**3**	**4**	**5**	**6**
Old	1.00	0.94	0.85	0.69	0.30	0.16
New	1.00	0.84	0.70	0.32	0.17	0.07

Cumulative Response Proportions: LS Groups

	Very sure new				*Very sure old*	
	1	**2**	**3**	**4**	**5**	**6**
Old	1.00	0.94	0.85	0.70	0.32	0.16
New	1.00	0.87	0.74	0.33	0.15	0.06

z-Scores: HS Groups

	Very sure new				*Very sure old*	
	1	**2**	**3**	**4**	**5**	**6**
Old	3.09	1.55	0.99	0.47	-0.52	-0.99
New	3.09	0.99	0.52	-0.50	-0.99	-1.47

z-Scores: LS Groups

	Very sure new				*Very sure old*	
	1	**2**	**3**	**4**	**5**	**6**
Old	3.09	1.47	1.04	0.49	-0.49	-1.04
New	3.09	1.13	0.61	-0.47	-1.08	-1.65

APPENDIX F

FREQUENCY TABLES, EXPERIMENT 2

Low-Similarity Condition 1

True	**Response**				
Source	"Art"	"Chloe"	"Erik"	"Whitney"	"New"
Art	**34**	33	36	28	109
Chloe	27	**23**	34	42	114
Erik	24	23	**48**	40	105
Whitney	39	23	48	**40**	105
New	50	31	49	62	**288**

Low-Similarity Condition 2

True	**Response**				
Source	"Anne"	"Don"	"Emma"	"Steve"	"New"
Anne	**70**	34	32	31	93
Don	28	**67**	31	27	107
Emma	38	22	**58**	30	112
Steve	36	27	40	**49**	108
New	40	52	45	43	**340**

High-Similarity Condition 1

True	**Response**				
Source	"Art"	"Don"	"Erik"	"Steve"	"New"
Art	**50**	24	36	29	121
Don	34	**45**	43	22	116
Erik	38	34	**45**	28	115
Steve	20	35	49	**43**	113
New	40	65	74	41	**300**

High-Similarity Condition 2

True	**Response**				
Source	"Anne"	"Chloe"	"Emma"	"Whitney"	"New"
Anne	**34**	14	19	19	114
Chloe	27	**31**	17	25	100
Emma	28	17	**31**	25	99
Whitney	28	13	27	**25**	107
New	41	27	41	34	**257**

APPENDIX G

FREQUENCY TABLE PERCENTAGES, EXPERIMENT 2

Low-Similarity Condition 1

True Source	**Response**				
	“Art”	“Chloe”	“Erik”	“Whitney”	“New”
Art	**14.16**	13.75	15.00	11.66	41.92
Chloe	11.25	**9.58**	14.16	17.50	43.84
Erik	10.00	9.58	**20.00**	16.66	40.38
Whitney	16.25	9.58	12.08	**23.75**	35.38
New	10.42	6.46	10.21	12.92	**60.00**

Low-Similarity Condition 2

True Source	**Response**				
	“Anne”	“Don”	“Emma”	“Steve”	“New”
Anne	**26.92**	13.07	12.31	11.92	35.77
Don	10.77	**25.77**	11.92	10.38	41.15
Emma	14.62	8.46	**22.31**	11.54	43.07
Steve	13.85	10.38	15.38	**18.85**	41.54
New	7.69	10.00	8.65	8.27	**65.38**

High-Similarity Condition 1

True Source	**Response**				
	“Art”	“Don”	“Erik”	“Steve”	“New”
Art	**19.23**	9.23	13.85	11.15	60.50
Don	13.08	**17.31**	16.54	8.46	58.00
Erik	14.62	13.07	**17.31**	10.77	57.50
Steve	7.69	13.46	18.85	**16.54**	56.50
New	7.69	12.50	14.23	7.88	**57.69**

High-Similarity Condition 2

True Source	**Response**				
	“Anne”	“Chloe”	“Emma”	“Whitney”	“New”
Anne	**17.00**	7.00	9.50	9.50	57.00
Chloe	13.50	**15.50**	8.50	12.50	50.00
Emma	14.00	8.50	**15.50**	12.50	49.50
Whitney	14.00	6.50	13.50	**12.50**	53.50
New	10.25	6.75	10.25	8.50	**64.25**

APPENDIX H

ROC DATA, EXPERIMENT 2

Response Frequencies: HS Groups

	Very sure new				Very sure old		
	1	**2**	**3**	**4**	**5**	**6**	**∑**
Old	81	165	264	596	222	272	1600
New	121	106	292	146	86	49	800

Response Frequencies: LS Groups

	Very sure new				Very sure old		
	1	**2**	**3**	**4**	**5**	**6**	**∑**
Old	84	178	282	639	252	245	1680
New	137	116	316	133	81	57	840

Cumulative Response Proportions: HS Groups

	Very sure new				Very sure old	
	1	**2**	**3**	**4**	**5**	**6**
Old	1.00	0.94	0.85	0.68	0.31	0.17
New	1.00	0.85	0.72	0.35	0.17	0.06

Cumulative Response Proportions: LS Groups

	Very sure new				Very sure old	
	1	**2**	**3**	**4**	**5**	**6**
Old	1.00	0.95	0.84	0.68	0.30	0.15
New	1.00	0.84	0.70	0.32	0.16	0.07

z-Scores: HS Groups

	Very sure new				Very sure old	
	1	**2**	**3**	**4**	**5**	**6**
Old	3.09	1.55	0.99	0.47	-0.52	-0.95
New	3.09	0.99	0.55	-0.39	-0.99	-1.55

z-Scores: LS Groups

	Very sure new				Very sure old	
	1	**2**	**3**	**4**	**5**	**6**
Old	3.09	1.65	0.99	0.44	-0.55	-1.08
New	3.09	0.95	0.50	-0.47	-0.99	-1.55

APPENDIX I

FREQUENCY TABLES, EXPERIMENT 3

Low-Similarity Condition 1

True	Response				
Source	"Art"	"Chloe"	"Erik"	"Whitney"	"New"
Art	**53**	39	47	40	121
Chloe	38	**53**	42	39	128
Erik	42	26	**57**	40	135
Whitney	31	23	50	**63**	133
New	43	37	54	71	**395**

Low-Similarity Condition 2

True	Response				
Source	"Anne"	"Don"	"Emma"	"Steve"	"New"
Anne	**65**	44	42	31	158
Don	44	**87**	32	44	133
Emma	39	53	**58**	36	154
Steve	40	51	42	**58**	149
New	72	69	66	54	**419**

High-Similarity Condition 1

True	Response				
Source	"Art"	"Don"	"Erik"	"Steve"	"New"
Art	**26**	41	44	26	103
Don	26	**41**	27	26	120
Erik	31	32	**44**	19	114
Steve	26	41	34	**26**	113
New	49	63	65	37	**266**

High-Similarity Condition 2

True	Response				
Source	"Anne"	"Chloe"	"Emma"	"Whitney"	"New"
Anne	**48**	23	31	29	129
Chloe	40	**32**	33	37	118
Emma	34	17	**45**	36	128
Whitney	37	24	27	**53**	119
New	75	44	50	61	**290**

APPENDIX J

FREQUENCY TABLE PERCENTAGES, EXPERIMENT 3

Low-Similarity Condition 1

True	Response				
Source	"Art"	"Chloe"	"Erik"	"Whitney"	"New"
Art	**17.66**	13.00	15.66	13.33	40.33
Chloe	12.66	**17.66**	14.00	13.00	42.66
Erik	14.00	8.66	**19.00**	13.33	45.00
Whitney	10.33	7.66	16.66	**21.00**	44.33
New	7.16	6.16	9.00	11.83	**65.83**

Low-Similarity Condition 2

True	Response				
Source	"Anne"	"Don"	"Emma"	"Steve"	"New"
Anne	**19.12**	12.94	12.35	9.12	46.47
Don	12.94	**25.58**	9.41	12.94	39.12
Emma	11.47	15.58	**17.06**	10.59	45.29
Steve	11.76	15.00	12.35	**17.06**	43.82
New	10.58	10.14	9.70	7.94	**57.08**

High-Similarity Condition 1

True	Response				
Source	"Art"	"Don"	"Erik"	"Steve"	"New"
Art	**10.83**	17.08	18.83	10.83	42.92
Don	10.83	**17.08**	11.25	10.83	50.00
Erik	12.92	13.33	**18.33**	7.92	47.50
Steve	10.83	17.08	14.12	**10.83**	47.08
New	10.18	13.12	13.54	7.71	**55.41**

High-Similarity Condition 2

True	Response				
Source	"Anne"	"Chloe"	"Emma"	"Whitney"	"New"
Anne	**18.46**	8.85	11.92	11.15	49.62
Chloe	15.38	**12.31**	12.69	14.23	45.38
Emma	13.08	6.54	**17.31**	13.85	49.23
Whitney	14.23	9.23	10.38	**20.38**	45.77
New	14.42	8.46	9.62	11.73	**55.77**

APPENDIX K

ROC DATA, EXPERIMENT 3

Response Frequencies: HS Groups

	Very sure new				*Very sure old*		
	1	**2**	**3**	**4**	**5**	**6**	**∑**
Old	114	203	343	727	277	336	2000
New	149	155	394	148	95	59	1000

Response Frequencies: LS Groups

	Very sure new				*Very sure old*		
	1	**2**	**3**	**4**	**5**	**6**	**∑**
Old	131	256	413	965	367	428	2560
New	180	178	477	206	151	88	1280

Cumulative Response Proportions: HS Groups

	Very sure new				*Very sure old*	
	1	**2**	**3**	**4**	**5**	**6**
Old	1.00	0.94	0.84	0.67	0.31	0.17
New	1.00	0.85	0.70	0.30	0.15	0.06

Cumulative Response Proportions: LS Groups

	Very sure new				*Very sure old*	
	1	**2**	**3**	**4**	**5**	**6**
Old	1.00	0.95	0.85	0.69	0.31	0.17
New	1.00	0.86	0.72	0.35	0.19	0.07

z-Scores: HS Groups

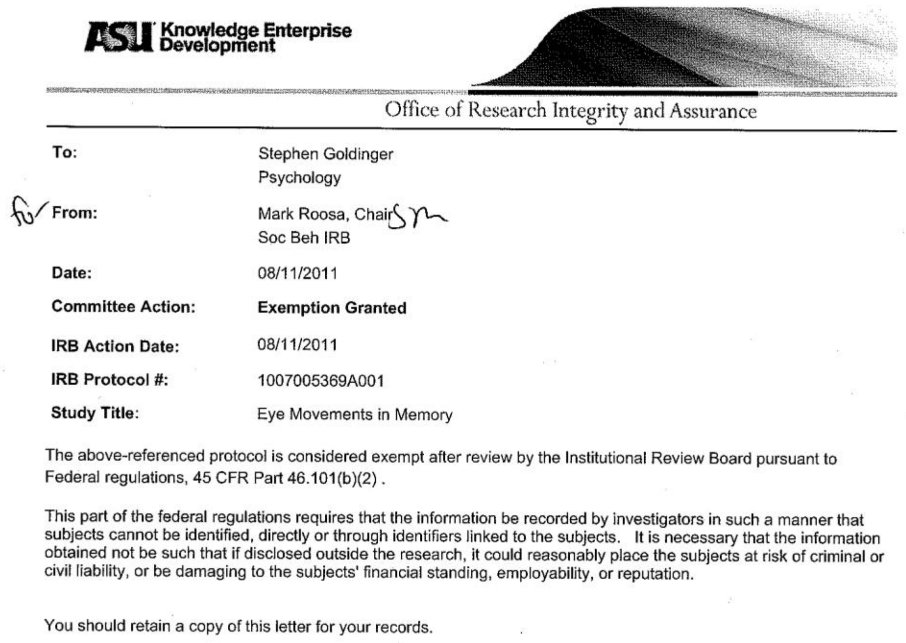
	Very sure new				*Very sure old*	
	1	**2**	**3**	**4**	**5**	**6**
Old	3.09	1.55	0.99	0.44	-0.52	-0.99
New	3.09	1.04	0.50	-0.52	-1.04	-1.65

z-Scores: LS Groups

	Very sure new				*Very sure old*	
	1	**2**	**3**	**4**	**5**	**6**
Old	3.09	1.55	0.99	0.47	-0.50	-0.99
New	3.09	1.04	0.58	-0.41	-0.92	-1.55

APPENDIX L

INSTITUTIONAL REVIEW BOARD APPROVAL



ASU Knowledge Enterprise Development

Office of Research Integrity and Assurance

To: Stephen Goldinger
Psychology

fo/ **From:** Mark Roosa, Chair
Soc Beh IRB

Date: 08/11/2011

Committee Action: **Exemption Granted**

IRB Action Date: 08/11/2011

IRB Protocol #: 1007005369A001

Study Title: Eye Movements in Memory

The above-referenced protocol is considered exempt after review by the Institutional Review Board pursuant to Federal regulations, 45 CFR Part 46.101(b)(2) .

This part of the federal regulations requires that the information be recorded by investigators in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. It is necessary that the information obtained not be such that if disclosed outside the research, it could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.

You should retain a copy of this letter for your records.