Psychological Review
2012, Vol. 119, No. 1, 40–79

0033-295X/11/$12.00 DOI: 10.1037/a0025464

Models of Recognition, Repetition Priming, and Fluency: Exploring a New Framework

Christopher J. Berry, David R. Shanks, and Maarten Speekenbrink
University College London

Richard N. A. Henson
Medical Research Council Cognition and Brain Sciences Unit, Cambridge, England

We present a new modeling framework for recognition memory and repetition priming based on signal detection theory. We use this framework to specify and test the predictions of 4 models: (a) a single-system (SS) model, in which one continuous memory signal drives recognition and priming; (b) a multiple-systems-1 (MS1) model, in which completely independent memory signals (such as explicit and implicit memory) drive recognition and priming; (c) a multiple-systems-2 (MS2) model, in which there are also 2 memory signals, but some degree of dependence is allowed between these 2 signals (and this model subsumes the SS and MS1 models as special cases); and (d) a dual-process signal detection (DPSD1) model, 1 possible extension of a dual-process theory of recognition (Yonelinas, 1994) to priming, in which a signal detection model is augmented by an independent recollection process. The predictions of the models are tested in a continuous-identification-with-recognition paradigm in both normal adults (Experiments 1–3) and amnesic individuals (using data from Conroy, Hopkins, & Squire, 2005). The SS model predicted numerous results in advance. These were not predicted by the MS1 model, though could be accommodated by the more flexible MS2 model. Importantly, measures of overall model fit favored the SS model over the others. These results illustrate a new, formal approach to testing theories of explicit and implicit memory.

Keywords: recognition, signal detection theory, repetition priming, fluency, computational model

Memory can express itself in different ways. Two memory phenomena that are often compared are recognition and repetition priming. *Recognition* refers to the capacity to judge whether a particular item (e.g., a word) has been previously presented in a particular context (e.g., with old–new judgments). *Repetition priming* (henceforth priming) refers to a long-term change in the identification, detection, or production of an item as a result of prior exposure to that item. This change often takes the form of facilitation in performance. For example, identification reaction times (RTs) to words that were presented on a recent study list are typically shorter than to words that have not been recently presented. Comparisons of recognition and priming have proven to be highly influential in the construction of multiple-systems theories of memory. A controversial issue is whether multiple-systems theories are necessary to explain recognition and priming, or whether a single-system (SS) theory will suffice. Here we take a new approach to investigating this question by testing the predictions of formal SS and multiple-systems models. We present a novel modeling framework that is based upon signal detection theory (SDT) and use this framework to specify models that differ in the number of memory signals driving recognition and priming, and also in the degree of dependence between these memory signals. We find that it is surprisingly difficult to reject a relatively simple SS model in favor of multiple-systems models. At the very least, we hope to motivate further formal modeling approaches to the study of recognition and priming.

This article was published Online First October 24, 2011.

Christopher J. Berry, David R. Shanks, and Maarten Speekenbrink, Department of Cognitive, Perceptual and Brain Sciences, University College London, London, England; Richard N. A. Henson, Medical Research Council Cognition and Brain Sciences Unit, Cambridge, England.

Preparation of this article was supported by U.K. Economic and Social Research Council Grants RES-063-27-0127 and RES-062-23-1511, by a Wellcome Trust Value in People award to Christopher J. Berry, and by the U.K. Medical Research Council (MC_US_A060_0046). We thank Hania Ghauri for assistance with data collection in Experiment 3. We also thank Klaus Oberauer for helpful discussions and David Huber for insightful comments. Finally, we thank Larry Squire and Jennifer Frascino for kindly sending us the raw data from Conroy et al. (2005).

Correspondence concerning this article should be addressed to Christopher J. Berry, Department of Cognitive, Perceptual and Brain Sciences, University College London, Gower Street, London, WC1E 6BT, England. E-mail: christopher.berry@ucl.ac.uk

Multiple-Systems Theory

An influential and largely dominant view is that recognition and priming are driven by functionally and neurally distinct explicit and implicit memory systems (e.g., Gabrieli, 1998; Squire, 1994, 2004, 2009; Tulving & Schacter, 1990). The terms *explicit* and *implicit memory* are actually used in several different ways: Implicit memory can be used to refer to memory that is revealed in the absence of conscious recollection, in contrast with expressions of explicit memory that are accompanied by conscious recollection (Graf & Schacter, 1985; Schacter, 1987). The terms can refer to particular classes of memory task (Roediger & McDermott, 1993) and are also often used to refer to hypothesized memory stores, systems, or sources (e.g., as in the terms *declarative* and *nondeclarative*, which are used instead of *explicit* and *implicit* by

Squire, 1994, 2004, 2009). Here we are primarily concerned with this latter usage. Squire (1994, 2004, 2009) has also proposed that the contents of declarative memories are accessible to awareness, whereas the contents of nondeclarative memories are not. A variety of dissociation evidence has been marshaled in support of multiple-systems views in general. We now review the main evidence.

Functional Dissociations

Functional dissociations—where a variable is shown to produce different effects on recognition and priming—have been found in normal adults and have been taken as evidence that explicit and implicit memory systems operate according to different principles (for an early and comprehensive review of the dissociation evidence, see Roediger & McDermott, 1993). For example, recognition is much greater for items that are processed semantically at encoding (e.g., by answering questions about the item's meaning) than those processed nonsemantically at encoding (e.g., by deciding whether the item contains a particular letter), whereas this type of "levels of processing" manipulation has either no effect on priming (e.g., Jacoby & Dallas, 1981; Richardson-Klavehn, Clarke, & Gardiner, 1999) or only a small effect (Brown & Mitchell, 1994; Challis & Brodbeck, 1992; Meier, Theiler-Bürgi, & Perrig, 2009).

Manipulations have also been identified that can produce the reverse pattern (viz. greater effects on priming than on recognition). These manipulations typically involve changing the physical form of an item between study and test. These findings have been taken as evidence for the highly specific nature of the (implicit) memory that supports priming, compared with recognition (Schacter, Dobbins, & Schnyer, 2004; Tulving & Schacter, 1990). For example, studies have found that presenting items in different modalities at study and test has little effect on recognition but affects priming (e.g., Craik, Moscovitch, & McDowd, 1994; Jacoby & Dallas, 1981; but see Lukatela, Moreno, Eaton, & Turvey, 2007, who argued that priming is unaffected by changes in modality when study–test asymmetries are controlled for, and Mulligan & Osborn, 2009, for evidence that recognition can be affected by modality).

Finally, certain variables have been shown to produce opposite effects on performance in each task. For example, Jacoby (1983) presented words at test in a perceptual identification task and a recognition task. Priming was greater when the word was read at study compared with when it was generated from its antonym (e.g., the word *cold* is generated from the cue *hot*–???), whereas recognition was greater for the generated words than read words (see Dew & Mulligan, 2008, for a recent replication in the auditory modality; but note that not all generate manipulations have been found to reduce priming; see studies by Masson & MacLeod, 1992, 2002, and Mulligan & Dew, 2009). The differences between these findings concerning functional effects on priming may relate to the range of different types of stimuli and tasks that have been used to measure priming.

Stochastic Independence

Findings of stochastic independence between performance in recognition and priming tasks have been viewed as strong evidence for multiple memory systems. Specifically, stochastic independence has been taken as evidence that the memory source driving an item's likelihood of being recognized is independent of the memory source driving the extent to which it shows priming (Tulving, 1985, 1999; Tulving & Schacter, 1990; Tulving, Schacter, & Stark, 1982). Stochastic independence refers to the relation between two events in which the probability of their joint occurrence is equal to the product of the probabilities of the occurrence of each event alone. For example, stochastic independence between recognition and priming in a word fragment completion task would be shown if the probability of recognizing a certain study item is independent of whether its word fragment will be successfully completed (for related findings where the priming measure is naming latency, see, e.g., Mitchell & Brown, 1988; Mitchell, Brown, & Murphy, 1990).

The use of stochastic independence as evidence for distinct memory systems has been widely criticized on methodological and theoretical grounds (e.g., Hintzman, 1990; Howe, Rabinowitz, & Grant, 1993; Ostergaard, 1992; see Poldrack, 1996, for a review). For example, this type of evidence relies on the acceptance of the null hypothesis, and Poldrack (1996) has shown that the statistical power needed to reasonably determine whether two measures are stochastically dependent is not achieved by many studies purporting to demonstrate stochastic independence. Furthermore, Ostergaard (1992) has argued that only a small proportion of the variance in performance on priming tasks is due to the influence of memory, and the influence of nonmemorial factors is greater on priming tasks than recognition, which is generally a more sensitive measure of memory. As a result, low or near-zero correlations between recognition and priming performance would be expected even if the same memory source drives them. Indeed, we have previously shown that an SS model of recognition and priming can produce very low correlations between recognition and priming performance (Berry, Henson, & Shanks, 2006; Kinder & Shanks, 2003).

Neuropsychological Dissociations

The evidence that is typically regarded as providing the most compelling support for multiple-systems theories comes from individuals with amnesia (e.g., arising from damage to the hippocampus or medial temporal lobes) who can show impairments in recognition and yet relatively normal levels of priming (e.g., Cermak, Talbot, Chandler, & Wolbarst, 1985; Graf, Squire, & Mandler, 1984; see also Fleischman & Gabrieli, 1998, and Fleischman, 2007, for discussion of this type of dissociation in Alzheimer's disease and normal aging). Particularly striking is that a profoundly amnesic individual, E.P., has been found to consistently perform at chance on recognition tests (Stefanacci, Buffalo, Schmolck, & Squire, 2000) despite showing comparable priming to controls (e.g., Conroy, Hopkins, & Squire, 2005; Hamann & Squire, 1997). Furthermore, individuals with damage to the right occipital lobe have been found to show the opposite pattern to amnesic individuals, demonstrating impaired (visual) priming despite relatively intact recognition (e.g., Gabrieli, Fleischman, Keane, Reminger, & Morrell, 1995; though this evidence has been questioned; see Yonelinas et al., 2001). Together with the dissociation in amnesia, this constitutes a double dissociation, providing support for distinct neural systems. The results of functional imaging studies also support the idea that recognition and priming are associated with distinct neural correlates (e.g., Schott et al., 2005, 2006; Woollams, Taylor, Karayanidis, & Henson, 2008).

Fluency

Investigations into the role of fluency in recognition have provided further evidence for the independence of the memory sources driving recognition and priming. *Fluency* in this context refers to the speed of processing an item. One possibility is that recognition and priming may be related via fluency. For example, an increased fluency, induced by prior exposure to an item, can lead to faster identification times (priming), and conscious detection of this unexpected fluency can also be attributed to the exposure, yielding recognition. Indeed, some theories have proposed that such fluency is one important basis of recognition judgments (Jacoby & Dallas, 1981; Mandler, 1980). This idea was recently investigated by Conroy et al. (2005) in a study in which both amnesic individuals and controls were tested in a continuous-identification-with-recognition (CID-R) paradigm (Feustel, Shiffrin, & Salasoo, 1983; Stark & McClelland, 2000). In this paradigm, an item (e.g., a word) gradually clarifies on each trial (e.g., by the slow removal of pixels that are obscuring the word), and participants press a button when they can identify the item (the identification RT forms the basis of priming and fluency measures). Participants then decide whether that item was presented in a prior study phase by making a yes–no recognition judgment. This CID-R paradigm therefore allows measures of recognition and an identification RT (the basis of priming and fluency) to be obtained for every item (for other studies that have used CID-R paradigms, see, e.g., Johnston, Dark, & Jacoby, 1985; Johnston, Hawley, & Elliott, 1991; Stark & McClelland, 2000; Verfaellie & Cermak, 1999).

Amnesic individuals in Conroy et al.'s (2005) study showed the typical dissociation between recognition and priming: Despite impaired levels of recognition, priming was comparable to that of controls. Furthermore, the amnesic individuals showed comparable fluency effects to controls, where *fluency effect* refers to the tendency for items judged old to have shorter identification RTs than items judged new (regardless of their actual study status). These findings were taken as evidence that amnesic individuals were sensitive to fluency but did not use it to support accurate recognition. Indeed, Conroy et al. conducted further analyses to estimate the potential contribution of fluency from priming to recognition (in patients and controls) and found that the estimates were very low indeed, and were much lower than would be required to account for the observed levels of recognition performance (see also Poldrack & Logan, 1997, for a similar analysis and result). The apparent lack of a contribution of fluency to recognition in their study is further evidence for the independence of recognition and priming (but see Berry, Shanks, & Henson, 2008a).

Priming in the Absence of Recognition

Evidence of priming in the absence of recognition has been taken as support for what is arguably a defining feature of implicit memory, that the contents of the memory source driving priming are not accessible to awareness (Roediger & McDermott, 1993; Stadler & Roediger, 1998). Priming without recognition has been taken as corroborating multiple-systems theories that make a fundamental distinction between explicit and implicit sources of memory more generally (Hamann & Squire, 1997). One approach has been to try to obtain priming when overall recognition performance is not reliably different from chance. However, previous research has shown that this is a difficult result to attain. For example, although some have found evidence of priming effects in the absence of recognition when using attentional manipulations at encoding, attempts to replicate these findings have been unsuccessful (see Berry, Shanks, & Henson, 2006; Berry, Shanks, Li, Rains, & Henson, 2010), and the results of other studies suggest that overall levels of priming diminish when recognition is close to chance levels (e.g., Berry, Henson, & Shanks, 2006; Moscovitch & Bentin, 1993; see also Shanks & St. John, 1994, for further discussion; for a review of previous attempts to demonstrate priming in the absence of recognition using attentional manipulations at encoding, see Mulligan, 2008). Another approach is to show that priming can occur even for items that are not recognized in a recognition task (e.g., that within the subset of items judged new, the identification RTs to studied items [misses] are shorter than those of new items [correct rejections]; Stark & McClelland, 2000; but see Berry et al., 2008a, and below, for a different interpretation of this result).

Alternative Accounts

The idea that dissociations imply multiple systems or processes, in general, is controversial (e.g., Benjamin, 2010; Buchner & Wippich, 2000; Dunn & Kirsner, 1988, 2003; Hintzman, 1990; Kinder & Shanks, 2001, 2003; Newell & Dunn, 2008; Plaut, 1995; Van Orden, Pennington, & Stone, 2001), as is the idea that memory systems or processes divide on consciousness (Dew & Cabeza, 2011; Henke, 2010; Reder, Park, & Kieffaber, 2009; Roediger & McDermott, 1993; Shanks & St. John, 1994). For example, it has become clear that single dissociations (cases where an independent variable affects one measure but has no detectable effect on another) can be artefactual in nature, and can arise because of differences in the reliabilities of the tasks and not necessarily because of differences in memory systems. Priming tasks tend to have a lower reliability than recognition (e.g., as measured by split-half correlations; Buchner & Brandt, 2003; Buchner & Wippich, 2000; Meier & Perrig, 2000). On purely statistical grounds, this means that priming is less likely to show a detectable effect of independent variables than is recognition, producing an apparent single dissociation.

Furthermore, to assert that a variable has no effect on a particular measure requires accepting the null hypothesis, and this can be difficult to justify given that the size of a true effect could be extremely small (Dunn, 2003). This same criticism applies when two (opposite) single dissociations are used together to constitute a double dissociation (e.g., the double dissociation between amnesiacs and individuals with damage to the right occipital lobe; Gabrieli et al., 1995), or to assert that a particular measure is at chance (e.g., recognition), where again it could be argued that the failure to detect an effect does not mean that the effect does not exist.

In addition, the notion that priming is normal (intact) in amnesia is controversial. For example, Ostergaard has claimed that when carefully assessed, impairments in priming are evident in amnesia (e.g., Jernigan & Ostergaard, 1993; Ostergaard, 1999; Ostergaard & Jernigan, 1993, 1996; see also Meier et al., 2009; but see Hamann, Squire, & Schacter, 1995). Priming effects are often proportional to baseline levels of performance in priming tasks (in controls, Ostergaard, 1998, and amnesic individuals, Ostergaard, 1994). Baseline levels of performance are often worse in amnesic individuals than controls (e.g., identification RTs are longer overall in perceptual identification and perceptual clarification tasks), which could lead to elevated levels of priming, effectively masking any priming deficit in these individuals. Indeed, when differences in baselines are equated between amnesic individuals and controls, amnesic individuals have been found to show lower levels of priming than controls (Ostergaard, 1994).

If baseline performance in priming tasks is affected by nonmemorial factors that are unrelated to memory (e.g., the amount of perceptual information available from a stimulus at test; Ostergaard, 1992, 1998, 1999), and these factors act to constrain priming effects, then reducing the influence of these factors should increase priming effects (see Ostergaard, 1998, for a formal model of priming that embodies these assumptions, the information availability model). In line with this, Ostergaard (1999) has shown that under presentation conditions that are more visually demanding (by gradually revealing words over a relatively long rather than short duration), clear impairments in priming are evident in amnesia. This evidence appears to undermine the view that the memory system crucial for priming is selectively spared in amnesia and suggests that amnesia affects a single system that is crucial for both priming and recognition.

Formal Models of Recognition and Priming

A limitation of multiple-systems accounts of recognition and priming is that there have been few attempts to develop and test formal multiple-systems models. Formal models offer many benefits. First, they promote theoretical transparency compared with purely descriptive theories, which are notoriously susceptible to alternative interpretation by their very nature. Second, competing models can be specified, and distinct empirical predictions can be derived. Thirdly, formal model selection measures can be used to select between models on the basis of their fit and number of free parameters (e.g., with the Akaike information criterion [AIC]; Akaike, 1973). Indeed, in other related research fields, tests of competing models have led to considerable theoretical development (e.g., as with signal detection models of recognition memory; Wixted, 2007; Wixted & Mickes, 2010; Yonelinas & Parks, 2007). Finally, even simple models can yield counterintuitive results that may have otherwise been taken as evidence for a more complex theory (see, e.g., Shanks, 2005).

The general goal of the present article is to present a new formal framework within which models that make different assumptions about contributing systems (e.g., single vs. multiple) can be formulated and compared. We employ this framework to derive several key predictions that discriminate between the models, and we report experiments testing these predictions. We introduce four new formal models of recognition and priming. We begin by describing the simplest model that can be expressed in this framework—an SS model—and then go on to describe the other models and general framework in more detail.

A Single-System Model

We have previously proposed an SS computational model that extends SDT (Green & Swets, 1966; Macmillan & Creelman, 2005) of recognition memory to priming (and also fluency; Berry, Henson, & Shanks, 2006; Berry et al., 2008a; see Berry, Shanks, & Henson, 2008b, for an overview).[1] The main assumption of the SS model (see Figure 1A) is that the same memory strength signal drives recognition, priming, and fluency. Each item in the test phase is associated with a memory-strength-of-evidence variable, f, which is a random variable drawn from a normal (Gaussian) distribution with standard deviation σ_f (i.e., $f \sim N(\mu_I, \sigma_f)$, where the I subscript stands for item type and I = old, new).

Because of prior exposure, the mean f of studied items is assumed to be greater than that of new items (i.e., $\mu_{\text{new}} = 0$ and $\mu_{\text{old}} \geq 0$). To simulate recognition, one value of f is sampled for each individual item from the relevant (old or new) distribution, and this value of f is combined with a randomly sampled, normally distributed noise value (e_r) to produce a recognition strength value, J_r:

$$J_r = f + e_r \quad e_r \sim N(0, \sigma_r). \tag{1}$$

The noise parameter e_r is specific to the recognition task, has a mean of 0, and the standard deviation of this noise (σ_r) is fixed to equal σ_f. As in SDT, old–new judgments are modeled by assuming that participants have some criterion of strength (C) that must be exceeded in order for an old judgment to be made: If an item's value of J_r exceeds C, then the item will be judged old, otherwise it will be judged new (see Figure 2). Adding e_r to f in Equation 1 is formally equivalent to adding e_r to C (see Benjamin, Diaz, & Wee, 2009, for further discussion on variability in the decision criterion in recognition).[2]

Performance for the priming task is modeled in a similar manner. Here we will describe the application of the model to a CID-R priming task, although we have previously applied the model to other commonly used priming tasks such as perceptual identification (Berry, Henson, & Shanks, 2006) and picture fragment identification (Berry et al., 2010). We focus upon the CID-R task here because this paradigm will be used to test the models later. It is important to note from the outset that we do not model all the various processes and mechanisms involved in identification of an item; rather, it is the influence of memory upon identification that is being modeled. To generate an item's identification RT, the same value of f that was used to calculate that item's J_r is combined with another source of normally distributed random noise (e_p). Importantly, e_p is not correlated with e_r and, like e_r, is a normally distributed random variable, with a mean of 0, and is drawn from the same distribution for old and new items. In other words, e_p and e_r represent task-specific noise that is completely independent of any "memory." The parameter e_p can be conceptualized as representing the influence of nonmemorial factors upon task performance. These factors decrease the sensitivity of priming tasks to

[1] In previous applications of the SS model, the parameters were fit to data averaged across trials and participants with (random) Monte Carlo simulation methods. In this article, we present a completely new formulation of the SS model in terms of a general framework in which (deterministic) maximum likelihood estimation techniques are used to fit the data from every recognition or priming trial for every individual.

[2] SDT models of item recognition are largely now believed to require unequal (greater old item) variance, whereas here we make the simplifying assumption of equal old and new item variance. We deal with this issue in greater detail in Experiment 2 (see Footnote 11).

the influence of memory, and may do so to a greater extent than in recognition tasks (see Ostergaard, 1992, 1998, 1999, for discussion). Indeed, in previous applications of the model, we showed how σ_p needs to be greater than σ_r in order to fit data patterns (e.g., Berry, Henson, & Shanks, 2006). Identification RT is assumed to be a decreasing function of *f*:

$$RT = b - sf + e_p \quad e_p \sim N(0, \sigma_p), \tag{2}$$

where the parameters *b* and *s* are scaling parameters whose value is greater than 0: The *b* parameter represents the identification RT intercept (and *b* is equal to the expected identification RT for new items predicted by the model), and the *s* parameter represents the rate of change in RT with *f* (and so *s* multiplied by μ is equal to the expected priming effect predicted by the model, i.e., the mean difference in identification RTs to old and new items). Thus, a greater value of *f* for an item increases the chance that it will be judged old and also increases the likelihood that it will have a relatively short identification RT. Because the mean *f* for old items is assumed to be greater than that of new items (i.e., $\mu_{old} \geq 0$), the model will simulate above-chance recognition discrimination performance (i.e., there will tend to be a greater proportion of old items with values of J_r that exceed *C* and are judged old); the model will also simulate the standard priming effect (i.e., identification RTs to old items will tend to be shorter than those of new items); and it will also reproduce overall fluency effects (i.e., items judged old will tend to have shorter identification RTs than items judged new, as shown below).[3]

The model can account for a number of basic findings. First, it can produce single dissociations in which a variable has a greater

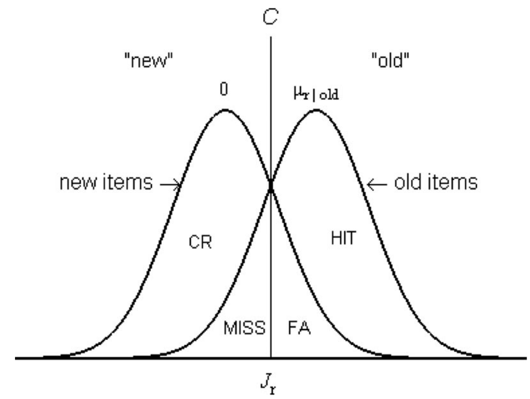

Figure 2. Signal detection theory (SDT) of old–new recognition judgments. J_r represents the recognition strength-of-evidence variable in the single-system (SS) model (analogous to the strength-of-evidence variable in standard SDT). Zero and $\mu_{r|old}$ denote the means of the new and old item distributions, respectively (in the SS and dual-process signal detection models, $\mu_{r|old} = \mu_{old}$). *C* represents the decision criterion. Items to the right of *C* will be judged "old," and those to the left of *C* will be judged "new." An old item is classified as a hit if judged old or a miss if judged new; a new item is classified as a false alarm (FA) if judged old or a correct rejection (CR) if judged new.

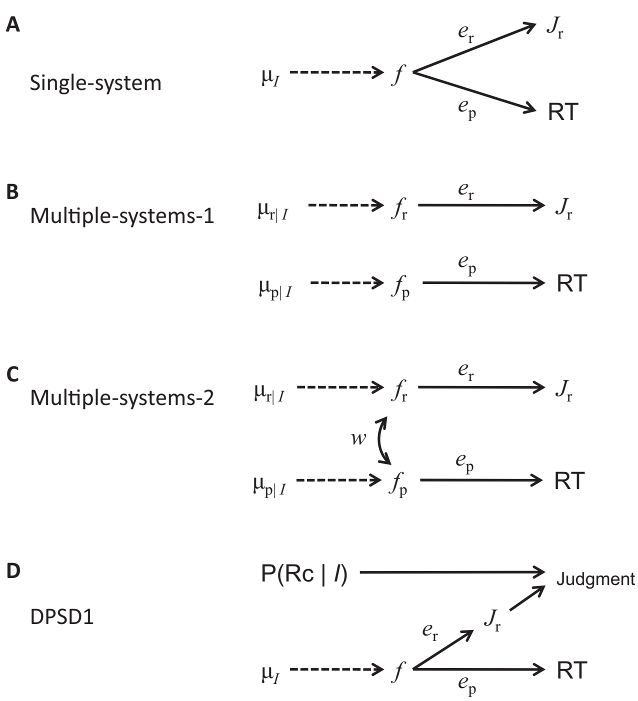

Figure 1. Four models of recognition and priming. (A) Single-system (SS) model: Each item's value of *f* is sampled (represented by the dotted line) from a distribution with mean μ_I (where the subscript *I* stands for item type and I = old, new). The same value of *f* is used to generate J_r and the identification reaction time (RT). (B) Multiple-systems-1 model: The values f_r and f_p are sampled from independent distributions of memory strength, and f_r and f_p are uncorrelated. (C) Multiple-systems-2 model: The values f_r and f_p are sampled from independent distributions of strength, but f_r and f_p can be correlated (by an amount equal to *w*). (D) Dual-process signal detection (DPSD1) model: As in the SS model, one value of *f* is used to calculate an item's J_r and RT, but there is an additional recollection process, $P(\text{Rc}|I)$ (where Rc stands for recollection), which can drive recognition judgments. If an item is not recollected, its recognition judgment is based on J_r.

[3] Equation 2 assumes that the identification RT distributions in the CID-R paradigm are normally distributed. Normality is not considered to be typical of RT distributions in general. However, in keeping with Equation 2, we found that the vast majority of the identification RT distributions from the CID-R data that we model here (i.e., the old and new item identification RT distributions of each participant in Experiments 1–3 and Conroy et al., 2005, in the Amnesia Modeling Study) were not significantly different from a normal distribution, as tested with Lilliefors (Kolmogorov–Smirnov) and Anderson–Darling tests of normality: With Lilliefors (Kolmogorov–Smirnov), 73% of RT distributions were nonsignificant; with Anderson–Darling, 69% of RT distributions were nonsignificant. The assumptions of normality were made for the sake of computational simplicity and ease of model specification; however, it is possible that the characteristics of the identification RT distributions are better described by more complex formulations of the model in which they are assumed to be nonnormal, or by sequential sampling accounts that consider the processes involved in RT production in greater detail (e.g., Ratcliff & Starns, 2009). Indeed, such formulations would enable the framework to be applied to other RT measures more generally. Equation 2 also assumes that identification RT is a linear function of *f*, and this can lead to some implausible results. For example, if *f* were sufficiently large, then identification RTs can be negative. However, for the range of *f* that can reasonably be expected to occur within an experiment, Equation 2 can be considered to be adequate for the purposes of modeling the influence of *f* on identification RTs.

effect on recognition than priming by assuming that the variance of the noise associated with the priming task (σ_p) is greater than that of recognition. Given this assumption, as the overall strength of the underlying memory signal increases, recognition will increase at a greater rate than priming (when performance is measured on the same scale, see, e.g., Berry, Henson, & Shanks, 2006). This assumption is supported by the aforementioned evidence showing that priming tasks typically have a lower reliability compared with recognition (e.g., Buchner & Wippich, 2000).

Second, because of the assumption that the noise associated with the priming task is typically greater than that of recognition, the model predicts that the sensitivity (e.g., d') of priming tasks will not typically exceed that of recognition. This is consistent with our previous research in which we have found that (a) the magnitude of priming does not exceed that of recognition when both are measured upon the same response metric, and (b) priming does not occur when recognition is observed to be at chance (Berry, Henson, & Shanks, 2006; Berry, Shanks, & Henson, 2006; Berry et al., 2010). Third, because of the uncorrelated, task-specific sources of noise associated with recognition and priming, the model is able to produce very low correlations between recognition and priming performance (Berry, Henson, & Shanks, 2006; Berry et al., 2008a; but see Rünger, Nagy, & Frensch, 2009). Fourth, the model predicts that priming will occur even for items that are not recognized (Berry et al., 2008a; cf. Stark & McClelland, 2000). From an SDT perspective, the reason for this is quite trivial: When $\mu_{old} > 0$, misses will tend to have a higher memory strength than correct rejections, even though that strength is not high enough to surpass the criterion for responding "old" (i.e., $J_r < C$; see Figure 2). Fifth, the random noise in the model explains why the relationship between the identification latencies to misses and false alarms can change as a function of overall memory strength (Johnston et al., 1985; see Berry et al., 2008a) and, sixth, explains why the estimate of the contribution of fluency to recognition can be extremely low (Conroy et al., 2005; see Berry et al., 2008a). Thus, somewhat counterintuitively, the SS model is able to account for many findings that prima facie appear to be indicative of the involvement of distinct memory systems in recognition and priming.

Rather than simply show that many observations can be accounted for by the SS model, clearly a far more compelling test of a model is to show that it makes specific predictions in advance and, moreover, that these predictions are not made by multiple-systems models. Formal model selection measures (e.g., AIC) can also be used to select between competing models. We now present three multiple-systems models that will be tested against the SS model. The first two multiple-systems models assume two independent continuous memory signals, the values of which are either uncorrelated for a given item or allowed to be correlated; the third model combines a single continuous memory signal with a probabilistic recollection process (one possible extension of the "dual-process model"; Yonelinas, 1994, 2002).

Two Formal Multiple-Systems Models

The first multiple-systems model that we present is one in which the memory strength signals driving recognition and priming are independent at the level of individual items (i.e., stochastically independent) and at the level of the mean strength of the memory signal in the explicit and implicit memory systems (i.e., functionally independent). Stochastic and functional independence of explicit and implicit memory systems has been previously claimed (see, e.g., Mitchell & Brown, 1988; Mitchell et al., 1990; Tulving & Schacter, 1990; Tulving et al., 1982). The multiple-systems-1 (MS1) model that we present (shown in Figure 1B) is one formalization of such claims.

Whereas one value of f is sampled in the SS model to derive an item's recognition and priming performance, in the MS1 model two values of f are sampled for every item at test. One value, f_r, is drawn from an "explicit memory" distribution of memory strengths and drives recognition, where $f_r \sim N(\mu_{r|I}, \sigma_f)$, the mean explicit memory strength for new items, $\mu_{r|new}$, is equal to 0 and the mean strength of old items, $\mu_{r|old}$, is greater than or equal to 0. The other value, f_p, is drawn from an "implicit memory" distribution of memory strengths and drives priming, where $f_p \sim N(\mu_{p|I}, \sigma_f)$, the mean implicit memory strength for new items, $\mu_{p|new}$, is equal to 0 and the mean implicit memory strength for old items, $\mu_{p|old}$, is greater than or equal to 0. Crucially, $\mu_{r|old}$ and $\mu_{p|old}$ are free parameters. This assumption enables mean recognition and priming performance to be functionally independent. For example, the model could predict a mean priming effect over trials even in the absence of above-chance recognition because $\mu_{p|old}$ can be greater than 0 even when $\mu_{r|old}$ is equal to 0. It is this assumption that allows an experimental manipulation to have dissociable effects on priming and recognition, such as affecting $\mu_{r|old}$ but not $\mu_{p|old}$. A second consequence of sampling f_p and f_r in this way is that priming and recognition are uncorrelated across trials: A particularly high value of f_r may happen to be sampled to calculate an item's value of J_r, but it is unlikely that a comparably high value of f_p will again be sampled to calculate the item's identification RT. Thus, the MS1 model differs from the SS model in two important ways: (a) mean levels of priming and recognition can vary independently of each other, and (b) identification RTs and recognition judgments are uncorrelated across trials.

A more relaxed version of the MS1 model allows f_p and f_r to be correlated across items. A correlation could arise, for example, via the distinctiveness of an item: A relatively distinctive item may be encoded relatively strongly into the explicit and implicit memory systems (even though the mean memory strengths in the two systems are unrelated). A second multiple-systems model that incorporates this assumption is presented in Figure 1C, and we refer to it as the multiple-systems-2 (MS2) model. In the MS2 model $\mu_{r|old}$ and $\mu_{p|old}$ are still uncorrelated, but now f_r and f_p are correlated (i.e., f_r and f_p can be viewed as random variables drawn from a bivariate normal distribution), and this correlation is represented by w, which is a free parameter. This means that (a) like the MS1 model but unlike the SS model, mean levels of recognition and priming in the MS2 model can vary independently of each other, and (b) unlike the MS1 model but like the SS model, an item's recognition judgment and identification RT can be correlated in the MS2 model.

A General Modeling Framework

It is possible to describe the SS, MS1, and MS2 models within the same general framework. This is useful for the purposes of identifying the mathematical relationships between the models and also for the purposes of fitting the models to data. The variables f_r,

f_p, J_r, and RT can be viewed as random variables in a multivariate normal distribution with mean vector

$$\boldsymbol{\mu} = \begin{array}{c} f_r \\ f_p \\ J_r \\ RT \end{array}\begin{bmatrix} \mu_{r|I} \\ \mu_{p|I} \\ \mu_{r|I} \\ b - s\mu_{p|I} \end{bmatrix}, \tag{3}$$

and covariance matrix

$$\boldsymbol{\Sigma}_{\text{old}} = \boldsymbol{\Sigma}_{\text{new}} = \boldsymbol{\Sigma} =$$

$$\begin{array}{c} f_r \\ f_p \\ J_r \\ RT \end{array}\begin{bmatrix} \sigma_f^2 & & & \\ w\sigma_f^2 & \sigma_f^2 & & \\ \sigma_f^2 & w\sigma_f^2 & \sigma_f^2 + \sigma_r^2 & \\ -ws\sigma_f^2 & -s\sigma_f^2 & -ws\sigma_f^2 & s^2\sigma_f^2 + \sigma_p^2 \end{bmatrix}. \tag{4}$$

The correlation between J_r and RT (within old–new item type) is thus

$$\rho(J_r, RT) = \frac{-ws\sigma_f^2}{\sqrt{\sigma_f^2 + \sigma_r^2}\sqrt{s^2\sigma_f^2 + \sigma_p^2}}. \tag{5}$$

Considering the MS2 model in this framework, $\mu_{r|old}$ is free to vary independently of $\mu_{p|old}$, and w, the correlation between f_r and f_p, is also free to vary. Similarly, in the MS1 model, $\mu_{r|old}$ is also free to vary independently of $\mu_{p|old}$, but the value of w is fixed to equal 0; this means that J_r and RT are not correlated in this model (i.e., $\rho(J_r, \text{RT}) = 0$ when $w = 0$ in Equation 5). Lastly, specifying the SS model in this framework, $w = 1$, which results in $f_r = f_p = f$, and results in a nonzero correlation between J_r and RT. Furthermore, in the SS model, $\mu_{r|old} = \mu_{p|old} = \mu_{old}$. Thus, it is evident that the SS and MS1 models are actually special cases of the MS2 model (and are therefore nested under it mathematically).

Dual-Process Signal Detection (DPSD1) Model

Later in this article, we also consider the predictions of another model, based on the dual-process model of recognition memory (Yonelinas, 1994, 2002). In this model, recognition is assumed to be based on either a probabilistic "recollection" memory process (which occurs with a fixed probability, $P(\text{Rc}|I)$, where Rc stands for recollection and the I stands for item type, where I = old, new), or a continuous "familiarity" signal (akin to the signal in SDT and as in the above SS, MS1, and MS2 models). Recognition is based on the familiarity signal if recollection does not occur. Though the dual-process model has been applied mainly to recognition data, it has been suggested that the familiarity signal may be related to the same causes as priming (e.g., Jacoby & Dallas, 1981). One possible DPSD1 model of priming and recognition can be created by adding a recollective process to the SS model (see Figure 1D). Identification RTs and hence priming are determined by f as in the SS model. This DPSD1 model is considered a multiple-systems model in the sense that it includes independent sources of memory that contribute to recognition and priming. The DPSD1 model assumes that when recollection occurs, the value of f that is used to generate the identification RT is simply another random sample of f from the relevant distribution, rather than one that is likely to be greater than average. Though the DPSD1 model is not expressed within the same general multivariate normal framework described in Equations 3–5, a likelihood function can still be obtained by extending this framework (see Appendix A), which we use to fit the DPSD1 model to the data of Experiments 2 and 3, and we test its predictions explicitly in Experiment 3.[4]

Fitting the Models to Empirical Data

A main aim in the experiments reported here is to test specific predictions of the models, that is, whether they predict certain a priori patterns in the data (as tested by planned comparisons). For example, the SS and MS1 models make distinct predictions that can be pit against each another (as we describe in the next section). Note, though, that because the MS2 model can produce any result that the other two models can, it is not possible to produce a result that is diagnostic of the SS model or MS1 model over the MS2 model. Because of its flexibility, the MS2 model does not make firm predictions in advance in the same way that the SS and MS1 models do. Importantly, however, it is possible to obtain results that are diagnostic of the MS2 model over the other two models when looking at conjunctions of results (see the next section).

Another main aim is to compare the models on how well they fit data. In fitting the experiments described below, we determined the parameter values for each model by fitting them to the data of every trial (i.e., every pair of identification RTs and recognition judgments) for every participant, using maximum likelihood procedures. Because of the Gaussian assumptions, the likelihood function for the general framework above is relatively straightforward (see Appendix A for details). To accommodate the different numbers of free parameters in the different (nonnested) models, we also report the AIC (Akaike, 1973) measure, which is one way to take into account the number of free parameters in a model. The greater the number of free parameters in a given model, the greater the penalty imposed by the AIC. Thus, even though it is not possible to find an empirical pattern that is predicted by the nested models (SS and MS1) but not the general (MS2) model, it is in principle possible to find a data pattern that indicates that a nested model is more likely than the general model. Importantly, we validate our use of the AIC measure for selecting between models in model recovery simulations, in which we show that this measure allows for the true generative model of the data to be identified against other models (see Appendix B).

Model Predictions

We test the models by applying them to empirical data from three new experiments with normal adults and one published experiment with amnesic individuals by Conroy et al. (2005). The CID-R paradigm is well suited for testing the models because a recognition judgment and an identification RT can be obtained for every item (trial). Furthermore, because these measures are taken concurrently rather than at different time points, they are less likely to be differentially affected by other factors such as forgetting.

Recognition is measured with old–new judgments in Experiment 1 (and in Conroy et al., 2005, in the Amnesia Modeling Study, presented after Experiment 3), 6-point recognition confi-

[4] We did not apply the DPSD1 model to Experiment 1 or to the data of Conroy et al. (2005) because the type of recognition judgment in both these data sets was old–new and, under these circumstances, stable estimates of $P(\text{Rc}|I)$ in the DPSD1 model could not be obtained (see Appendix A).

dence ratings in Experiment 2, and remember–know judgments in Experiment 3 (Gardiner, 1988; Tulving, 1985). We now outline three predictions of the SS, MS1, and MS2 models in a CID-R task with old–new judgments. Later in the article we present two more predictions for performance in a modified CID-R paradigm with different types of recognition judgments.

> *Prediction 1:* A fluency effect within old and new items is predicted by the SS model but not the MS1 model. The MS2 model can produce either result.

That is, the SS model predicts that RT(false alarm) < RT(correct rejection) and also that RT(hit) < RT(miss), where a false alarm is an old judgment to a new item, a correct rejection is a new judgment to a new item, a hit is an old judgment to an old item, and a miss is a new judgment to an old item. This is because $\rho(J_r, \mathrm{RT}) < 0$ in Equation 5: An item that receives a relatively high sampled value of f is likely to end up with a relatively high value of J_r that exceeds C (after being combined with e_r) and be judged old. If an item's value of f is relatively high, its identification RT is also likely to be relatively short (after f is combined with e_p). This is the case regardless of whether the item is actually old or new.[5]

In contrast, the MS1 model predicts that RT(false alarm) = RT(correct rejection) and RT(hit) = RT(miss). This is because $\rho(J_r, \mathrm{RT}) = 0$ in Equation 5. Within an item type, identification RTs are uncorrelated with recognition judgments. Consider a new item that elicits a relatively high sampled value of f_r for the calculation of J_r. The value of J_r is likely to be high, and the item is therefore more likely to be judged old (and be classified as a false alarm). However, when f_p is sampled to calculate the same item's RT, f_p is unlikely to be as comparably high; in fact, the value of f_p is most likely to be equal to $\mu_{p|new}$. The RT of this new item is therefore most likely to be equal to the mean identification RT of all new items. Thus, the MS1 model predicts RT(false alarm) = RT(correct rejection) = RT(new items). A similar logic applies to old items: RT(hit) = RT(miss) = RT(old items).

Note that all the models can predict fluency effects when old and new items are considered as one pooled stimulus set. Even the MS1 model makes this prediction: When $\mu_{r|old}$ and $\mu_{p|old}$ are greater than 0, old items tend to be items that are judged old and also tend to be the items with relatively short identification RTs.[6] Previous evidence concerning whether fluency effects occur within old and new stimuli, however, is often not reported, or is mixed (see Berry et al., 2008a, for discussion).

> *Prediction 2:* The SS model predicts that the magnitude of priming for items judged new is smaller than the overall priming effect. The MS1 model predicts that the magnitude of both priming effects is the same. The MS2 model can produce either data pattern.

In CID-R paradigms, the overall priming effect is normally shown by shorter mean identification RTs to old items than new items. Priming effects can also occur even for items that are not recognized (judged new), that is, RT(miss) < RT(correct rejection) (e.g., Berry et al., 2008a; Stark & McClelland, 2000). This "priming in the absence of recognition memory" is important because it has been taken as evidence for separate sources of memory underlying explicit and implicit memory (Stark & McClelland, 2000). Yet, this result falls quite naturally from the SS model because f for misses tends to be greater than f for correct rejections (see Berry et al., 2008a). Importantly, though, the SS model predicts that the magnitude of priming for items judged new (misses) will be smaller than the normal "full" priming effect. This can be deduced from Figure 2: The difference in mean J_r (and by implication, the difference in mean f) between misses and correct rejections is smaller than the difference in mean J_r between old and new items overall. The difference in f of these items will tend to mirror this, meaning that the same relationship will be present in the generation of the identification RTs, and so RT(correct rejection) − RT(miss) < RT(new) − RT(old).

In the MS1 model, because RT(false alarm) = RT(correct rejection) = RT(new items), and because RT(hit) = RT(miss) = RT(old items) (as explained in Prediction 1 above), it necessarily follows that RT(correct rejection) − RT(miss) = RT(new) − RT(old). That is, the MS1 model predicts that the magnitude of priming for items judged new will be equal to the normal priming effect.

> *Prediction 3:* The MS1 and MS2 models can produce a priming effect when recognition is at chance. The SS model predicts that an overall priming effect will never occur in the absence of overall recognition.

A reliable, replicable finding of an overall priming effect when overall recognition performance is (truly) at chance (i.e., RT(new) − RT(old) > 0, when $d' = 0$) would be strong evidence against the SS model, and would be evidence for multiple-systems theory. It is worth noting that such a result would also be evidence for what is arguably a defining feature of implicit memory—that its contents are not accessible to awareness (Squire, 1994, 2009). In terms of the general modeling framework outlined above, for this result to occur, $\mu_{r|old}$ must equal 0 when $\mu_{p|old}$ is greater than 0. In other words, there would need to be a memory strength signal that can drive priming even though there is a complete absence of a memory signal to drive recognition. This cannot occur in the SS model because $\mu_{r|old} = \mu_{p|old} = \mu_{old}$, and so when $\mu_{r|old} = 0$, $\mu_{p|old}$ must also equal 0; when overall recognition is at chance, priming should similarly be absent. However, this result could

[5] The SS model does not always predict that differences in the mean identification RT of items mirror differences in J_r. For example, from looking at Figure 2, one might expect to observe RT(false alarm) < RT(miss). However, the relationship between the identification RTs to misses and false alarms has been shown to be variable, in that it depends on whether overall memory strength is high or low (Johnston et al., 1985). In Berry et al. (2008a), we described simulations of this result with the SS model, where this variable relationship is explained simply by the principle of regression to the mean.

[6] The calculation of the fluency effect across all items is equivalent to using the weighted means (rather than the unweighted means) of the RTs to hits, misses, false alarms, and correct rejections to calculate the fluency effect. Thus, it is not the case that RT(hit) + RT(false alarm) = RT(miss) + RT(correct rejection) (i.e., that the MS1 model does not predict a fluency effect across all items), as might be suggested given that the MS1 model predicts that RT(false alarm) = RT(correct rejection) and RT(hit) = RT(miss).

occur in the MS1 and MS2 models because $\mu_{r|old}$ and $\mu_{p|old}$ are free to vary independently of each other. Furthermore, if (a) associations are found between recognition judgments and identification RTs, as predicted by the SS model in Predictions 1 and 2, and (b) overall priming in the absence of recognition is found, then this pattern of results would be indicative of the MS2 model, because it would be evidence that (a) a nonzero correlation exists between f_p and f_r, and (b) $\mu_{r|old}$ and $\mu_{p|old}$ are free to vary independently of each other.

A related prediction of the SS model is that the sensitivity of priming measures (e.g., as indexed by d') will never exceed that of recognition when priming and recognition tasks are comparable and performance is assessed with the same response metric (see Berry, Henson, & Shanks, 2006). This is because, as previously discussed, the variance of the noise associated with priming tasks (e.g., perceptual identification) is typically greater than that of the recognition task. Thus, whereas the SS model does not predict that priming will occur in the absence of recognition, a finding that recognition is greater than chance when levels of priming are too low to be detected would not be inconsistent with the model.

Experiment 1

The aim of Experiment 1 was to test Predictions 1–3 using the CID-R task to measure recognition, priming, and fluency. To allow Prediction 3 to be tested, strong and weak memory conditions were created by using a manipulation of selective attention at encoding. Pairs of words were presented on each trial, one above the other, for 1,000 ms, and participants were instructed to read aloud the word that was cued by arrows while ignoring the uncued word. We have previously shown that this manipulation can produce strong and weak memory conditions: Recognition and priming (as measured in a perceptual identification task) for cued items is typically greater than for uncued items (Berry, Henson, & Shanks, 2006). Similar manipulations of attention have been claimed to produce priming effects in the absence of recognition (e.g., Merikle & Reingold, 1991; but see Berry, Shanks, & Henson, 2006).

At test, previously cued-study, uncued-study, and new (unseen) words were presented in the CID-R task. On each CID-R trial, a word was gradually revealed over 7,500 ms. Participants pressed a button when they could identify the word (and this latency, the identification RT, was recorded), which terminated the gradual exposure of the word. They typed the word on the keyboard and then made an old–new recognition judgment to the word. Cued and uncued words were presented in different test conditions with different new items in each condition. The new items in each condition served as the appropriate baseline (within condition) for the calculation of measures of recognition, priming, and fluency.

Method

Participants. Thirty-two participants (20 female, 12 male; mean age = 23 years; range: 18–34) were recruited from a University College London participant database. Each received £6 for participating. All participants completed one test condition containing cued and new items and another test condition containing uncued and new items. The order of presentation of these test conditions was counterbalanced across participants. All participants in this and subsequent experiments reported normal or corrected-to-normal vision and English as their first language.

Materials. The stimuli were 288 four-letter words, selected from the Medical Research Council psycholinguistic database (Coltheart, 1981). All words had a low frequency of occurrence (1–10 per million; Kučera & Francis, 1967), and archaic and colloquial terms were excluded. Four 72-word lists were constructed. Each list served as either the cued, uncued, new-cued (new words presented in the cued condition), or new-uncued words (new words presented in the uncued condition). The assignment of lists to each type of item was counterbalanced across participants according to a Latin square.

Another 24 words were selected with the same constraints to serve as the stimuli for the first and final six study trials (included to minimize primacy and recency effects). These words were not presented again in the experiment. A further 25 words were selected as words to be presented in a continuous identification (CID) practice phase. Again, these were selected with similar constraints to the other words and were not subsequently presented in the experiment. All words were presented in 20-point Courier font.

Procedure. The procedure for all participants is detailed below.

Study. At the start of each study trial a "+" fixation was presented for 500 ms in the center of the screen that was viewed from approximately 75 cm. The screen was then blanked for 200 ms. Next, two words were presented simultaneously for 1,000 ms, one 0.9 cm (subtending a vertical visual angle of approximately 0.69°) above the central fixation point and the other 0.9 cm below. The cued word had an arrow two spaces from each end of it, pointing toward it (e.g., > BEAD <). Participants were instructed to read the cued word aloud on each trial. The cued word appeared an equal number of times above and below the fixation point across trials. The two words presented on each trial were randomly selected from the relevant list. The screen was then blanked for 2,000 ms before the next trial began. There were 84 study trials: 72 target trials and six primacy and six recency buffer trials.

CID practice. After the study phase, participants completed 25 practice CID trials in order to familiarize themselves with the task prior to the test phase. The CID procedure used was very similar to that of Stark and McClelland (2000). On each CID trial a single word was flashed for longer and longer durations, becoming clearer over time. Participants were instructed to press the Return key as soon as they were sure that they could identify the word. Accuracy and speed were emphasized in the task instructions. At the start of each trial a fixation mask "####" was presented in 24-point Courier font for 1,000 ms. Next, the word was presented in 20-point Courier font for 16.7 ms (one screen refresh at 60 Hz). The mask was then presented for 233.3 ms, forming a 250-ms presentation block. There were thirty 250-ms presentation blocks. The stimulus duration increased by 16.7 ms on each alternate block, and the mask was always presented for the remainder of the 250-ms block. Thus, each CID trial was potentially 7,500 ms long, but each trial could be terminated prematurely by the participant pressing the Return key. When the Return key was pressed, the mask was then re-presented for 16.7 ms. Next, a white outlined box was presented that indicated to the participant that he or she must type the word on the keyboard. Key presses were displayed in the box. Participants were told to press Return after typing the word to advance to the next trial.

Test phases. Next, instructions were presented for the first CID-R test phase. Participants were told that they would again complete identification trials but that some of the words would be words from the first stage (either words they read aloud or words that were not cued, depending on the condition). They were told that they must indicate whether they thought that the word was old (one from the first stage) or new (one that had not been presented before in the experiment) after each identification.

There were 144 trials in each test phase. On each trial a word was presented with the same CID procedure as in the practice trials. Test trials were arranged into four blocks, each containing 18 old and 18 new items. This was done to evenly distribute old and new items across the test phase. There was no indication of block transition to the participant. After participants made their identification, the word was replaced by a recognition probe ("Is the word Old or New? Press O or N"). After making their judgment, a prompt was presented instructing participants to press the Return key to start the next trial.

After the first test phase was complete, instructions were presented for the next test phase. The instructions for each phase were identical except for the reference to the type of old word that would be presented (cued or uncued). Cued and uncued items were presented in different conditions because we thought that this arrangement would encourage a relatively neutral criterion placement, and therefore be the most likely arrangement to yield approximately equal frequencies of (cued and uncued) misses and false alarms in each condition. In an effort to boost general levels of motivation, participants were told in advance that they would receive feedback at the end of the experiment concerning how fast they were at pressing the Return key during the word clarification and also how accurate they were at making old or new judgments. Accordingly, participants had the opportunity to receive this feedback if they wished by pressing a button at the end of the experiment (this option was also provided in subsequent experiments).

Results

The results are presented in three parts: First, general recognition, priming, and fluency findings are given for the cued and uncued conditions; second, the application of the SS, MS1, and MS2 models to the data is described; and third, the experimental and model results relevant to Predictions 1–3 are presented.

In this and subsequent experiments, an alpha level of .05 was used for all statistical tests, and all t tests were two-tailed. We applied the Greenhouse–Geisser formula to correct for nonsphericity on repeated-measures analysis of variance tests with factors with more than two levels.

General recognition, priming, and fluency results. The CID-R trials were analyzed to provide basic measures of recognition, priming, and fluency, as detailed below.

Initial screening of CID-R trials. In this experiment and subsequent ones, a CID-R trial was not included in the analysis if a word was misidentified. If a word was misidentified at study, then it was not analyzed at test. Identification responses were corrected for minor typographical errors. The proportion of misidentified trials was low: cued condition, $M = 2.4\%$; uncued condition, $M = 2.3\%$. Trials on which the identification RT was less than 200 ms or greater than 3 standard deviations above the mean identification RT (within participant, within condition) were not analyzed (cued condition, $M = 0.56\%$ of trials; uncued condition, $M = 0.91\%$ of trials).

Recognition. Recognition discrimination performance (measured by d', the difference between the z-transformed hit and false-alarm rates) in the cued condition was significantly greater than chance (i.e., $d' = 0$), $t(31) = 13.25$, $p < .001$, as was recognition in the uncued condition, $t(31) = 2.28$, $p = .03$, even though performance in the latter was very close to chance (see Figures 3A and 3B). Recognition in the cued condition was significantly better than in the uncued condition, $t(31) = 10.87$, $p < .001$.

Priming. For each participant, the priming effect was calculated as the difference in mean identification RTs to new and old items (see Figures 4A and 4B). There was a significant priming

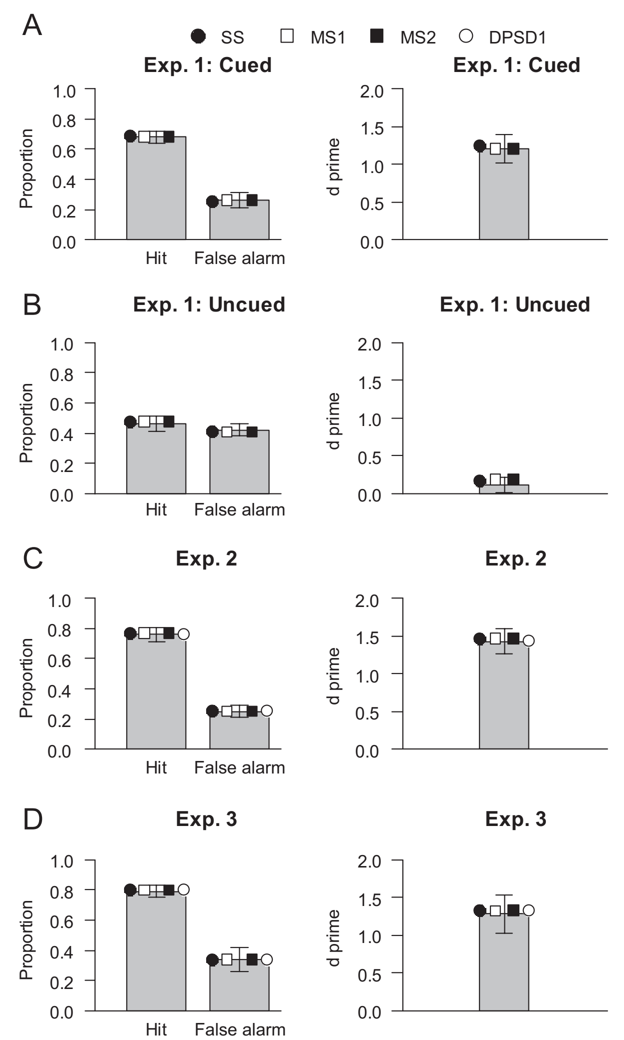

Figure 3. Recognition in Experiments 1–3 showing hit and false-alarm rates (left panels) and d' (right panels). Bars indicate experimental data (error bars indicate 95% confidence intervals of the mean), and symbols indicate the expected result from each model when fit to the data. Exp. = Experiment; SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model.

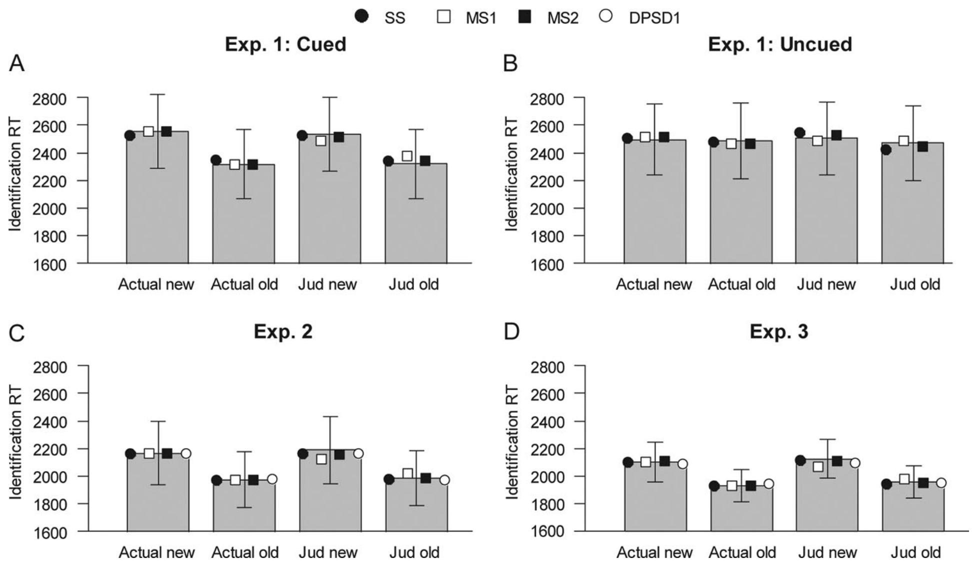

Figure 4. Identification reaction times (RTs) in Experiments 1–3. Priming is indicated by the difference in the mean identification RTs to actual new versus actual old items. The fluency effect is indicated by the shorter mean identification RTs to judged (jud) old versus judged new items. Bars indicate experimental data (error bars indicate 95% confidence intervals of the mean), and symbols indicate the expected result from each model when fit to the data. Exp. = Experiment; SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model.

effect (i.e., greater than 0 ms) in the cued condition, $t(31) = 8.97$, $p < .001$ (see Figure 5A). Indeed, priming in the cued condition was extremely robust: All 32 participants showed a priming effect. However, there was no significant priming effect in the uncued condition (see Figure 5B), $t(31) = 0.43$, $p = .67$. As with the recognition data, priming was significantly greater in the cued condition than the uncued condition, $t(31) = 7.43$, $p < .001$.

Fluency. For each participant, the fluency effect was calculated as the difference between the mean identification RT of items judged new and judged old (see Figures 4A and 4B; old and new stimuli were treated as one stimulus set). There was a significant fluency effect (i.e., reliably greater than 0 ms) in the cued condition ($M = 218$ ms, $SD = 177$), $t(31) = 6.95$, $p < .001$, but the fluency effect in the uncued condition only approached significance ($M = 36$ ms, $SD = 112$), $t(31) = 1.84$, $p = .075$. As with recognition and priming, the magnitude of the fluency effect in the cued condition was significantly greater than in the uncued condition, $t(31) = 4.69$, $p < .001$.

RTs classified by recognition response. Identification RTs were classified according to whether the item was a correct rejection, miss, false alarm, or hit (see Figures 6A and 6B). The RTs to each of the four response types significantly differed in the cued condition, $F(2.2, 69.1) = 29.64$, $p < .001$, but not in the uncued condition, $F(2.9, 89.4) = 1.12$, $p = .34$.

Model fits. The parameters of the general model in Equations 3–5 are $\mu_{r|old}$, which represents the mean difference between the old and new distributions of f_r; $\mu_{p|old}$, which represents the mean difference between the old and new distributions of f_p; σ_f, the standard deviation of the old and new distributions of f_r and f_p; σ_r, the standard deviation of the noise associated with the recognition task; σ_p, the standard deviation of the noise associated with the identification task; b, the identification RT intercept; s, the scaling parameter that represents the rate of change in the identification RT with f_p; w, the correlation between f_p and f_r; and C, the criterion of J_r that must be exceeded for an old judgment to occur.

The parameters of the models were estimated separately for every participant's data. Certain parameter values were fixed: The standard deviation of the overall distribution of the recognition strength variable, σ_{Jr}, was fixed to 1, as in standard SDT of recognition judgments. In previous applications of the model, for the sake of simplicity, we have assumed that $\sigma_f = \sigma_r$, and we make the same assumption here. Thus, $\sigma_f = \sigma_r = 1/\sqrt{2}$ (because $\sigma^2_{Jr} = \sigma^2_f + \sigma^2_r$).

Furthermore, some additional constraints were placed upon the s and w parameters across the cued and uncued conditions. In the SS model, only one value of s was estimated per participant. Thus, we assumed that the rate of change in f in the CID-R task remains constant for a given participant across conditions. Also, when fitting the MS1 and MS2 models to the data, the value of s for each participant was fixed to that of the SS model.[7] Finally, for the sake of simplicity, we assumed that there was only one value of w per participant in the MS2 model (i.e., that the correlation between

[7] Including the s parameter in all models allows us to specify the SS, MS1, and MS2 models in the same framework (s is also a free parameter for the DPSD1 model). Fixing s in the MS1 and MS2 models is necessary to obtain stable estimates of $\mu_{p|old}$. In the SS model, the value of s affects both the mean level of priming and the variance of the identification RT distributions. In the MS1 model, the value of s can be offset by varying the μ_p and σ_p parameters to fit the mean priming effect and the variance of identification RTs, respectively. To a lesser extent, this also holds for

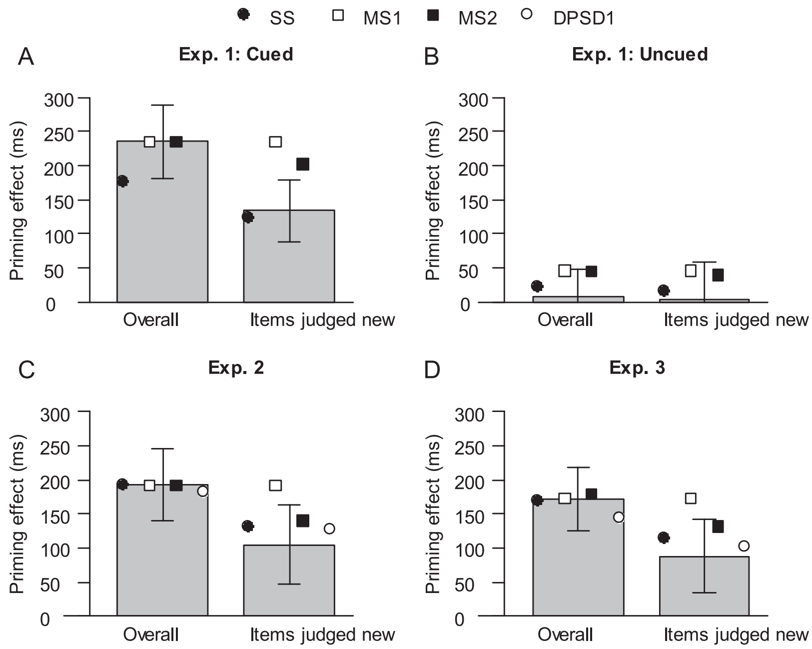

Figure 5. Overall priming versus priming for items judged new (i.e., the difference in identification reaction time to correct rejections and misses) in Experiments 1–3. Bars indicate experimental data (error bars indicate 95% confidence intervals of the mean), and symbols indicate the expected result from each model. Exp. = Experiment; SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model.

explicit and implicit item strengths is the same for cued and uncued items).

Recall that in the SS model, $\mu_{r|old} = \mu_{p|old} = \mu_{old}$ and $w = 1$ (and so $f_r = f_p$). Thus, in the SS model, there were nine free parameters per participant to fit the data of Experiment 1: μ_{old}, σ_p, b, and C for the cued condition; μ_{old}, σ_p, b, and C for the uncued condition; and s. In the MS1 model, the value of w is fixed to 0; therefore, the MS1 model has 10 free parameters for Experiment 1: $\mu_{r|old}$, $\mu_{p|old}$, σ_p, b, and C for the cued condition and $\mu_{r|old}$, $\mu_{p|old}$, σ_p, b, and C for the uncued condition. Finally, the MS2 model has 11 free parameters for Experiment 1: $\mu_{r|old}$, $\mu_{p|old}$, σ_p, b, and C for the cued condition; $\mu_{r|old}$, $\mu_{p|old}$, σ_p, b, and C for the uncued condition; and w.

The mean parameter estimates across participants are shown in Table 1, and the associated log-likelihood values (over all participants) are presented together with the AIC in Table 2. The SS model fit the data the best according to the AIC. Furthermore, the AIC measures were calculated for each individual, and Figure 7 shows the proportion of participants that were best fit by each model according to these measures. The data from the majority of participants were best fit by the SS model.

The expected recognition, priming, and fluency results for each model were determined with the equations described in Appendix C (in this and subsequent experiments). The results of the models for the basic recognition and identification RT data from Experiment 1 are shown in Figures 3A, 3B, 4A, 4B, 6A, and 6B; the overall priming effects are shown in Figures 5A and 5B. All three models fit the conventional measures in the two tasks reasonably well, and all models reproduced the basic trends in the data, that is, greater recognition, priming, and fluency in the cued condition than the uncued condition. We now turn to the results of the more specific predictions made by the models.

Model predictions. The results relevant to the specific predictions of the models are given below.

Prediction 1: Fluency effects within old and new items. The left and middle bars of Figures 8A and 8B show the fluency effects within new (i.e., RT(correct rejection) − RT(false alarm)) and old (i.e., RT(miss) − RT(hit)) stimuli. In the data of the cued condition

the MS2 model, although in this model the s parameter also interacts with the w parameter (see Appendix A) such that the extent to which the fit of the MS2 model is affected by fixing the s parameter is not as easy to discern. Importantly, though, the pattern of AIC across models is unchanged when s is allowed to be a free parameter in the MS1 and MS2 models (see Table E1). Furthermore, when the value of s was fixed (to the mean value of s across all participants in Experiments 1–3 when each participant had been fit by the SS model) in all models across all participants and experiments, the pattern of AIC still favored the SS model (except for the AIC in Experiment 2, for which the MS2 model was preferred). Finally, we also fit the models to the aggregated data across participants for each experiment, both when s was fixed (to the estimate of s when the SS model was fit to the aggregated data) and when s was free. Again, whether s was free or fixed, the SS model fit the data from Experiments 2 and 3 the best according to the AIC; however, the MS1 model fit the data from Experiment 1 the best (see Table E2). These results show that the better AIC of the SS model over the MS1 and MS2 models is largely unaffected by whether the parameter s is free or fixed, or whether the models are fit to data aggregated across individuals or on an individual participant basis.

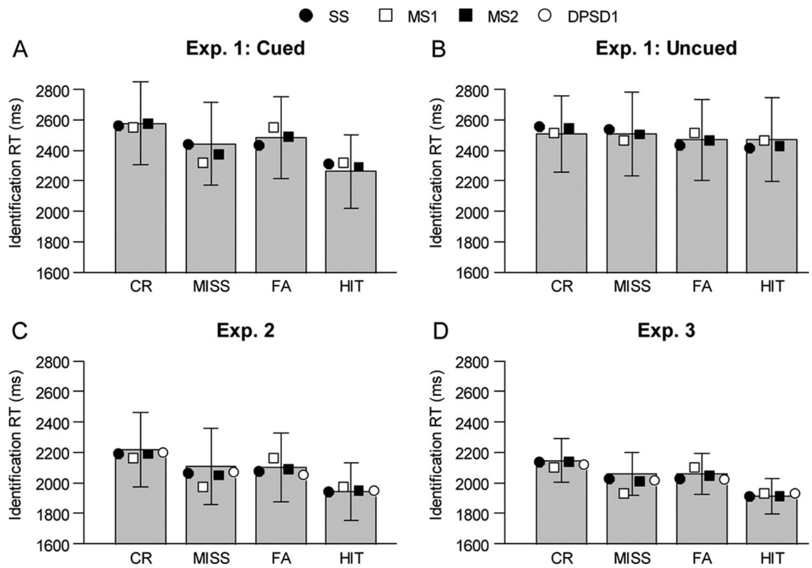

Figure 6. Identification reaction times (RTs) classified according to the recognition response (correct rejection [CR], miss, false alarm [FA], hit) in Experiments 1–3. Bars indicate experimental data (error bars indicate 95% confidence intervals of the mean), and symbols indicate the expected result from each model. Exp. = Experiment; SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model.

(Figure 8A), clear fluency effects occurred within new and old items. The identification RTs to false alarms were significantly shorter than those of correct rejections, $t(31) = 3.07$, $p = .004$, and the identification RTs to hits were significantly shorter than those of misses, $t(31) = 4.89$, $p < .001$. (There was also a significant Item (cued, new) × Judgment (old, new) interaction in the data, $F(1, 31) = 8.26$, $p = .007$, indicating that the fluency effect within old items was greater than within new items. This was not predicted by the models, was not found in subsequent experiments, and so is not commented upon further.)

The SS model correctly predicted fluency effects within new and old items (see Figure 8A) and the model results fell within the 95% confidence intervals. The MS1 model incorrectly predicted an absence of fluency effects within old and new items. Finally, the MS2 model produced a fluency effect within new and old items, but the model result for old items fell below the lower confidence interval.

Similar numerical trends to the cued condition were observed in the data of the uncued condition (see Figure 8B); however, these trends were not reliable. There was no significant difference in the identification RTs to correct rejections and false alarms, $t(31) = 1.30$, $p = .20$, or hits and misses, $t(31) = 1.28$, $p = .21$. The null findings in this condition might be expected given the nonsignificant RT-by-recognition-response one-way analysis of variance and nonsignificant fluency effects reported previously. The MS1 model predicted an absence of fluency effects within old and new items, and small fluency effects were expected under the MS2 model; both models' results fell within the confidence intervals. However, the SS model predicted fluency effects that were slightly lower than those in the cued condition (i.e., 130 ms for new items and 125 ms for old items in the cued condition vs. 121 ms for new items and 121 ms for old items in the uncued condition), and the predicted fluency effects fell outside the confidence intervals.

In sum, the presence of fluency effects within old and new items (in the cued condition, at least) was predicted by the SS model, was not predicted by the MS1 model, and was not inconsistent with the MS2 model. Fluency effects in the uncued condition were expected under the SS and MS2 models, but were not reliably observed (although the numerical trends in the data were in the correct direction).

Prediction 2: Priming for items judged new versus overall priming. The priming effect for items judged new was significant in the data from the cued condition, $t(31) = 5.90$, $p < .001$ (see Figure 5A). Moreover, as predicted by the SS model but not the MS1 model, the magnitude of this priming effect was significantly smaller than the overall priming effect, $t(31) = 4.08$, $p < .001$ (see Figure 8A), and the magnitude of the difference predicted by the SS model also fell within the confidence interval. In the MS2 model, the expected overall priming effect was greater than that of items judged new, but the model result fell below the lower confidence interval.

There was no significant priming effect for items judged new in the uncued condition, $t(31) = 0.15$, $p = .88$ (see Figure 5B). There was a numerical trend for this priming effect to be less than the overall priming effect, but the difference was not significant, $t(31) = 0.23$, $p = .82$ (see Figure 8B). Again, the nonsignificant results in this condition might be expected given that the overall priming effect was not significant to begin with.

In sum, the smaller magnitude of priming for items judged new (in the cued condition, at least) was predicted by the SS model and could be accounted for (to a lesser extent) by the MS2 model, but this difference was not predicted by the MS1 model.

Prediction 3: Overall priming in the absence of recognition. As mentioned previously, in the uncued condition, (a) discriminability in the recognition task was reliably different from, although very close to, chance levels, but (b) priming was not reliable. This suggests that the sensitivity of the recognition task is

Table 1
Means and Standard Deviations of the Parameter Values of the Models Across Participants in Experiments 1–3

Parameter	SS Experiment 1 Cued	SS Experiment 1 Uncued	SS Experiment 2 confidence	SS Experiment 3 R–K	MS1 Experiment 1 Cued	MS1 Experiment 1 Uncued	MS1 Experiment 2 confidence	MS1 Experiment 3 R–K	MS2 Experiment 1 Cued	MS2 Experiment 1 Uncued	MS2 Experiment 2 confidence	MS2 Experiment 3 R–K	DPSD1 Experiment 2 confidence	DPSD1 Experiment 3 R–K
$\mu_{r\|old}$	1.25 (0.50)	0.17 (0.18)	1.47 (0.34)	1.33 (0.48)	1.21 (0.51)	0.18 (0.20)	1.47 (0.34)	1.33 (0.50)	1.21 (0.52)	0.19 (0.20)	1.47 (0.34)	1.33 (0.50)	1.12 (0.32)	1.05 (0.34)
$\mu_{p\|old}$	$= \mu_{r\|old}$	$= \mu_{r\|old}$	$= \mu_{r\|old}$	$= \mu_{r\|old}$	1.70 (0.73)	0.36 (0.50)	1.50 (0.43)	1.34 (0.59)	1.70 (0.73)	0.35 (0.50)	1.50 (0.43)	1.38 (0.57)	$= \mu_{r\|old}$	$= \mu_{r\|old}$
w	= 1	= 1	= 1	= 1	= 0	= 0	= 0	= 0	0.60 (0.41)	0.60 (0.41)	0.78 (0.35)	0.75 (0.34)	= 1	= 1
C_1	—	—	−0.26 (0.65)	—	—	—	−0.26 (0.65)	—	—	—	−0.27 (0.65)	—	−0.29 (0.66)	—
C_2	—	—	0.32 (0.42)	—	—	—	0.33 (0.42)	—	—	—	0.32 (0.42)	—	0.31 (0.44)	—
$C_3/C/C_G$	0.72 (0.47)	0.25 (0.34)	0.69 (0.25)	0.45 (0.47)	0.70 (0.47)	0.25 (0.34)	0.69 (0.25)	0.46 (0.47)	0.70 (0.47)	0.25 (0.34)	0.69 (0.25)	0.46 (0.47)	0.69 (0.26)	0.57 (0.44)
C_4/C_K	—	—	1.06 (0.28)	0.99 (0.47)	—	—	1.06 (0.28)	0.99 (0.47)	—	—	1.06 (0.28)	0.99 (0.47)	1.11 (0.31)	1.27 (0.48)
C_5/C_R	—	—	1.54 (0.35)	1.71 (0.48)	—	—	1.54 (0.35)	1.71 (0.48)	—	—	1.54 (0.36)	1.71 (0.48)	1.77 (0.48)	—
$P(\text{Rc}\|\text{old})$	—	—	—	—	—	—	—	—	—	—	—	—	0.29 (0.17)	0.37 (0.18)
$P(\text{Rc}\|\text{new})$	—	—	—	—	—	—	—	—	—	—	—	—	—	0.06 (0.08)
b	2,529 (741)	2,504 (737)	2,166 (440)	2,101 (296)	2,552 (746)	2,511 (732)	2,165 (435)	2,103 (287)	2,552 (745)	2,510 (733)	2,165 (434)	2,109 (292)	2,160 (436)	2,088 (295)
s	150 (96)	150 (96)	138 (78)	131 (70)	= SS	= SS	= SS	= SS	= SS	= SS	= SS	= SS	168 (103)	128 (82)
σ_p	566 (161)	590 (154)	502 (119)	452 (93)	565 (160)	585 (153)	502 (119)	452 (95)	563 (159)	587 (155)	500 (119)	450 (93)	497 (115)	454 (93)

Note. In Experiment 1, C_3 corresponds to C, and in Experiment 3, C_3, C_4, and C_5 correspond to C_G, C_K, and C_R, respectively. Standard deviations are shown in parentheses. A dash indicates that the parameter was not used to model the data. A value preceded by an equal sign indicates that the value was fixed. The parameters σ_f and σ_r are not shown, but in all models their value was fixed to $\sigma_f = \sigma_r = 1/\sqrt{2}$. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model; R–K = remember–know.

greater than that of the priming task, a finding that is consistent with previous ones (e.g., Berry, Henson, & Shanks, 2006). Thus, there was no evidence that overall priming occurred in the absence of recognition, a finding that would be evidence against the SS model and would favor the MS1 or MS2 models.

Discussion

There were two main positive results from Experiment 1 that were predicted by the SS but not the MS1 model; the MS2 model also produced these results: In the cued condition, (a) identification RTs to items judged old were shorter than those of items judged new within both old and new items (Prediction 1), and (b) the magnitude of the priming effect for items that were not recognized was smaller than the magnitude of the overall priming effect (Prediction 2). Although the corresponding results in the uncued condition were not as clear-cut, there were numerical trends in the same direction; that is, there were trends for fluency effects within old and new items, and for priming for items judged new to be smaller than the overall priming effect (the reason that these trends did not reach significance is likely to be low power, because the memory representation for uncued words was extremely weak, as indicated by the very low level of recognition discriminability and the small and nonsignificant overall priming and fluency effects in this condition).

The result concerning Prediction 3—whether priming would occur in the absence of recognition—did not discriminate between the models. Because the MS1 and MS2 models have functionally independent memory signals for priming and recognition, priming in the absence of recognition can be produced: an empirical claim that has been made previously but that does not in fact seem to hold when the conditions and measures of priming and recognition are closely matched (Berry, Shanks, & Henson, 2006). In the present uncued condition, where recognition was very weak, there was again no evidence of priming. This is at least consistent with the SS model, which predicts that priming will not occur in the (true) absence of recognition.

Finally, when the models were fit at the level of individual participants, the AIC favored the SS model over the MS1 and MS2 models. Indeed, the majority of participants were best fit by the SS model.

Experiment 2

In Experiment 2, the predictions of the models were extended to a CID-R task with 6-point recognition confidence judgments (from 1 = *most confident a test item was not studied* to 6 = *most confident that a test item was studied*). According to SDT, participants assign a confidence rating to an item by comparing its strength-of-evidence value against multiple decision criteria (see, e.g., Macmillan & Creelman, 2005). With six possible confidence ratings, five criteria are necessary to model the confidence ratings (i.e., C_1–C_5). If confidence ratings are modeled in the SS, MS1, and MS2 models as they are in SDT, it is possible to discern another prediction of the models.

> *Prediction 4:* The SS model predicts that identification RTs will decrease as recognition confidence increases, within both old and new items. The MS1 model predicts that identifica-

Table 2
Goodness-of-Fit Values for the Models in Experiments 1–3

	SS			MS1			MS2			DPSD1		
Experiment	*p*	ln(*L*)	AIC	*p*	ln(*L*)	AIC	*p*	ln(*L*)	AIC	*p*	ln(*L*)	AIC
1 (N = 32)	9	−74738	**150053**	10	−74737	150114	11	−74698	150101	—	—	—
2 (N = 16)	9	−41949	84187	9	−41987	84261	10	−41943	84206	10	−41918	**84156**
3 (N = 18)	7	−30203	**60658**	7	−30231	60714	8	−30196	60679	8	−30217	60721
Pooled	9	−156731.3	**313480.6**	9	−156770.2	313558.3	10	−156730.9	313481.8	—	—	—

Note. The Akaike information criterion (AIC) is calculated as AIC = $-2\ln(L) + 2P$, where $P = p \times N$ is the total number of free parameters for each fit, where p is the number of free parameters for each model, and N is the number of participants modeled in each experiment. A smaller AIC value indicates a relatively better model. A dash indicates that the model was not applied to the data. The ln(L) value for Experiment 1 is based upon fits to the cued and uncued conditions. Bold indicates that the model fit the data best according to the AIC measure. The pooled row shows the results of fitting the models to the data that had been pooled from all experiments, and when the parameters were fixed across participants and experiments. The total number of data points fit in each experiment was as follows: Experiment 1 = 8,932; Experiment 2 = 4,656; Experiment 3 = 3,519. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model; L = maximum likelihood.

tion RTs and recognition confidence ratings are not related. The MS2 model can produce either pattern.[8]

According to the SS model, because the same value of f is used to calculate an item's recognition judgment and its identification RT, an item that elicits a high value of f is likely to end up with a relatively high confidence rating and a relatively short identification RT. It predicts this regardless of whether the item is old or new (because $\rho(J_r, \text{RT}) < 0$ in Equation 5). In the MS1 model, on the other hand, an item that receives a high confidence rating will not necessarily have a comparably short identification RT (see Prediction 2), and so the MS1 model predicts that identification RTs will not vary with confidence ratings, and predicts this, regardless of whether the items are old or new.[9]

Furthermore, the recognition confidence ratings acquired in this experiment allowed the DPSD1 model (as described in the introduction) to be fit to the data and hence compared with the other models. In the DPSD1 model, recognition judgments and identification RTs are modeled in the same way as they are in the SS model (thus, in terms of the general model in Equations 3–5, w = 1 and $\mu_{r|old} = \mu_{p|old} = \mu_{old}$), except that an item can be recollected with probability $P(\text{Rc}|I)$, where Rc stands for recollection and I stands for item type, where I = old, new.

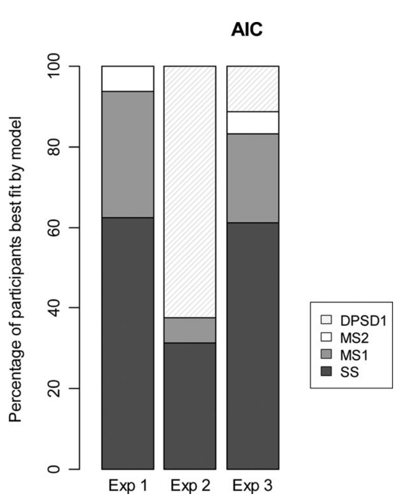

Figure 7. Model selection results. Each bar represents the proportion of participants best fit by each model (according to the Akaike information criterion [AIC]) in Experiments 1–3. Exp = Experiment; SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model.

The DPSD1 model predicts a similar trend to the SS and MS2 models regarding Prediction 4: Identification RTs will generally tend to decrease as confidence increases. However, because of the additional recollection process, the precise prediction of the DPSD1 model differs slightly for the highest confidence rating. According to dual-process theory (Yonelinas, 1994), recollection will normally result in the highest confidence rating (e.g., 6). In the DPSD1 model that we implemented,[10] recollected items do not necessarily have high values of f (because the probability of recollection is independent of f). Instead, the mean f of recollected items will tend to be equal to μ_I, and the identification RTs of these items will therefore tend to be equal to the mean identification RT of old or new items. Because a portion of 6-rated items are recollected, the overall mean identification RT of 6-rated items will not be as short as it would be had the 6-rated items arisen

[8] Prediction 4 (and also Prediction 5; see ahead to Experiment 3) can be viewed as an extension of Prediction 1 to a situation where there are multiple decision criteria rather than one criterion.

[9] Stark and McClelland (2000) also collected (4-point) confidence ratings after each CID trial in their Experiments 2–5. However, they did not compare the mean identification RTs when classified according to these ratings because there were insufficient frequencies of responses at each confidence level. Accordingly, in our Experiment 2 we use a much larger number of old and new word stimuli at test than Stark and McClelland (i.e., 150 in Experiment 2 vs. 30–60 in the experiments of Stark and McClelland).

[10] The DPSD1 model is just one possible instantiation of dual-process theory. Another possible instantiation would allow recollection (cued by partial stimulus information during clarification of the word) to affect the identification RTs; that is, it would include another arrow from $P(\text{Rc}|I)$ to RT in Figure 1D. This would imply that the identification RTs for recollected items are modeled with a separate distribution (e.g., a normal distribution with two extra free parameters for the mean and variance). To make the DPSD1 model not too complex, we did not build this assumption into the model. Further consideration is given to this model and other possible implementations in the Discussion of Experiment 3.

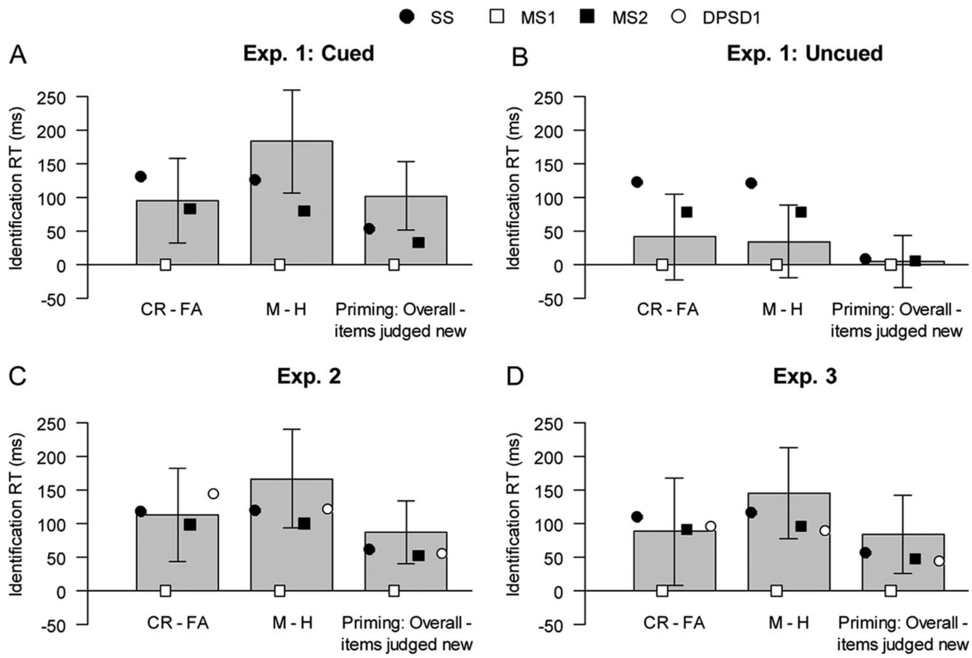

Figure 8. Data and model results concerning Predictions 1 and 2. Correct rejection–false alarm (CR–FA) refers to the fluency effect within new items (Prediction 1), miss–hit (M–H) refers to the fluency effect within old items (Prediction 1), and "Priming: Overall − items judged new" refers to the difference in the overall priming effect and the priming effect for items judged new (Prediction 2). Bars denote data, and symbols indicate the expected result from each model. Error bars denote 95% confidence interval of the difference in means. Exp. = Experiment; SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model; RT = reaction time.

solely from high familiarity. Thus, similar to the SS model, the DPSD1 model predicts that identification RTs will decrease with increasing confidence, but it predicts that the mean identification RT for highest confidence ratings will not be as short as that predicted by the SS model. (A more qualitative test of the DPSD1 model is presented in Experiment 3.)

Thus, Experiment 2 was similar to Experiment 1 but with two changes to allow Prediction 4 to be tested: (a) recognition judgments were measured with a 6-point confidence rating scale, rather than with binary old–new judgments, and (b) all items at encoding were presented in a CID procedure and there was no manipulation of attention. The design of Experiment 2 also allowed Predictions 1–3 to be tested again by collapsing all the new (Ratings 1–3) and all the old judgment ratings together (Ratings 4–6), in addition to allowing the fit of a fourth model, the DPSD1 model, to be determined.

Method

Participants. Twenty individuals (10 female, 10 male) were recruited through a University College London participant database. Their ages ranged from 19 to 35 years with a mean of 22.4 years. Each received £5 in return for participation.

Materials. The same 337 four-letter words that were used in Experiment 1 were stimuli in this experiment. For each participant, 150 words were randomly selected from this pool to be the old stimuli, another 150 were selected to be the new stimuli, and a further 10 were selected to be stimuli on primacy and recency trials.

Procedure. Participants completed a study phase and then a test phase. Each word at study and test was presented via the CID procedure, as described in Experiment 1. There were 160 study trials. The first and last five trials were considered primacy and recency trials, and none of the words from these trials appeared at test. In the test phase, participants were told that they must indicate whether they thought the word was old or new after each identification by using a 6-point confidence scale where 1 = *sure new*, 2 = *probably new*, 3 = *guess new*, 4 = *guess old*, 5 = *probably old*, 6 = *sure old*. After participants made their identification, a recognition probe was presented ("Is the word Old or New?"), and the numbers 1–6 were presented with the appropriate label (above). After making their judgment, a prompt was presented instructing participants to press the Return key to start the next trial. Test trials were arranged into four blocks, each containing an equal number of old and new items. There was no indication of block transition to the participant.

Results

General recognition, priming, and fluency results. The basic recognition, priming, and fluency results are given below.

Initial screening of CID-R trials. Few trials were misidentified: study phase, $M = 1.6\%$, $SD = 1.6$; test phase, $M = 2.3\%$, $SD = 2.0$. Of the remaining trials, $M = 0.75\%$ had identification

RTs that were less than 200 ms or greater than 3 standard deviations above the mean identification RT and were not analyzed. Four participants failed to make use of all confidence rating response options (within new and old stimuli). This meant that their identification RT data classified according to rating could not be fully analyzed. The data from these participants were therefore not included in any subsequent analysis, although the inclusion of these participants did not change the qualitative pattern of results (where their inclusion was possible).

Recognition, priming, and fluency. Recognition discrimination performance was significantly greater than chance (see Figure 3C), $t(15) = 18.31$, $p < .001$.[11] There was a significant overall priming effect (i.e., greater than 0 ms; see Figure 5C), $t(15) = 7.72$, $p < .001$. There was also a significant fluency effect (i.e., reliably greater than 0 ms; $M = 206$ ms, $SD = 116$), $t(15) = 7.07$, $p < .001$.

RTs classified by recognition response. Identification RTs were classified according to whether the item was a correct rejection, miss, false alarm, or hit. The RTs to each of the four response types differed significantly (see Figure 6C), $F(2.5, 37.1) = 22.12$, $p < .001$.

Modeling. Five criteria (C_1–C_5) were required to model the six confidence ratings in Experiment 2. Thus, the SS model had nine free parameters per participant: s, μ_{old}, σ_{p}, b, C_1, C_2, C_3, C_4, C_5. The MS1 model also had nine: $\mu_{\text{r|old}}$, $\mu_{\text{p|old}}$, σ_{p}, b, C_1, C_2, C_3, C_4, C_5. The MS2 model had 10: $\mu_{\text{r|old}}$, $\mu_{\text{p|old}}$, σ_{p}, b, w, C_1, C_2, C_3, C_4, C_5. In fitting the DPSD1 model to this data set, 10 free parameters were required: s, μ_{old}, σ_{p}, b, C_1, C_2, C_3, C_4, C_5, and $P(\text{Rc}|\text{old})$. The likelihood functions given in Appendix A were used to obtain the maximum likelihood estimates of the parameters of the DPSD1 model in a similar fashion to the other models. Identification RTs in the DPSD1 model are based on f (as in Equation 2), and recognition responses in the DPSD1 model are based on either recollection or f. If an item is not recollected, it receives a 1–6 confidence rating by comparing J_{r} with C_1–C_5.

The mean and standard deviation of the parameter estimates are shown in Table 1, and the expected model results are shown in Figures 3–6 and Figures 8 and 9. The parameter estimate for w in the MS2 model was relatively high. The associated AIC values in Table 2 indicate that, as in Experiment 1, the SS model should be preferred over the MS1 and MS2 models. The qualitative pattern of results predicted by the DPSD1 model was very similar to that of the SS and MS2 models. As shown in Table 2 and Figure 7, the DPSD1 model fit the data better according to the AIC.

The expected recognition, identification RT, and priming results from the models are shown in Figures 3C, 4C, 5C (left bar), and 6C. (Note that when we calculated the expected hit and false-alarm rates, we assumed that $C_3 = C$ for each model.) All models gave close, comparable fits to these data. We now turn to the results of the specific predictions made by the models.

Predictions 1 and 2. Similar outcomes to Experiment 1 were obtained in Experiment 2 with regard to Predictions 1–3. First, fluency effects occurred within old and new items, as predicted by the SS, MS2, and DPSD1 models but not the MS1 model (Prediction 1): The RTs to false alarms were significantly shorter than those of correct rejections (see Figure 8C), $t(15) = 3.44$, $p = .004$, and the identification RTs to hits were significantly shorter than those of misses, $t(15) = 4.82$, $p < .001$. (The Item (old, new) × Judgment (old, new) interaction was not reliable, $F(1, 15) = 2.58$, $p = .13$.) Second, the priming effect for items judged new was significant, $t(15) = 3.78$, $p < .001$ (see Figure 5C), and as predicted by the SS and DPSD1 models but not the MS1 model, the magnitude of this priming effect was significantly smaller than the overall priming effect (Prediction 2), $t(15) = 3.97$, $p = .001$ (see Figure 8C).

Prediction 4: Identification RTs classified according to confidence rating. The SS, MS2, and DPSD1 models predict that identification RTs decrease as recognition confidence increases within both old and new stimuli (see Figure 9A). In contrast, the MS1 model predicts that identification RTs should not vary with confidence within old and new items. The relevant data from Experiment 2 are shown in the leftmost panel of Figure 9A. Although the data for the guess responses (3 and 4) appear to be somewhat variable, a linear trend analysis confirmed that identification RTs significantly decreased as recognition confidence increased within old and new items: old items, $F(1, 15) = 13.61$, $p = .002$; new items, $F(1, 15) = 23.73$, $p < .001$. No other higher order trend components were significant within either type of stimulus (all Fs < 2.3). The interaction between old and new linear trends was not significant, $F(1, 15) = 2.7$, $p = .12$. This confirms Prediction 4 of the SS and MS2 models.

The SS, MS2, and DPSD1 models all predicted that identification RTs for items with the highest confidence rating (6) would be lower than those at the next highest level (5). This difference was reliable within new items, $t(15) = 2.14$, $p = .049$, but not old items, $t(15) = 0.74$, $p = .47$. Finally, given that $P(\text{Rc}|\text{old})$ is greater than $P(\text{Rc}|\text{new})$ ($= 0$) in the DPSD1 model, it predicts an interaction between old–new items and the 5–6 confidence ratings

[11] The collection of confidence ratings also allowed us to plot receiver operator characteristic (ROC) curves for the recognition data. ROC curves plot the hit rate and false-alarm rate at each level of confidence. The analysis of ROC curves has featured prominently in the testing of different models of recognition memory. For example, traditional SDT models of recognition, in which the variances of the old and new item strength distributions are assumed to be equal, predict that ROCs will be curvilinear in shape and that the z-transformed ROC (zROC) will have a slope equal to 1. Because $\sigma_{J_r}(\text{new}) = \sigma_{J_r}(\text{old})$ in the SS, MS1, and MS2 models, they all predict that the slope of the zROC is equal to 1. However, the slope of the zROC in Experiment 2 was calculated as 0.75, which suggests that the equal-variance assumption in the models is false. Unequal-variance signal detection theory (UVSDT) accounts for such a finding by assuming that the variance of the old item strength distribution is greater than that of the new item distribution. Incorporating an UVSDT assumption into the SS model (e.g., by fixing $\sigma_{J_r}(\text{new}) = 0.75 \times \sigma_{J_r}(\text{old})$, which is achieved by setting σ_f (new) to 0.25 in the SS, MS1, and MS2 models) did not alter any of the SS, MS1, or MS2 model results regarding Predictions 1–5 in this study. (This was determined via the method of simulation: With the exception of σ_f (new), the parameter values used to simulate data for each participant were the same as those estimated with maximum likelihood procedures; see Table 1.) Interestingly, with this unequal-variance assumption, the SS and MS2 models now predict that the magnitude of the fluency effects for old words is slightly greater than that of new words. This trend was generally observed (see Figures 8 and 16), and there is some indication that this was actually the case in the cued condition of Experiment 1, at least (i.e., the Item × Judgment interaction was significant in Experiment 1). In future investigations it would be interesting to extend the modeling framework to allow for this unequal-variance assumption and to compare the fit of this model with the others more formally. The dual-process model of recognition (Yonelinas, 1994) also predicts that the slope of the ROC will be less than 1. We present a specific test of this model in Experiment 3.

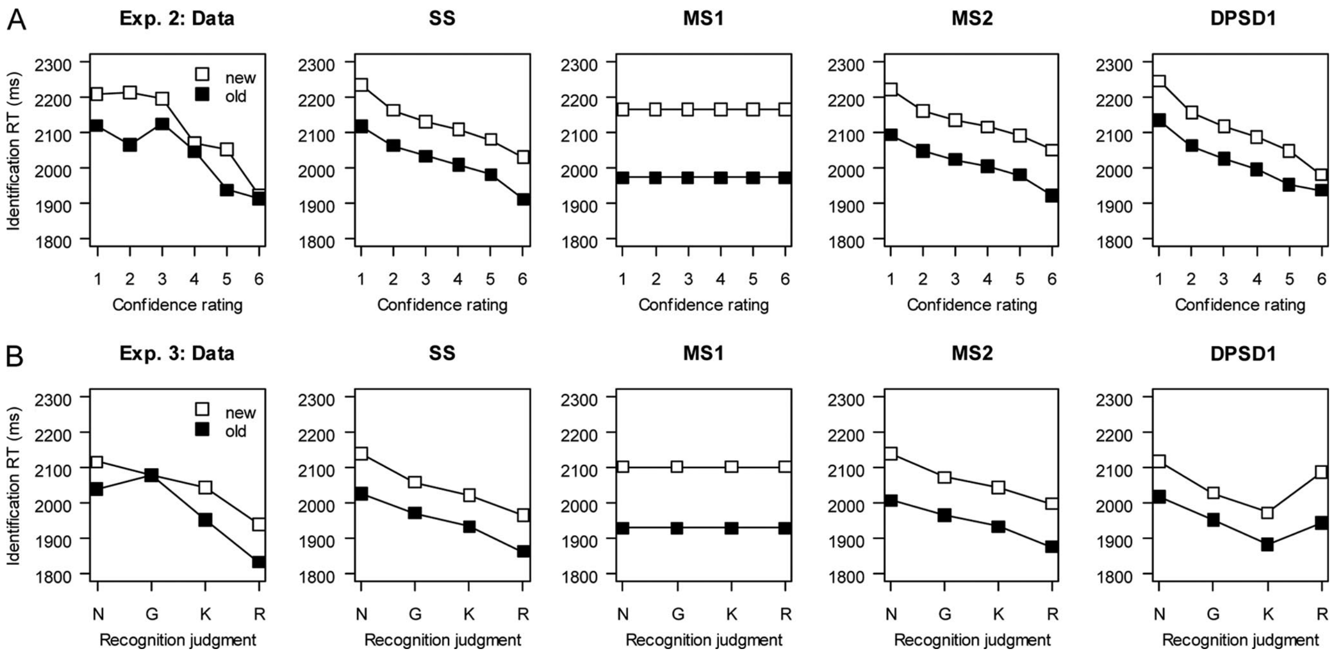

Figure 9. Mean identification reaction times (RTs) classified according to the recognition response in Experiments 2 and 3. Data from each experiment are shown along with single-system (SS), multiple-systems-1 (MS1), multiple-systems-2 (MS2), and dual-process signal detection (DPSD1) model results. (A) Numbers 1–6 refer to the six possible recognition confidence ratings in Experiment 2. (B) N, G, K, and R refer to the four possible types of recognition response in Experiment 3, where N = "new," G = "guess," K = "know," R = "remember."

(see Figure 9A); though a trend, this interaction did not reach significance, $F(1, 15) = 2.57$, $p = .13$.

Discussion

Two results from Experiment 2 were predicted by the SS, MS2, and DPSD1 models but not the MS1 model: (a) identification RTs decreased as recognition confidence increased within both old and new items (Prediction 4; also, as in Experiment 1, fluency effects occurred within both old and new items: Prediction 1), and (b) as in Experiment 1, the magnitude of the priming effect for items judged new was smaller than the overall priming effect (Prediction 2). Finally, the SS model fit the data better than the MS1 and MS2 models according to the AIC, but the DPSD1 model fit the data better than the SS model.

One trend that was predicted by the SS model may be counterintuitive and requires explanation. As is evident in Figure 9A, the SS model predicts that identification RTs to old items will be shorter than those of new items at each confidence level. However, it might be expected that the SS model would predict that identification RTs to old and new items will be the same at each rating because the value of J_r that must be exceeded for a given confidence rating to occur is fixed and does not depend upon whether an item is old or new.

The reason for this seemingly counterintuitive prediction of the SS model is subtle: It arises because *f* is combined with random noise (e_r) for the generation of J_r (in Equation 1), and then *f* is combined with another uncorrelated source of noise (e_p) for the generation of the identification RT (in Equation 2). Consider a new item that is assigned a relatively high confidence rating of 5. To receive this rating, the value of *f* of an "average" new item needs to be combined with a relatively high noise value (e_r) to offset its relatively low value of *f*, and end up with a J_r value greater than C_4. When the identification RT is then generated for the item by combining *f* with e_p, e_p is unlikely to be as extreme in its value as e_r because e_p is not correlated with e_r. This new item will therefore have a relatively high recognition confidence rating, but its identification RT will be more similar to that of all new items. In contrast, an old item that is assigned a rating of 5 is unlikely to require as high a value of e_r because its value of *f* is already relatively high to begin with. This old item's identification RT is likely to be appropriately short (for an old item) after combining *f* with e_p—shorter than that of the new item's identification RT. Thus, the process of combining a single value of *f* with uncorrelated sources of noise produces this counterintuitive trend in the SS model. (This is the same reason for the variable relationship between RT(miss) and RT(false alarm) in Footnote 5.) The noise parameters are crucial for the SS model to produce this trend because without them the model would predict that identification RTs to old and new items will be the same at each confidence level (see also Shanks, Wilkinson, & Channon, 2003, for another example and further discussion).

Experiment 3

In Experiment 3, the predictions of the models were extended to a CID-R paradigm with remember–know judgments (Gardiner, 1988; Tulving, 1985), another widely used method of measuring recognition. In this type of recognition procedure, participants are instructed to respond "remember" (R) if they can recollect specific contextual information relating to an item's presentation in the study phase at the time of retrieval (e.g., "I can remember seeing that word . . . it was the first one on the list"), and they are asked to respond "know" (K) if they believe an item was presented in the study phase, but is only familiar, in the absence of any specific contextual recollection.

R and K judgments have been hypothesized to be driven by independent recollection and familiarity memory processes (e.g., as proposed in the dual-process model of recognition; Yonelinas, 1994, 2002). In the dual-process model, the familiarity process is assumed to drive K responses and is characterized by a standard equal-variance signal detection process with a single strength-of-evidence axis. Recollection is assumed to drive R judgments, and is assumed to be a distinct memory process in the sense that it is probabilistic: An item is either recollected with a fixed probability or not recollected at all. If an item is recollected, it receives an R response (or a high confidence response). If an item is not recollected, then the recognition judgment for that item is based on familiarity.

According to the SDT interpretation of R–K judgments, R and K judgments index high and low strength of evidence, rather than distinct memory processes per se. The SDT interpretation says that R–K judgments can be modeled in a similar manner to confidence ratings: There is a relatively high criterion for R judgments (C_R) and a relatively low criterion for K judgments (C_K). Items will receive an R response if their strength value exceeds C_R, and items will receive a K response if their strength value exceeds C_K but falls below C_R (Donaldson, 1996; Dunn, 2004; Wixted & Stretch, 2004). Thus, by this alternative account, R–K judgments are driven by a single (unidimensional) memory strength signal. If R–K judgments are modeled in the SS, MS1, and MS2 models as they are in SDT, then this would lead the models to make distinct predictions, similar to Prediction 4.

> *Prediction 5:* Within old and new items, the SS model predicts that identification RTs classified according to the recognition judgment will be ordered (from longest to shortest) "new" (N), K, R. The MS1 model predicts that identification RTs are unrelated to recognition judgments. The MS2 model can produce either result.

Furthermore, this experiment allows a specific prediction of the DPSD1 model to be tested. If recollection is independent of familiarity (f in the model) as dual-process theory of recognition proposes, then a recollected item (i.e., an item receiving an R judgment) will not necessarily have a high value of f. A recollected item's value of f is most likely to be equal to the mean f of the distribution it was drawn from (μ_f), and our instantiation of the DPSD1 model would predict that the identification RT of an old item assigned an R response is equal to the mean identification RT of all old items. Thus the DPSD1 model makes some specific predictions: (a) like the SS model, but unlike the MS1 model, it predicts (a priori) that identification RTs for K responses will be shorter than for N judgments (since the DPSD1 and SS models assume a single underlying f distribution contributing to these judgments), and (b) unlike the SS model, but like the MS1 model, it does not predict shorter identification RTs for R than K judgments (in fact, the DPSD1 model predicts longer identification RTs for R than K judgments to old items).

To test Prediction 5, the procedure in Experiment 3 was the same as in Experiment 2 except that participants were asked to indicate their recognition response by responding N if they thought that the word had not been presented before, "guess" (G) if they thought the item was old but had no recollection or familiarity associated with the item, K if they thought the item was familiar but not recollected, and R if they recollected the item. The G option was included following the recommendation of others (e.g., Gardiner & Richardson-Klavehn, 2000) who have suggested that without the G option, K responses can also include random guesses, which can dilute the process purity of these (hypothesized familiarity-based) responses. According to SDT, on the other hand, the G response is simply accommodated by an additional criterion that must be exceeded in order for a G response to occur (C_G; i.e., N, G, K, and R responses reflect responses made to items with low to high strength of evidence). The design of Experiment 3 also allowed us to test Predictions 1–3 of the SS, MS1, and MS2 models again by collapsing across G, K, and R responses for "old" judgments.

Method

Participants. Twenty individuals (14 female, six male) were recruited. Their ages ranged from 19 to 25 years with a mean of 20.4 years.

Materials. The same 337 four-letter words that were used in Experiment 1 were stimuli in this experiment. For each participant, 105 words were randomly selected from this pool to be the old stimuli, another 105 were selected to be the new stimuli, and a further 10 were selected to be stimuli on primacy and recency trials.

Procedure. Each word at study and test was presented via the CID procedure, as described in Experiment 1. There were 115 study trials. The first and last five trials were considered primacy and recency trials, and none of the words from these trials appeared at test. In the test phase, participants were additionally told that they must indicate whether they thought that the word was old or new after each identification by pressing the numbers 1–4 where 1 = *remember old*, 2 = *know old*, 3 = *guess old*, 4 = *new*. The instructions for using each of these responses were adapted for the current paradigm from Gardiner and Richardson-Klavehn (2000; see Appendix D). After participants made their identification, a recognition probe was presented ("Is the word Old or New?"), and the numbers 1–4 were presented with the appropriate label (above). After making their judgment, a prompt was presented instructing participants to press the Return key to start the next trial. Test trials were arranged into four blocks, each containing an equal number of old and new items. There was no indication of block transition to the participant.

Results

General recognition, priming, and fluency results. The basic recognition, priming, and fluency results are given below.

Initial screening of CID-R trials. In the study phase, M = 3.4% (SD = 2.4) of trials were misidentified, and in the test phase, M = 3.8% (SD = 3.2) of trials were misidentified. Of the remaining trials, M = 1.3% had identification RTs that were less than 200 ms or greater than 3 standard deviations above the mean identification RT and were not analyzed. Four participants failed to make any responses in at least one of the response options for new and old stimuli, meaning that their identification RT data classified according to the judgment could not be fully analyzed. The data from these participants were not included in any subsequent analysis, although the inclusion of these participants did not change the qualitative pattern of results (where their inclusion was possible; one exception to this is noted below).

Recognition, priming, and fluency. Recognition discrimination performance was significantly greater than chance (see Figure 3D), $t(17) = 10.99$, $p < .001$. The proportion of R, K, G, and N responses to old and new stimuli are shown in Figure 10. There was a significant overall priming effect (i.e., greater than 0 ms; see Figure 5D), $t(17) = 7.92$, $p < .01$. There was also a significant fluency effect (i.e., greater than 0 ms; $M = 168$ ms, $SD = 136$), $t(17) = 5.26$, $p < .001$.

RTs classified by recognition response. Identification RTs were classified according to whether the item was classified as a correct rejection, miss, false alarm, or hit (by initially classifying items that received G, K, or R responses as items judged old). The RTs to each of these four response types significantly differed, $F(2.2, 37.7) = 18.34$, $p < .001$ (see Figure 6D).

Modeling. The parameters of the SS, MS1, and MS2 models were the same as in Experiment 2 except that three decision criteria were required to model the N, G, K, and R responses: C_G, C_K, and C_R. Thus, for this data set, the SS model had seven free parameters per participant: s, μ_{old}, σ_p, b, C_G, C_K, C_R. The MS1 model had seven: $\mu_{r|old}$, $\mu_{p|old}$, σ_p, b, C_G, C_K, C_R. The MS2 model had eight: $\mu_{r|old}$, $\mu_{p|old}$, σ_p, b, w, C_G, C_K, C_R. As in Experiment 2, we assumed that the DPSD1 model is essentially the same as the SS model but with the addition of a recollection process to model R judgments. There is no criterion for the R response (C_R) in the DPSD1 model, and so eight free parameters per participant are required by the DPSD1 model for Experiment 3: s, μ_{old}, σ_p, b, C_G, C_K, $P(\text{Rc}|\text{old})$, and $P(\text{Rc}|\text{new})$.[12]

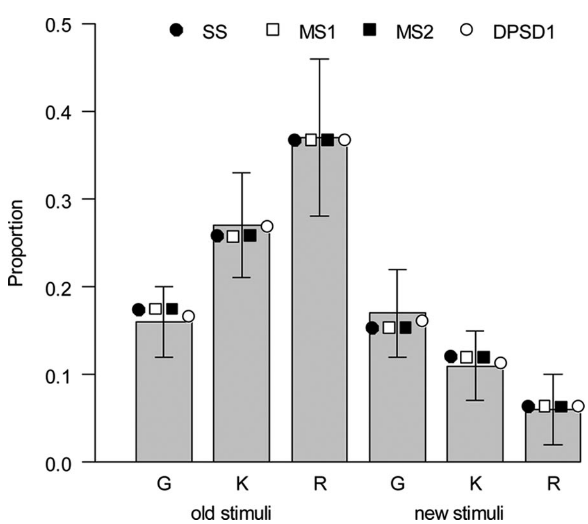

Figure 10. Mean proportion of G, K, and R judgments in Experiment 3, where G = "guess," K = "know," R = "remember." Bars denote data, and symbols indicate the expected result from each model. Error bars indicate 95% confidence intervals of the mean. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model.

The mean and standard deviation of the estimates of the parameters across participants are given in Table 1. The measures of fit are given in Table 2 and Figure 7 and indicate that the SS model produced better fits to the data of the majority of participants than the MS1, MS2, and DPSD1 models according to the AIC. The expected data for the overall recognition, identification RT, priming, and fluency results of the models are shown in Figures 3D, 4D, 5D (left bar), 6D, and 10. All models gave close, comparable fits to the basic recognition, priming, and fluency data.

Model predictions. The results relevant to the specific predictions of the models are given below.

Predictions 1 and 2. The results concerning Predictions 1 and 2 were largely similar to those of Experiments 1 and 2. As predicted by the SS, MS2, and DPSD1 models, but not the MS1 model, the identification RTs to items judged old were significantly shorter than to those judged new within old and new stimuli (Prediction 1; see Figure 8D): The identification RTs to hits were significantly shorter than those of misses, $t(17) = 4.56$, $p < .001$, and RTs to false alarms were significantly shorter than those of correct rejections, $t(17) = 2.33$, $p = .03$; however, this latter comparison was not reliable when the data from the four excluded subjects were included, $t(21) = 1.70$, $p = .105$. (There was no significant Item (old, new) × Judgment (old, new) interaction in the data, $F(1, 17) = 2.69$, $p = .12$.) Second, as predicted by the SS, MS2, and DPSD1 models, but not the MS1 model, the magnitude of the priming effect for items judged new was significantly smaller than the magnitude of the overall priming effect (Prediction 2), $t(17) = 3.03$, $p = .008$ (see Figure 8D).

Prediction 5: Identification RTs classified according to R–K response. As shown in Figure 9B, the SS and MS2 models predict that identification RTs classified according to the recognition response would be ordered (from longest to shortest) N, G, K, R, within both old and new stimuli. In contrast, the MS1 model did not predict any difference between the identification RTs when classified according to the recognition judgment within old and new stimuli.

The data are shown in the leftmost panel of Figure 9B and confirm the trends predicted by the SS and MS2 models. There was a significant linear trend for identification RTs to decrease across N, G, K, and R judgments within old items, $F(1, 17) = 42.36$, $p < .001$; the linear trend for new items was marginally significant, $F(1, 17) = 4.11$, $p = .059$. No other higher order trend components were significant within either type of stimulus (all Fs < 1.04), except for an unexpected significant quadratic trend within old items, $F(1, 17) = 5.36$, $p = .03$. The interaction between old and new linear trends was not significant ($F < 1$).

Finally, the identification RTs to old items receiving the R response were significantly shorter than those of old items receiving K responses, $t(17) = 2.95$, $p = .009$, and were also signifi-

[12] Dual-process theory of recognition does not include a parameter analogous to $P(\text{Rc}|\text{new})$ because recollection is hypothesized to occur only for old items. The $P(\text{Rc}|\text{new})$ parameter was included in the DPSD1 model to allow trials on which new items received an R response to be included in the parameter estimation procedures (see Appendix A). If $P(\text{Rc}|\text{new})$ is not included in the DPSD1 model (i.e., by setting $P(\text{Rc}|\text{new}) = 0$), the maximum likelihood estimation routine cannot return a likelihood and cannot be used to derive estimates of the other parameters.

cantly shorter than the mean identification RT for all old items, $t(17) = 4.44$, $p < .001$. This contradicts the DPSD1 model (see Figure 9B), which predicts that the mean identification RT of old items receiving an R response would be longer than that of K items and would be equal to the mean identification RT of all old items. A similar data trend was evident in new items, but the differences were not reliable: identification RTs for R versus K responses, $t(17) = 1.31$, $p = .21$; identification RTs for R responses versus the mean RT for all new items, $t(17) = 2.06$, $p = .055$.

Discussion

Three findings in Experiment 3 were predicted by the SS and MS2 model but not the MS1 model; the third finding is inconsistent with the DPSD1 model. First, within old items, identification RTs decreased linearly across N, G, K, and R judgments (Prediction 5); a similar but only marginally significant effect was found for new items. Second, as in Experiments 1 and 2, the priming effect for items judged new was smaller than the overall priming effect (Prediction 2), and third, the identification RTs for old items receiving an R response were shorter than those of old items receiving a K response and were also shorter than the mean identification RT for all old items. This pattern was not reliable for new items; however, we regard the result for old items to be the crucial one because according to dual-process theory (Yonelinas, 1994), recollection only occurs for old items.

It is important to note that our DPSD1 model is just one possible application of dual-process theory to priming. Other instantiations might include an effect of recollection on priming (i.e., an arrow directly from $P(\text{Rc}|I)$ to identification RT in Figure 1D), which might arise, for example, if a partial cue during gradual clarification of the item was sufficient to recollect a studied word, hence decreasing identification RTs (consistent with the third result above). Alternatively, greater familiarity of items might lead to a more effective or faster search of episodic memory (Woollams et al., 2008), which could increase the likelihood that an item will be recollected (i.e., an arrow from f to $P(\text{Rc}|I)$ in Figure 1D), again potentially explaining the third result. Such alternative DPSD1 models could be explored in future work, as part of the same modeling framework that we introduce here.

Another possible interpretation of the third result, which may also be consistent with dual-process theory, is that R responses do not arise solely because of recollection. An R response may also occur if an item is not recollected but has a relatively high value of memory strength (f). In other words, R judgments may not be process pure (as suggested by Parks & Yonelinas, 2007; Rotello, Macmillan, Reeder, & Wong, 2005), contrary to previous formalizations of R–K judgments with dual-process theory (Yonelinas, 2002). The implementation of such a modified DPSD1 model would require the addition of an extra parameter C_R, the criterion of J_r that needs to be exceeded for an R judgment to be made. Modeling R judgments in this manner is similar to the way in which confidence ratings of 6 were modeled with the DPSD1 model in Experiment 2.

Amnesia Modeling Study

The results of Experiments 1–3 suggest that the SS model is preferable to the MS1 and MS2 models when fit to a CID-R paradigm in normal adults with three types of recognition judgment (and is preferable to the DPSD1 model in Experiment 3). However, as noted in the introduction, the bedrock of evidence for multiple-systems views has not traditionally come from studies with normal adults, but from neuropsychological studies of individuals with amnesia arising from damage to the medial temporal lobes (see Squire, 2004, for a review). More specifically, the pattern of relatively spared priming despite impaired levels of recognition in amnesia has been regarded by many as the strongest evidence for multiple-systems views. Clearly, it is important to compare the performance of the models when fit to amnesic patient data.

One data set that is well suited for comparing the models is that of Conroy et al. (2005), Experiment 2, described in the introduction. This study is well suited because of the similarity in procedures to the experiments in this article and because it reports a dissociation between priming and recognition in amnesia. The participants tested by Conroy et al. were eight control participants (CON group), three individuals with focal damage to the hippocampus (HIP group), and two individuals with more extensive damage to the medial temporal lobes (MTL group). A key finding from this study was that recognition was impaired in the MTL and HIP groups relative to the CON group, but priming and overall fluency effects were not different from those of the CON group. The two patients in the MTL group were G.P. and E.P. (described in Bayley & Squire, 2005). As described in the introduction, patient E.P. poses a particular challenge to the SS model because he shows relatively normal priming but has repeatedly performed at chance in recognition tests. As described in Prediction 3, a finding of priming in the absence of recognition is evidence against the SS model and in favor of the MS1 and MS2 models.

Conroy et al.'s (2005, Experiment 2) participants first completed a study phase in which they read 40 words. At test, the 40 studied items and 40 new items were presented to all participants via a CID-R procedure. In this procedure an item gradually clarified from a mask of pixels over a period of 11 s. Participants pressed a button to halt the clarification (the identification RT was recorded), and then they made a verbal identification of the word. An old–new recognition judgment was made after each item was identified.

We modeled the data of each participant in the CON, HIP, and MTL groups with the SS, MS1, and MS2 models in the same manner as each participant was modeled in Experiments 1–3, and with the same constraints on parameters. The DPSD1 model was not applied here (see Footnote 4). The SS model had five free parameters per participant: s, μ_{old}, σ_p, b, C. The MS1 model also had five free parameters: $\mu_{r|\text{old}}$, $\mu_{p|\text{old}}$, σ_p, b, C. The MS2 model had six free parameters: $\mu_{r|\text{old}}$, $\mu_{p|\text{old}}$, σ_p, b, C, w. The means and standard deviations of the parameter estimates across participants are given in Table 3, and the goodness-of-fit measures are given in Table 4. Bearing in mind the theoretical weight that is often given to patient E.P., we report E.P.'s results individually with G.P.'s to allow closer inspection. The AIC values in Table 4 indicated that the SS model provided the best fit to the data from the CON and HIP groups. Inspection of the AIC for each individual in these groups indicated that in the CON group, the SS model was the best fitting model for three participants; the remaining five participants were best fit by the MS1 model. Two of the patients in the HIP group were best fit by the SS model, and the third patient was best

Table 3
Means and Standard Deviations of the Parameter Estimates of Conroy et al. (2005), Experiment 2

	SS				MS1				MS2			
Parameter	CON	HIP	G.P.	E.P.	CON	HIP	G.P.	E.P.	CON	HIP	G.P.	E.P.
$\mu_{r\|old}$	1.22 (0.50)	0.68 (0.34)	0.36	0.22	1.22 (0.50)	0.60 (0.34)	0.45	0.00	1.22 (0.50)	0.61 (0.34)	0.46	0.00
$\mu_{p\|old}$	$= \mu_{r\|old}$	$= \mu_{r\|old}$	$= \mu_{r\|old}$	$= \mu_{r\|old}$	1.26 (0.54)	1.15 (0.59)	0.35	0.84	1.25 (0.54)	1.12 (0.55)	0.34	0.85
w	= 1	= 1	= 1	= 1	= 0	= 0	= 0	= 0	0.54 (0.35)	0.66 (0.22)	0.96	0.86
C	0.67 (0.45)	0.18 (0.49)	−0.06	0.02	0.67 (0.45)	0.14 (0.50)	0.00	−0.09	0.67 (0.44)	0.14 (0.50)	−0.01	−0.09
b	8,832 (1025)	7,637 (792)	9,694	8,932	8,813 (970)	7,638 (814)	9,707	9,161	8,813 (971)	7,635 (817)	9,700	9,160
s	498 (232)	635 (381)	1,106	744	= SS	= SS	= SS	= SS	= SS	= SS	= SS	= SS
σ_p	1,111 (240)	1,235 (63)	922	1,150	1,106 (248)	1,221 (80)	917	1,106	1,101 (245)	1,221 (80)	917	1,106

Note. Standard deviations are shown in parentheses. A value preceded by an equal sign indicates that the value was fixed. The parameters σ_f and σ_r are not shown, but in all models their value was fixed to $\sigma_f = \sigma_r = 1/\sqrt{2}$. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; CON = control group; HIP = focal hippocampal lesion group; G.P. and E.P. = patients in medial temporal lobe group.

fit by the MS1 model. With regard to the MTL group, the AIC for G.P. was best for the SS model, but the AIC of E.P. was best for the MS2 model.

The expected recognition results from the models are shown with the data from Conroy et al. (2005, Experiment 2) in Figure 11. All models reproduced the trend for recognition to decrease across the CON group, HIP group, and patients G.P. and E.P., and this is reflected in the decrease in the μ_r parameter across these groups in all models. Note that although the 95% confidence interval overlaps 0 in the HIP group, d' for all three patients was greater than 0. Patient G.P.'s recognition performance was below the lower 95% confidence interval of the CON group (see Figure 11) and was closely fit by all models. Patient E.P.'s recognition performance was just below chance ($d' = -0.06$). The MS1 and MS2 models produced a $d' = 0$ in E.P. (because the estimate of the parameter $\mu_{r|old}$ was 0 in both models; see Table 3), but recognition for E.P. under the SS model was slightly greater than chance ($d' = 0.22$).

The identification RT data, priming results, and model fits are shown in Figures 12, 13, and 14. In the data, priming across the CON group, HIP group, and patient E.P. did not differ.[13] However, priming is impaired in G.P. relative to the CON group; indeed, G.P.'s priming fell below the lower 95% confidence interval of the CON mean priming (see Figure 13C). The SS model gave a close fit to the priming effects in the CON group and to patient G.P. The model slightly underestimated the mean priming effect in the HIP group, but the estimate was still within the 95% confidence interval of the group mean. The SS model also underestimated the mean priming effect in patient E.P., but again the estimate still fell within the 95% confidence interval of E.P.'s data. The MS1 and MS2 models closely fit the mean priming effects in the CON group, HIP group, and patients E.P. and G.P.

Figure 15 shows the data and model fits with regard to overall fluency effects. The SS and MS2 models closely fit the observed fluency effects, but the MS1 model underestimated all the mean fluency effects. The SS and MS2 models are therefore able to reproduce the relatively spared fluency effects in amnesia.

The data and model fits regarding Predictions 1 and 2 of the models are shown in Figure 16. Regarding Prediction 1, the SS and MS2 models produced fluency effects within new and old items for the CON and HIP groups and patients E.P. and G.P., whereas the MS1 model predicted no fluency effects within new and old items. As accounted for by the SS and MS2 models, the CON group showed a reliable fluency effect within old items, $t(7) = 4.32$, $p = .003$ (though the MS2 estimate for the fluency effect within old items in the CON group fell below the lower 95% confidence interval). However, unlike the results of Experiments 1–3, there was no reliable fluency effect in the CON group within new items, $t(7) = -0.27$, $p = .83$. The HIP group also showed fluency effects within old and new items, though these were not reliable: fluency within old items, $t(2) = 2.28$, $p = .15$; fluency within new items, $t(2) = 1.43$, $p = .29$ (though all three patients in the HIP group showed a fluency effect within old items, and two out of three showed a fluency effect within new items). Patients G.P. and E.P. both showed fluency effects within old and new items (though E.P.'s confidence interval overlapped 0 in both cases).

Regarding Prediction 2, as expected by the SS and MS2 models, but not the MS1 model, the magnitude of priming overall was significantly greater than that of items judged new in the CON group, $t(7) = 4.13$, $p = .004$ (though both SS and MS2 estimates were below the lower 95% confidence interval). All three patients in the HIP group also showed this trend, but the effect was not reliable, $t(2) = 2.12$, $p = .17$. G.P. and E.P. also showed greater priming effects than priming for items judged new. Greater priming overall than for items judged new was correctly predicted by the SS model in these individuals. The MS1 model incorrectly predicted no differences. The MS2 model correctly produced the differences in the HIP group and G.P. However, according to the MS2 model, the magnitude of overall priming in E.P. should be identical to the priming effect for items judged new. This is because the estimate of $\mu_{r|old}$ for E.P. was equal to 0. E.P.'s overall priming effect was 211 ms greater than his priming effect for items judged new (see Figure 16D), though the overlap in 95% confidence intervals in Figure 13D indicates that this difference was not reliable in E.P.

In sum, the MS1 model is clearly rejected by the data of Conroy et al. (2005) and the AIC results. The SS model performed sur-

[13] We detected an outlier in E.P.'s identification RTs to new items (i.e., RT > $M + 2.5 \times SD$). If this single outlier is removed, E.P.'s priming estimate decreases by 84 ms (from 632 ms to 548 ms).

Table 4
Goodness-of-Fit Values for the Models in Conroy et al. (2005), Experiment 2

	SS			MS1			MS2		
Group	p	ln(L)	AIC	p	ln(L)	AIC	p	ln(L)	AIC
CON (N = 8)	5	−5772	**11624**	5	−5776	11633	6	−5769	11634
HIP (N = 3)	5	−2211	**4452**	5	−2212	4454	6	−2210	4455
G.P.	5	−729	**1468**	5	−734	1478	6	−729	1470
E.P.	5	−738	1487	5	−738	1485	6	−736	**1484**

Note. The Akaike information criterion (AIC) was calculated as in Table 2. Bold indicates that the model fit the data best according to the AIC. The total number of data points fit in each group was as follows: control group (CON) = 1,040; focal hippocampal lesion group (HIP) = 240; G.P. = 80; E.P. = 80. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; L = maximum likelihood.

prisingly well, closely fitting the recognition, priming, and fluency data from the CON group, HIP group, and patient G.P. (though with a slight underestimation of mean priming in the HIP group). The advance predictions of the SS model were borne out in these groups (Predictions 1 and 2, though there was no fluency effect within new items in the CON group). The MS2 model also fit the CON, HIP, and G.P. data as well as the SS model, though the SS model had the superior fit by AIC for these data.

With regard to patient E.P., his normal priming and recognition were more closely fit by the MS2 model than the SS model, and both models performed equally well with regard to the fluency effects overall and within items. The AIC indicated that the MS2 model gave the best fit to E.P; the MS2 model even produced a priming effect when recognition was at chance (providing support for the MS2 model over the SS model with regard to Prediction 3), whereas when the SS model was fit to E.P., it predicted a small residual amount of recognition and a smaller priming effect than was observed. There was, however, positive evidence against the MS2 model with regard to patient E.P.: According to the fit of the MS2 model, there should be no difference in priming overall and priming for items judged new (Prediction 2), though E.P. appeared to show one. The SS model did predict this difference. Future research should investigate whether the priming effect is indeed greater than the priming effect for items judged new in profoundly amnesic individuals. Thus, neither the SS model nor the MS2 model could completely account for E.P.'s data.

General Discussion

In this article we proposed a new modeling framework of recognition and repetition priming based on signal detection theory and used this framework to specify and test four models. In the SS model (see Figure 1A), one continuous memory signal drives recognition, priming, and fluency, whereas in the simplest multiple system model (MS1), two functionally and stochastically independent continuous memory signals drive recognition and priming (see Figure 1B). In the more complex MS2 model (see Figure 1C), the means of the two memory signals can vary independently, though the values drawn for each item can be correlated across items (conforming to a bivariate distribution with separate means but nonzero covariance). This correlation captures factors such as item distinctiveness or sustained attention across trials, which might be expected to affect both recognition and priming, while the independence of the means allows other experimental manip-

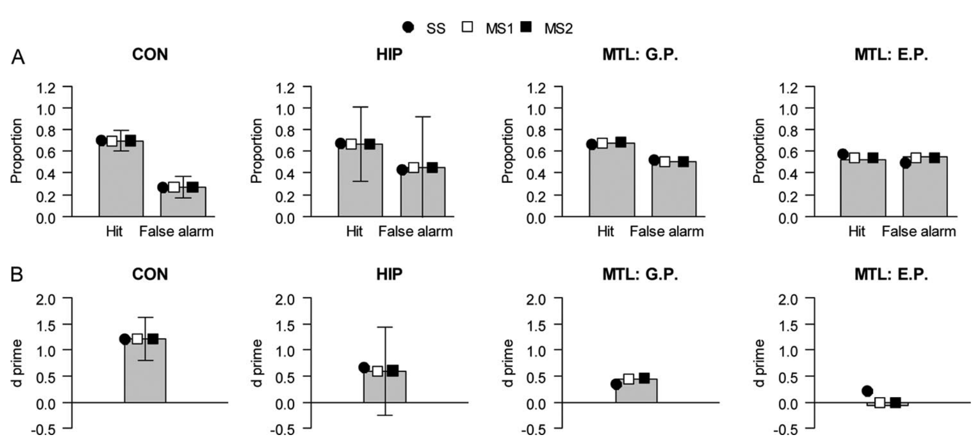

Figure 11. Recognition in Conroy et al. (2005), Experiment 2. (A) Hit and false-alarm rates. (B) Sensitivity of recognition measures (d'). Control (CON) group: n = 8; focal hippocampal lesion (HIP) group: n = 3. The results for the two patients (G.P. and E.P.) in the medial temporal lobe (MTL) group are presented individually. Bars indicate experimental data (error bars indicate 95% confidence intervals of the mean), and symbols indicate the expected result from each model. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model.

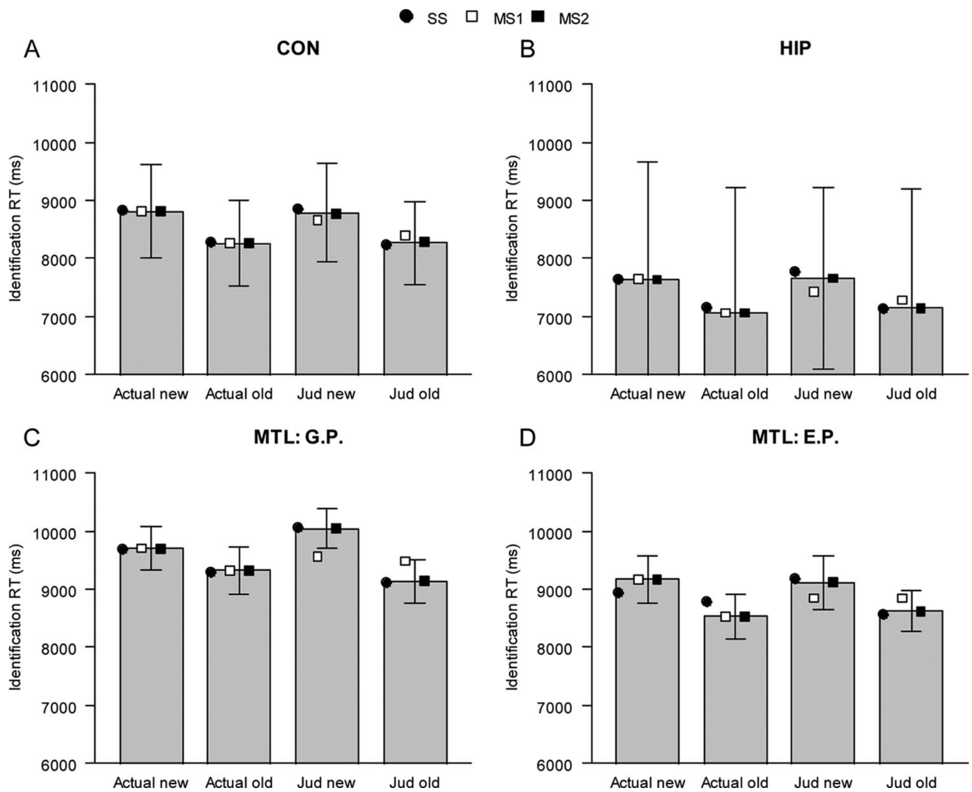

Figure 12. Mean identification reaction times (RTs) in Conroy et al. (2005), Experiment 2, classified according to whether the stimuli are actually old or new and whether they are judged (jud) old or new. Bars indicate experimental data (error bars indicate 95% confidence intervals of the group mean in Figures 12A and 12B and individual participant mean in Figures 12C and 12D), and symbols indicate the expected result from each model. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; CON = control group; HIP = focal hippocampal lesion group; MTL = medial temporal lobe group with patients G.P. and E.P.

ulations to affect recognition or priming selectively. Finally, the DPSD1 model (see Figure 1D) is similar to the SS model except that it contains an additional, probabilistic process, which can contribute to recognition. The qualitative predictions of the models were compared, as was the fit of the models (with the AIC) to the normal adult data of Experiments 1–3 and amnesic patient data of Conroy et al. (2005, Experiment 2).

With regard to the qualitative predictions, the SS and MS1 models are nested mathematically under the MS2 model; this means that it is not possible to obtain qualitative evidence that is diagnostic of the SS model or MS1 model over the MS2 model. Importantly, though, it is possible to obtain evidence for the MS2 model over the SS and MS1 models, in principle. Several specific empirical results supported the SS and MS2 models, and provided evidence against the MS1 model. Notably, the SS model made these predictions in advance: (a) fluency effects tended to occur within old and new stimuli in normal adults (Prediction 1, Experiments 1–3) and in the majority of amnesic patients tested by Conroy et al. (2005), that is, identification RTs to items judged old tended to be shorter than those of items judged new within both old and new classes of stimuli; (b) the magnitude of the priming effect for items not recognized tended to be smaller than that of the overall priming effect in both normal adults (Prediction 2, Experiments 1–3) and the amnesic patients of Conroy et al.; (c) identification RTs decreased as recognition confidence increased in normal adults, within both old and new stimuli (Prediction 4, Experiment 2); and (d) identification RTs decreased across N, G, K, and R judgments in normal adults, within old and new stimuli (Prediction 5, Experiment 3).

A third prediction concerned findings of priming in the absence of above-chance recognition. Such a finding would be positive evidence for both the MS1 and MS2 models, and would be evidence against the SS model. The results of Experiment 1 pertaining to this prediction did not discriminate the models. Although manipulations of attention at encoding have sometimes produced priming in the absence of awareness (see review in Mulligan, 2008; but see Berry, Shanks, & Henson, 2006; Berry et al., 2010), this pattern was not found in Experiment 1. In the uncued condition in Experiment 1, overall priming did not occur in the absence of overall recognition; instead priming could not be detected as recognition approached chance and all models could explain this result. The elusiveness of priming in the absence of recognition is, at least, consistent with the SS model because it predicts that overall priming will not (truly) occur in the absence of recognition (Prediction 3).

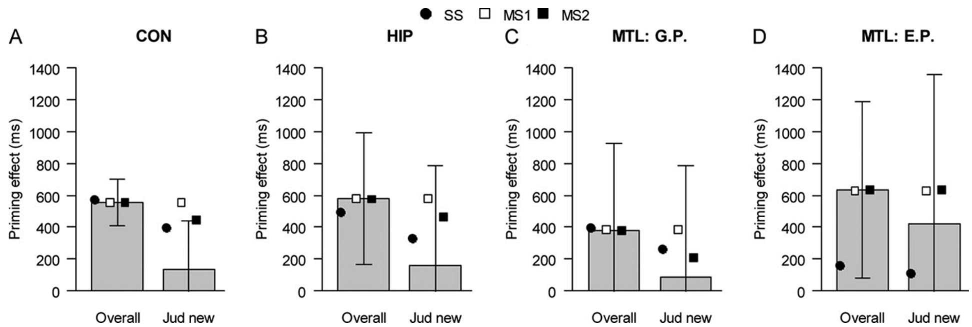

Figure 13. Overall priming versus priming for items judged (jud) new in Conroy et al. (2005), Experiment 2. Bars indicate experimental data. The error bars in Figures 13A and 13B indicate 95% confidence intervals of the group mean; the error bars for patients G.P and E.P in Figures 13C and 13D indicate the 95% confidence interval of the difference in mean identification reaction time to old and new items ("Overall" bar) and 95% confidence interval of the difference in mean identification reaction time across correct rejections and misses ("Jud new" bar). Symbols indicate the expected result from each model. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; CON = control group; HIP = focal hippocampal lesion group; MTL = medial temporal lobe group.

However, patient E.P. from the study by Conroy et al. (2005) did show a pattern of priming in the absence of recognition (see Amnesia Modeling Study), and the MS2 model was able to produce chance recognition and above-chance priming, supporting this model with regard to Prediction 3. In order to fit chance recognition, however, the $\mu_{r|old}$ parameter in the MS2 model needed to be equal to 0, and this prevented the MS2 model from accounting for other aspects of E.P.'s data (e.g., E.P.'s trend for his normal priming effect to be greater than the priming effect for items judged new: Prediction 2). The SS model predicted a small residual amount of recognition memory in E.P., and did predict a difference in E.P.'s two priming effects, although it provided a worse fit than the MS2 model according to the AIC. Thus, crucially, neither the SS nor the MS2 model provided a complete account of E.P.'s data. Clearly, E.P. is an interesting case, but we think that it is probably unwise to draw strong theoretical conclusions from this individual alone. We would ideally like to see further cases of densely amnesic individuals, and further evidence to determine whether there are differences in the normal priming effect and the priming effect for items judged new in individuals with extremely dense amnesia.

The failure of the MS1 model to explain the results relating to Predictions 1, 2, 4, and 5 suggests that this model of recognition, priming, and fluency should be rejected. Specifically, the results suggest that the sources of memorial evidence driving an item's priming and recognition judgment are not uncorrelated. Claims of this kind have been previously made (cf., e.g., Conroy et al. 2005; Tulving et al., 1982; Tulving & Schacter, 1990). For example, in an influential and still widely cited article, Tulving et al. (1982) wrote:

> Whatever it is that is transferred from the episodic study of a word to the subsequent fragment completion task *is not identical or even correlated* [emphasis added] with whatever it is that makes it possible for the subjects to distinguish between words previously encountered in the experiment and words not encountered. The information that subjects use in completing the fragments of primed words is not the

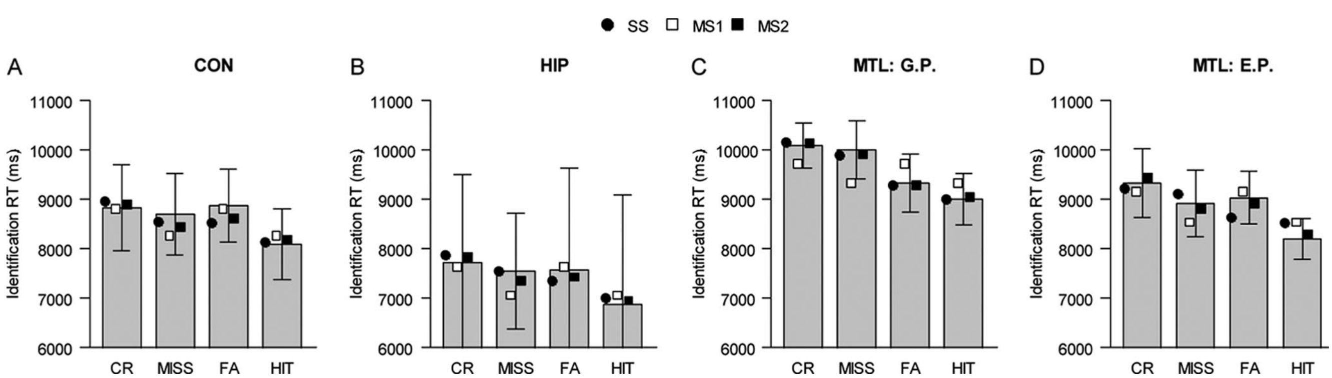

Figure 14. Identification reaction times (RTs) classified according to the recognition response (correct rejection [CR], miss, false alarm [FA], hit) in Conroy et al. (2005), Experiment 2. Bars indicate experimental data (error bars indicate 95% confidence intervals of the group mean in Figures 14A and 14B or individual participant mean in Figures 14C and 14D), and symbols indicate the expected result from each model. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; CON = control group; HIP = focal hippocampal lesion group; MTL = medial temporal lobe group with patients G.P. and E.P.

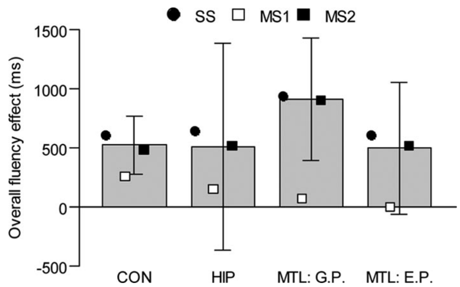

Figure 15. Mean fluency effect in Conroy et al. (2005), Experiment 2. Bars indicate experimental data. The error bars for the control (CON) and focal hippocampal lesion (HIP) groups indicate the 95% confidence interval of the mean fluency effect; the error bars for patients G.P and E.P indicate the 95% confidence interval of the difference in mean identification reaction time to judged new and judged old items. Symbols indicate the expected result from each model. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; MTL = medial temporal lobe group with patients G.P. and E.P.

> same kind of information on which people rely in remembering events from their past. (p. 341)

To the contrary, our findings suggest that multiple-systems models of recognition and priming should allow for the explicit and implicit memory strengths of an item to be correlated (and the results with the SS model go further and suggest a common memory source driving recognition and priming).

The MS2 model was able to account for many of the empirical results primarily because of the greater degree of flexibility permitted by the *w* parameter, which is the correlation between explicit and implicit item strengths (f_r and f_p). The parameter *w* is free to vary and enables the model to act in a manner similar to the SS model (when the value of *w* is close to 1), or the MS1 model (when the value of *w* is close to 0). In all experiments, the mean estimated value of *w* across participants was greater than 0 and was moderately high (i.e., between 0.54 and 0.96; see Tables 1 and 3). However, to properly identify this free parameter, it would be important to define a priori experimental manipulations that should increase or decrease its value; this is a challenge for future exploration of the MS2 model. Furthermore, diagnostic evidence of the MS2 model (over the SS and MS1 models) could have been obtained in this study had both of the following types of evidence been found simultaneously: (a) the best fitting value of *w* was greater than 0 (as was found in Experiments 1–3, and in the data of Conroy et al., 2005; this would be evidence against the MS1 model), and (b) evidence was found for priming in the absence of recognition (Prediction 3; this would be evidence against the SS model). However, with the exception of patient E.P., this conjunction was not obtained in this study and so remains a challenge for proponents of the MS2 model.

Even though it is not possible to find an empirical pattern that is predicted by the nested models (SS and MS1) but not the general (MS2) model, it is possible to evaluate the models with selection criteria that take into account the complexity of each model (as in the AIC). In this case, there was strong evidence in favor of the SS model (see Tables 2 and 4 and Figure 7). As a further test of this claim, we fit the models to the data when pooled across all experiments, that is, enforcing one set of parameter values across all experiments. This entailed nine free parameters in the SS model (μ_{old}, σ_p, b, C_1, C_2, C_3, C_4, C_5, s), nine free parameters in the MS1 model ($\mu_{r|old}$, $\mu_{p|old}$, σ_p, b, C_1, C_2, C_3, C_4, C_5), and 10 free parameters in the MS2 model ($\mu_{r|old}$, $\mu_{p|old}$, σ_p, b, w, C_1, C_2, C_3, C_4, C_5). As with the fits to individual participants' data in each experiment separately (as reported in the main text), the value of *s* in the MS1 and MS2 models was fixed to be the same as the value of *s* from the SS model. We also assumed that $C = C_3$ in Experiment 1; that $C_G = C_3$, $C_K = C_4$, and $C_R = C_5$ in Experiment 3; and that the value of *w* in the MS2 model was constant across all experiments. The "pooled" row of Table 2 shows the corresponding fit of the models. As when fitting the individual participants or experiments, the AIC indicated that the SS model should be preferred over the MS2 and MS1 models. The estimates of the parameters are shown in Table 5. The values of the parameters across models were very similar, supporting the results of the fits at the level of individuals (see also Footnote 7 and Appendix E). Most interestingly, the estimate of *w* in the MS2 model was very high (0.93), and $\mu_{r|old}$ and $\mu_{p|old}$ were very similar (0.96 and 0.99), suggesting near mimicry of the SS model (where $\mu_{r|old} = \mu_{p|old}$ and $w = 1$) by the MS2 model (but see Appendix B).

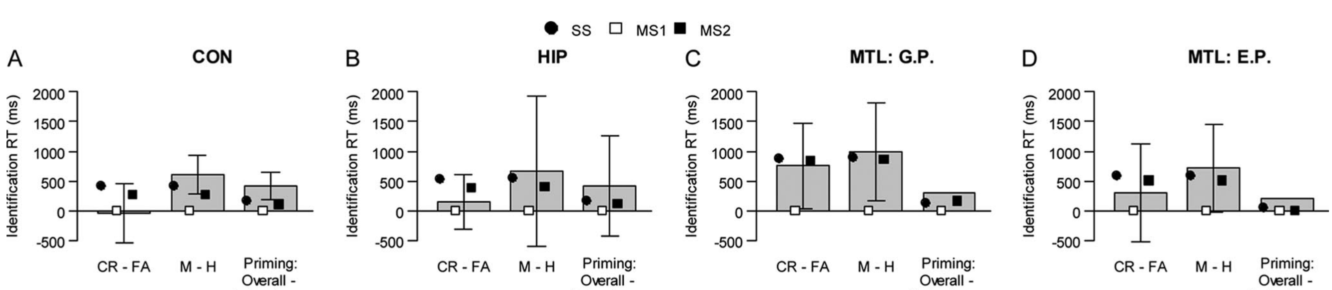

Figure 16. Conroy et al. (2005), Experiment 2, data and model results concerning Predictions 1 and 2. Bars indicate experimental data. Error bars indicate 95% confidence intervals of the group mean difference in Figures 16A and 16B and individual participant mean difference in Figures 16C and 16D. Symbols indicate the expected result from each model. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; CON = control group; HIP = focal hippocampal lesion group; MTL = medial temporal lobe group with patients G.P. and E.P.; RT = reaction time; CR–FA = correct rejection–false alarm; M–H = hit–miss.

Table 5
Parameters of the Models When Fit to the Pooled Data From Experiments 1–3

Parameter	SS	MS1	MS2
$\mu_{r\|old}$	0.96	0.96	0.96
$\mu_{p\|old}$	$= \mu_{r\|old}$	1.00	0.99
w	= 1	= 0	0.93
C_1	−0.30	−0.31	−0.30
C_2	0.17	0.16	0.17
$C_3/C/C_G$	0.49	0.49	0.49
C_4/C_K	0.86	0.85	0.85
C_5/C_R	1.34	1.33	1.34
b	2,342	2,343	2,342
s	150	= 150	= 150
σ_p	824	825	824

Note. A value preceded by an equal sign indicates that the value was fixed. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model.

The results have implications for extensions of dual-process theory of recognition (Yonelinas, 1994, 2002) to priming and fluency phenomena such as those investigated in the present CID-R paradigm. The DPSD1 model that we implemented assumes that the same signal drives the familiarity component of recognition and also priming (similar to the proposal of Jacoby & Dallas, 1981) and that an independent probabilistic recollection process also contributes to recognition. The model was tested in Experiments 2 and 3. This DPSD1 model performed well in Experiment 2, providing a better AIC fit than the other models. However, in Experiment 3, we found that the identification RTs to old items receiving an R response were shorter than those that received a K response (and were shorter than the mean identification RT of all old items overall; see Sheldon & Moscovitch, 2010, for related findings with other priming tasks). This was predicted by the SS model (Prediction 5), but not by the DPSD1 model, and the AIC of the SS model was better. One interpretation of this is that R responses are not process-pure measures of recollection (see also, e.g., Rotello et al., 2005; Wixted & Stretch, 2004). This finding is, however, consistent with the SDT interpretation of R–K judgments, according to which R–K judgments index different levels of memory strength, rather than distinct memory processes (e.g., Donaldson, 1996; Dunn, 2004). If R items have a greater strength than K items, and if the same memory strength variable drives priming (as in the SS model; or implicit–explicit memory strengths can be correlated as in the MS2 model), then identification RTs to R items will tend to be shorter than those of K items, as was observed in Experiment 3. Nonetheless, there may be alternative instantiations of dual-process theory that could be applied more successfully to conjoint recognition and priming data, as mentioned in the Discussion to Experiment 3.

The results of the amnesic patient data from Conroy et al. (2005) in our Amnesia Modeling Study suggest that the dissociation between recognition and priming in amnesic patients' data is not inconsistent with the SS model. It is possible for the SS model to closely reproduce the dissociation between recognition and priming in amnesia without any distinction between explicit and implicit memory, provided that one accepts that there may be a small impairment in priming in amnesia that often goes undetected. Indeed, as reviewed in the introduction, the issue of whether priming is completely normal in amnesia has proven to be controversial, and there is evidence to suggest that priming is impaired in amnesia (e.g., Ostergaard & Jernigan, 1993, 1996; Ostergaard, 1994, 1999). In the SS model, reductions in the mean memory strength signal μ_r affect recognition and priming, but the effect on recognition can be more pronounced.

Contamination of Recognition and Priming Measures?

One possible alternative explanation for the general relation between identification RTs and recognition judgments reported in this article is that the interleaved nature of the identification and recognition trials in the CID-R task encouraged the use of fluency to make the recognition judgment. For example, if an item is identified relatively quickly, participants may be more likely to attribute this relative ease of identification to its prior exposure at study (e.g., as proposed by Jacoby & Dallas, 1981). Indeed, this type of perceptual fluency effect has been demonstrated on recognition judgments (e.g., Huber, Clark, Curran, & Winkielman, 2008; Jacoby & Whitehouse, 1989; Westerman, Lloyd, & Miller, 2002).

We do not think that the associations evident between recognition judgment and identification RTs can be explained solely by such "contamination" between tests. Importantly, studies have reported similar associations to those reported here even when the priming and recognition tasks are presented in different blocks, rather than interleaved. For example, Sheldon and Moscovitch (2010) found that lexical decision latencies to studied items subsequently receiving remember responses were shorter than those of know responses in a design in which recognition judgment and lexical decision trials were presented in different blocks. They also found a similar result when the priming measure was word stem completion time and provided evidence that these associations were not due to item characteristics. Furthermore, Ostergaard (1998, Experiment 2) used a design in which the priming measurement (as indexed by identification latency of gradually clarifying words) and recognition trials were presented in separate blocks. Ostergaard found that priming for hits was greater than that of misses (as in Prediction 2). These findings suggest that the interleaved nature of the CID-R task is not the crucial factor for finding the associations between recognition and priming in this study.

Another possibility is that participant's identification of studied items in the CID portion of the CID-R trial is aided by attempts to recollect items from the study phase. That is, it is possible that the priming measure is contaminated by explicit retrieval strategies. There is evidence to suggest that this does not typically occur. For example, Brown, Neblett, Jones, and Mitchell (1991) found that priming did not differ under conditions that encouraged the use of explicit retrieval strategies. They found that the magnitude of priming observed when old and new items were presented in different blocks—and subjects were told whether a block contained old items or new items—was not significantly different from the magnitude of priming when old and new trials were interleaved. Furthermore, studies have found that the magnitude of priming does not differ when the identification trials are presented in a block on their own, or when interleaved with recognition trials (Brown, Jones, & Mitchell, 1996; Stark & McClelland, 2000, Experiment 5), suggesting that the presence of the recognition task does not lead to increased priming (however, Stark & McClelland, 2000, did find that overall priming differed under analogous conditions in their Experiment 4; for other arguments against contamination accounts, see the General Discussion section of Ostergaard, 1998). Unpublished studies from our laboratory also echo this latter finding; we find that priming for pictures of objects does not reliably differ when identification trials are blocked or interleaved with recognition trials. The modeling framework presented here could, in principle, be extended to formalize and test contamination accounts (we suspect that such formalization is likely to have much in common with the MS2 model). Indeed, we argue that it is essential to formalize models to allow such detailed testing.

Limitations

A limitation of the SS model is that it cannot produce reversed associations (Dunn & Kirsner, 1988). A reversed association between recognition and priming would be demonstrated if, under one set of conditions, a variable is shown to produce opposite effects on recognition and priming (i.e., what is commonly referred to as a "crossed dissociation") and yet, under another set of conditions, another variable is shown to affect priming and recognition in a similar manner. Indeed, reverse associations do exist when pooling results across studies in the literature. For example, generating an item from its antonym at encoding leads to greater recognition compared with simply reading the item at encoding, but the opposite is true for priming (a crossed dissociation; Jacoby, 1983; Dew & Mulligan, 2008); yet other variables like selective attention can exert similar effects on recognition and priming (Berry, Henson, & Shanks, 2006; Berry, Shanks, & Henson, 2006; Berry et al., 2010). A reversed association has also been reported within a single word stem completion study (Richardson-Klavehn et al., 1999). The MS2 model, however, can produce reversed associations because the μ_r and μ_p parameters are free to vary across conditions independently of each other. We have previously speculated that the SS model may be able to account for reversed associations by decomposing the single memory signal into separate modality-specific and amodal conceptual memory signals (rather than implicit and explicit memory signals), resembling a transfer-appropriate processing account of recognition and priming (e.g., Blaxton, 1989). Although such a model might be structurally similar to the MS2 model, importantly, because both signals are assumed to be accessible to awareness, the model would not predict that priming could occur in the absence of recognition. Exploration of such an account remains an issue for future research.

Another limitation is that the modeling framework also does not explicitly incorporate variability in parameters across participants or across items. Future developments of the framework could extend it hierarchically to take into account the variability in item characteristics and individual participants (as in hierarchical signal detection models; e.g., Pratte, Rouder, & Morey, 2010). Indeed, accounting for such characteristics has proven helpful in nonlinear computational models to help delineate observations that are due to mnemonic processes and surface variables such as item characteristics (Rouder, Lu, Morey, Sun, & Speckman, 2008).

It should be noted that the majority of evidence for multiple memory systems has come from between-task comparisons, whereas here we have specifically focused upon modeling performance in CID-R paradigms in which the recognition and priming trials are interleaved. Does the modeling framework generalize to between-task comparisons? First, there is nothing in the modeling framework to preclude its application to between-task comparisons. For example, the framework could still be applied if identification and recognition of each item were measured in different blocks of trials. However, one potential concern with using such a design is that obtaining measures of recognition and priming for a given item at two nonproximal points in time may mean that the measures are differentially affected by other factors such as differences in retention interval or changes in participant motivation. Such factors could be modeled within the framework, but doing so would considerably increase its complexity. We believe that a major advantage of using the CID-R paradigm is that it allows a recognition and priming measurement for each item at test, and these measures can be taken for each item relatively concurrently. Secondly, since others have used the results of CID-R tasks to argue for the independence of recognition and priming (e.g., Conroy et al., 2005), it seems only fair for us to determine whether findings from this task can be accounted for by an SS model. Finally, in previous work with a simpler implementation of the SS model, we have simulated a basic dissociation between recognition and priming from a between-task design (Berry, Henson, & Shanks, 2006).

Lastly, our focus in this article has been on particular manifestations of explicit and implicit memory, namely recognition and priming, respectively. However, there are many forms of learning and memory that have generated data relevant to the memory systems debate. Do these provide more robust evidence against the notion of a single memory signal driving performance across explicit and implicit tasks? The literature is so large that we can only confine ourselves here to comments on one particularly influential example. Karni and Sagi (1991) studied perceptual learning of a texture discrimination skill and were able to demonstrate not only long-term retention of this skill but also that it was highly specific: Notably, there was no transfer between retinal locations. Karni and Sagi interpreted this finding as evidence that the locus of perceptual (implicit) skill learning is in primary visual cortex, where neurons code information retinotopically. Later research showed this conclusion to be incorrect, though, and implicated instead a "central" locus for such perceptual learning. The crucial evidence comes from a study by Xiao et al. (2008), who modified Karni and Sagi's task in a small but important way: Their participants learned one perceptual discrimination (e.g., contrast) at one retinal location and concurrently learned a quite different discrimination (e.g., orientation) at a second location. Under these conditions, complete transfer of each task to the location at which the other task was trained (but no transfer when no task had been trained at that location) was observed. Xiao et al. inferred that a central locus plays an intrinsic role in perceptual learning, by controlling changes in spatial attention to relevant retinal locations. The implication is that Karni and Sagi's findings arose not because learning was localized to retinotopic neurons in primary visual cortex (which would provide no benefit for other retinal locations), but because the task led to changes in the distribution of spatial attention (which in turn led to transfer decrements). Thus, in line with our conclusions, this research suggests that many examples of explicit and implicit learning can be explained in terms of a single memory process.

Conclusion

To conclude, we can reject the MS1 model (primarily on the basis of positive evidence in Predictions 1, 2, 4, and 5) and the DPSD1 model (on the basis of positive evidence in Prediction 5) as accounts of priming and recognition in the present CID-R paradigm. The MS2 and SS models cannot be disambiguated on the basis of our qualitative empirical findings, but, according to the AIC, the SS model should be preferred over the MS2 model. Moreover, this simplicity allows it to make clear predictions in advance, as exemplified by Predictions 1–5 that were tested here. Thus, the idea that a single memory strength signal drives recognition, priming, and fluency is at least a viable alternative to the prevailing notion that there are functionally and stochastically independent explicit and implicit memory signals. Most importantly, we hope that the new formal modeling framework presented here will serve as a useful platform for others to develop and test formal theories of explicit and implicit memory.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Budapest, Hungary: Academiai Kiado.

Arnold, B. C., Beaver, R. J., Groeneveld, R. A., & Meeker, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika, 58,* 471–488. doi:10.1007/BF02294652

Bayley, P. J., & Squire, L. R. (2005). Failure to acquire new semantic knowledge in patients with large medial temporal lobe lesions. *Hippocampus, 15,* 273–280. doi:10.1002/hipo.20057

Benjamin, A. S. (2010). Representational explanations of "process" dissociations in recognition: The DRYAD theory of aging and memory judgments. *Psychological Review, 117,* 1055–1079. doi:10.1037/a0020810

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review, 116,* 84–115. doi:10.1037/a0014351

Berry, C. J., Henson, R. N. A., & Shanks, D. R. (2006). On the relationship between repetition priming and recognition memory: Insights from a computational model. *Journal of Memory and Language, 55,* 515–533. doi:10.1016/j.jml.2006.08.008

Berry, C. J., Shanks, D. R., & Henson, R. N. A. (2006). On the status of unconscious memory: Merikle and Reingold (1991) revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 925–934. doi:10.1037/0278-7393.32.4.925

Berry, C. J., Shanks, D. R., & Henson, R. N. A. (2008a). A single-system account of the relationship between priming, recognition, and fluency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 97–111. doi:10.1037/0278-7393.34.1.97

Berry, C. J., Shanks, D. R., & Henson, R. N. A. (2008b). A unitary signal-detection model of implicit and explicit memory. *Trends in Cognitive Sciences, 12,* 367–373. doi:10.1016/j.tics.2008.06.005

Berry, C. J., Shanks, D. R., Li, S., Rains, L. S., & Henson, R. N. A. (2010). Can "pure" implicit memory be isolated? A test of a single-system model of recognition and repetition priming. *Canadian Journal of Experimental Psychology, 64,* 241–255. doi:10.1037/a0021525

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 657–668. doi:10.1037/0278-7393.15.4.657

Brown, A. S., Jones, T. C., & Mitchell, D. B. (1996). Single and multiple test repetition priming in implicit memory. *Memory, 4,* 159–173. doi:10.1080/096582196389022

Brown, A. S., & Mitchell, D. B. (1994). A reevaluation of semantic versus nonsemantic processing in implicit memory. *Memory & Cognition, 22,* 533–541. doi:10.3758/BF03198392

Brown, A. S., Neblett, D. R., Jones, T. C., & Mitchell, D. B. (1991). Transfer of processing in repetition priming: Some inappropriate findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 514–525. doi:10.1037/0278-7393.17.3.514

Buchner, A., & Brandt, M. (2003). Further evidence for systematic reliability differences between explicit and implicit memory tests. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*A, 193–209. doi:10.1080/02724980244000260

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology, 40,* 227–259. doi:10.1006/cogp.1999.0731

Cermak, L. S., Talbot, N., Chandler, K., & Wolbarst, L. R. (1985). The perceptual priming phenomenon in amnesia. *Neuropsychologia, 23,* 615–622. doi:10.1016/0028-3932(85)90063-6

Challis, B. H., & Brodbeck, D. R. (1992). Level of processing affects priming in word fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 595–607. doi:10.1037/0278-7393.18.3.595

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

Conroy, M. A., Hopkins, R. O., & Squire, L. R. (2005). On the contribution of perceptual fluency and priming to recognition memory. *Cognitive, Affective, & Behavioral Neuroscience, 5,* 14–20. doi:10.3758/CABN.5.1.14

Craik, F. I. M., Moscovitch, M., & McDowd, J. M. (1994). Contributions of surface and conceptual information to performance on implicit and explicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 864–875. doi:10.1037/0278-7393.20.4.864

Dew, I. T. Z., & Cabeza, R. (2011). The porous boundaries between explicit and implicit memory: Behavioral and neural evidence. *Annals of the New York Academy of Sciences, 1224,* 174–190. doi:10.1111/j.1749-6632.2010.05946.x

Dew, I. T. Z., & Mulligan, N. W. (2008). The effects of generation on auditory implicit memory. *Memory & Cognition, 36,* 1157–1167. doi:10.3758/MC.36.6.1157

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523–533. doi:10.3758/BF03200940

Dunn, J. C. (2003). The elusive dissociation. *Cortex, 39,* 177–179. doi:10.1016/S0010-9452(08)70096-0

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

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95,* 91–101. doi:10.1037/0033-295X.95.1.91

Dunn, J. C., & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex, 39,* 1–7. doi:10.1016/S0010-9452(08)70070-4

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112,* 309–346. doi:10.1037/0096-3445.112.3.309

Fleischman, D. A. (2007). Repetition priming in aging and Alzheimer's disease: An integrative review and future directions. *Cortex, 43,* 889–897. doi:10.1016/S0010-9452(08)70688-9

Fleischman, D. A., & Gabrieli, J. D. E. (1998). Repetition priming in normal aging and Alzheimer's disease: A review of findings and theories. *Psychology and Aging, 13,* 88–119. doi:10.1037/0882-7974.13.1.88

Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology, 49,* 87–115. doi:10.1146/annurev.psych.49.1.87

Gabrieli, J. D. E., Fleischman, D. A., Keane, M. M., Reminger, S. L., & Morrell, F. (1995). Double dissociation between memory systems underlying explicit and implicit memory in the human brain. *Psychological Science, 6,* 76–82. doi:10.1111/j.1467-9280.1995.tb00310.x

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313. doi:10.3758/BF03197041

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229–244). New York, NY: Oxford University Press.

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

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 164–178. doi:10.1037/0278-7393.10.1.164

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York, NY: Wiley.

Hamann, S. B., & Squire, L. R. (1997). Intact perceptual memory in the absence of conscious memory. *Behavioral Neuroscience, 111,* 850–854. doi:10.1037/0735-7044.111.4.850

Hamann, S. B., Squire, L. R., & Schacter, D. L. (1995). Perceptual thresholds and priming in amnesia. *Neuropsychology, 9,* 3–15. doi:10.1037/0894-4105.9.1.3

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience, 11,* 523–532. doi:10.1038/nrn2850

Hintzman, D. L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology, 41,* 109–139. doi:10.1146/annurev.ps.41.020190.000545

Howe, M. L., Rabinowitz, F. M., & Grant, M. J. (1993). On measuring (in)dependence of cognitive processes. *Psychological Review, 100,* 737–747. doi:10.1037/0033-295X.100.4.737

Huber, D. E., Clark, T. F., Curran, T., & Winkielman, P. (2008). Effects of repetition priming on recognition memory: Testing a perceptual fluency–disfluency model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1305–1324. doi:10.1037/a0013370

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 485–508. doi:10.1016/S0022-5371(83)90301-8

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340. doi:10.1037/0096-3445.110.3.306

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118,* 126–135. doi:10.1037/0096-3445.118.2.126

Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *Journal of Experimental Psychology: General, 138,* 291–306. doi:10.1037/a0015525

Jernigan, T. L., & Ostergaard, A. L. (1993). Word priming and recognition memory are both affected by mesial temporal lobe damage. *Neuropsychology, 7,* 14–26. doi:10.1037/0894-4105.7.1.14

Johnston, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 3–11. doi:10.1037/0278-7393.11.1.3

Johnston, W. A., Hawley, K. J., & Elliott, M. G. (1991). Contribution of perceptual fluency to recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 210–223. doi:10.1037/0278-7393.17.2.210

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America, 88,* 4966–4970. doi:10.1073/pnas.88.11.4966

Kinder, A., & Shanks, D. R. (2001). Amnesia and the declarative/nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience, 13,* 648–669. doi:10.1162/089892901750363217

Kinder, A., & Shanks, D. R. (2003). Neuropsychological dissociations between priming and recognition: A single-system connectionist account. *Psychological Review, 110,* 728–744. doi:10.1037/0033-295X.110.4.728

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

Lukatela, G., Eaton, T., Moreno, M. A., & Turvey, M. T. (2007). Equivalent inter- and intramodality long-term priming: Evidence for a common lexicon for words seen and words heard. *Memory & Cognition, 35,* 781–800. doi:10.3758/BF03193315

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

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

Masson, M. E. J., & MacLeod, C. M. (1992). Reenacting the route to interpretation: Enhanced perceptual identification without prior perception. *Journal of Experimental Psychology: General, 121,* 145–176. doi:10.1037/0096-3445.121.2.145

Masson, M. E. J., & MacLeod, C. M. (2002). Covert operations: Orthographic recoding as a basis for repetition priming in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 858–871. doi:10.1037/0278-7393.28.5.858

Meier, B., & Perrig, W. J. (2000). Low reliability of perceptual priming: Consequences for the interpretation of functional dissociations between explicit and implicit memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53*A, 211–233. doi:10.1080/027249800390745

Meier, B., Theiler-Bürgi, M., & Perrig, W. (2009). Levels of processing and amnesia affect perceptual priming in fragmented picture naming. *International Journal of Neuroscience, 119,* 1061–1075. doi:10.1080/00207450802336691

Merikle, P. M., & Reingold, E. M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 224–233. doi:10.1037/0278-7393.17.2.224

Mitchell, D. B., & Brown, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 213–222. doi:10.1037/0278-7393.14.2.213

Mitchell, D. B., Brown, A. S., & Murphy, D. R. (1990). Dissociations between procedural and episodic memory: Effects of time and aging. *Psychology and Aging, 5,* 264–276. doi:10.1037/0882-7974.5.2.264

Moscovitch, M., & Bentin, S. (1993). The fate of repetition effects when recognition approaches chance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 148–158. doi:10.1037/0278-7393.19.1.148

Mulligan, N. W. (2008). Attention and memory. In H. L. Roediger (Ed.), *Learning and memory: A comprehensive reference* (Vol. 2, pp. 7–22). Oxford, England: Elsevier. doi:10.1016/B978-012370509-9.00134-0

Mulligan, N. W., & Dew, I. T. Z. (2009). Generation and perceptual implicit memory: Different generation tasks produce different effects on perceptual priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1522–1538. doi:10.1037/a0017398

Mulligan, N. W., & Osborn, K. (2009). The modality-match effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 564–571. doi:10.1037/a0014524

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology, 49,* 47–84. doi:10.1016/j.cogpsych.2003.11.001

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences, 12,* 285–290. doi:10.1016/j.tics.2008.04.009

Ostergaard, A. L. (1992). A method for judging measures of stochastic dependence: Further comments on the current controversy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 413–420. doi:10.1037/0278-7393.18.2.413

Ostergaard, A. L. (1994). Dissociations between word priming effects in normal subjects and patients with memory disorders: Multiple memory systems or retrieval? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 47*A, 331–364.

Ostergaard, A. L. (1998). The effects on priming of word frequency, number of repetitions, and delay depend on the magnitude of priming. *Memory & Cognition, 26,* 40–60. doi:10.3758/BF03211369

Ostergaard, A. L. (1999). Priming deficits in amnesia: Now you see them, now you don't. *Journal of the International Neuropsychological Society, 5,* 175–190. doi:10.1017/S1355617799533018

Ostergaard, A. L., & Jernigan, T. L. (1993). Are word priming and explicit memory mediated by different brain structures? In P. Graf & M. E. J. Masson (Eds.), *Implicit memory: New directions in cognitive, development, and neuropsychology* (pp. 327–349). Hillsdale, NJ: Erlbaum.

Ostergaard, A. L., & Jernigan, T. L. (1996). Priming and baseline perceptual identification performance in amnesia: A comment on Hamann, Squire, and Schacter. *Neuropsychology, 10,* 125–130. doi:10.1037/0894-4105.10.1.125

Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review, 114,* 188–201. doi:10.1037/0033-295X.114.1.188

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17,* 291–321. doi:10.1080/01688639508405124

Poldrack, R. A. (1996). On testing for stochastic dissociations. *Psychonomic Bulletin & Review, 3,* 434–448. doi:10.3758/BF03214547

Poldrack, R. A., & Logan, G. D. (1997). Fluency and response speed in recognition judgments. *Memory & Cognition, 25,* 1–10. doi:10.3758/BF03197280

Pratte, M. S., Rouder, J. N., & Morey, R. D. (2010). Separating mnemonic process from participant and item effects in the assessment of ROC asymmetries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 224–232. doi:10.1037/a0017682

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review, 116,* 59–83. doi:10.1037/a0014086

R Development Core Team. (2008). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Reder, L. M., Park, H., & Kieffaber, P. D. (2009). Memory systems do not divide on consciousness: Reinterpreting memory in terms of activation and binding. *Psychological Bulletin, 135,* 23–49. doi:10.1037/a0013974

Richardson-Klavehn, A., Clarke, A. J. B., & Gardiner, J. M. (1999). Conjoint dissociations reveal involuntary "perceptual" priming from generating at study. *Consciousness and Cognition, 8,* 271–284. doi:10.1006/ccog.1999.0382

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63–131). Amsterdam, the Netherlands: Elsevier.

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

Rouder, J. N., Lu, J., Morey, R. D., Sun, D., & Speckman, P. L. (2008). A hierarchical process–dissociation model. *Journal of Experimental Psychology: General, 137,* 370–389. doi:10.1037/0096-3445.137.2.370

Rünger, D., Nagy, G., & Frensch, P. A. (2009). Do recognition and priming index a unitary knowledge base? Comment on Shanks et al. (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 572–585. doi:10.1037/a0014543

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 501–518. doi:10.1037/0278-7393.13.3.501

Schacter, D. L., Dobbins, I. G., & Schnyer, D. M. (2004). Specificity of priming: A cognitive neuroscience perspective. *Nature Reviews Neuroscience, 5,* 853–862. doi:10.1038/nrn1534

Schott, B. H., Henson, R. N., Richardson-Klavehn, A., Becker, C., Thoma, V., Heinze, H.-J., & Düzel, E. (2005). Redefining implicit and explicit memory: The functional neuroanatomy of priming, remembering, and control of retrieval. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 1257–1262. doi:10.1073/pnas.0409070102

Schott, B. H., Richardson-Klavehn, A., Henson, R. N. A., Becker, C., Heinze, H.-J., & Düzel, E. (2006). Neuroanatomical dissociation of encoding processes related to priming and explicit memory. *Journal of Neuroscience, 26,* 792–800. doi:10.1523/JNEUROSCI.2402-05.2006

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464. doi:10.1214/aos/1176344136

Shanks, D. R. (2005). Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202–220). London, England: Sage.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17,* 367–447. doi:10.1017/S0140525X00035032

Shanks, D. R., Wilkinson, L., & Channon, S. (2003). Relationship between priming and recognition in deterministic and probabilistic sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 248–261. doi:10.1037/0278-7393.29.2.248

Sheldon, S. A. M., & Moscovitch, M. (2010). Recollective performance advantages for implicit memory tasks. *Memory, 18,* 681–697. doi:10.1080/09658211.2010.499876

Squire, L. R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems* (pp. 203–231). Cambridge, MA: MIT Press.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology Learning and Memory, 82,* 171–177. doi:10.1016/j.nlm.2004.06.005

Squire, L. R. (2009). Memory and brain systems: 1969–2009. *Journal of Neuroscience, 29,* 12711–12716. doi:10.1523/JNEUROSCI.3575-09.2009

Stadler, M. A., & Roediger, H. L., III. (1998). The question of awareness in research on implicit learning. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 105–132). Thousand Oaks, CA: Sage.

Stark, C. E. L., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 945–972. doi:10.1037/0278-7393.26.4.945

Stefanacci, L., Buffalo, E. A., Schmolck, H., & Squire, L. R. (2000). Profound amnesia after damage to the medial temporal lobe: A neuroanatomical and neuropsychological profile of patient E. P. *Journal of Neuroscience, 20,* 7024–7036.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12. doi:10.1037/h0080017

Tulving, E. (1999). Study of memory: Processes and systems. In J. K. Foster & M. Jelicic (Eds.), *Memory: Systems, process, or function?* (pp. 11–30). New York, NY: Oxford University Press.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247,* 301–306. doi:10.1126/science.2296719

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 336–342. doi:10.1037/0278-7393.8.4.336

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science, 25,* 111–172. doi:10.1016/S0364-0213(00)00036-7

Verfaellie, M., & Cermak, L. S. (1999). Perceptual fluency as a cue for recognition judgments in amnesia. *Neuropsychology, 13,* 198–205. doi:10.1037/0894-4105.13.2.198

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology, 48,* 28–50. doi:10.1016/j.jmp.2003.11.004

Westerman, D. L., Lloyd, M. E., & Miller, J. K. (2002). On the attribution of perceptual fluency in recognition memory: The role of expectation. *Journal of Memory and Language, 47,* 607–617. doi:10.1016/S0749-596X(02)00022-0

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., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117,* 1025–1054. doi:10.1037/a0020874

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11,* 616–641. doi:10.3758/BF03196616

Woollams, A. M., Taylor, J. R., Karayanidis, F., & Henson, R. N. (2008). Event-related potentials associated with masked priming of test cues reveal multiple potential contributions to recognition memory. *Journal of Cognitive Neuroscience, 20,* 1114–1129. doi:10.1162/jocn.2008.20076

Xiao, L.-Q., Zhang, J.-Y., Wang, R., Klein, S. A., Levi, D. M., & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Current Biology, 18,* 1922–1926. doi:10.1016/j.cub.2008.10.030

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354. doi:10.1037/0278-7393.20.6.1341

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

Yonelinas, A. P., Kroll, N. E. A., Baynes, K., Dobbins, I. G., Frederick, C. M., Knight, R. T., & Gazzaniga, M. S. (2001). Visual implicit memory in the left hemisphere: Evidence from patients with callosotomies and right occipital lobe lesions. *Psychological Science, 12,* 293–298. doi:10.1111/1467-9280.00354

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

(*Appendices follow*)

Appendix A

Parameter Estimation: Single-System, Multiple-Systems-1, and Multiple-Systems-2 Models

The parameters of the single-system (SS), multiple-systems-1 (MS1), and multiple-systems-2 (MS2) models (as formalized in Equations 4–6) were determined with maximum likelihood estimation procedures. The likelihood for a pair of observations, Z and reaction time (RT), where Z denotes the recognition judgment on a given continuous-identification-with-recognition (CID-R) trial, is given as

$$P(Z, RT) = P(Z|RT)P(RT) = (\int P(Z|J_r)P(J_r|RT)dJ_r)P(RT). \tag{A1}$$

Hence

$$L(Z, RT|I) = [\Phi(C_j|\mu_{Jr|\mathrm{RT},I}, \sigma^2_{Jr|\mathrm{RT}}) - \Phi(C_{j-1}|\mu_{Jr|\mathrm{RT},I}, \sigma^2_{Jr|\mathrm{RT}})] \times \phi(RT|b - s\mu_{\mathrm{p}|I}, \sigma^2_{\mathrm{RT}}), \tag{A2}$$

where I = old, new; Φ is the cumulative normal distribution function; ϕ is the normal density function; $\sigma^2_{\mathrm{RT}} = s^2\sigma^2_f + \sigma^2_{\mathrm{p}}$ (from Equation 2); $\mu_{Jr|\mathrm{RT},I}$ and $\sigma^2_{Jr|\mathrm{RT}}$ are the mean and variance of the conditional distribution of J_r given RT. For Experiment 1 and Conroy et al. (2005, Experiment 2), $j = 1$ when Z = "new" (N), and $j = 2$ when Z = "old" (O); $C_0 = -\infty$, $C_1 = C$, and $C_2 = \infty$. For Experiment 2, $j = Z = 1, \ldots, 6$; $C_0 = -\infty$; C_1–C_5 are the decision criteria; and $C_6 = \infty$. For Experiment 3, $j = 1, \ldots, 4$; $j = 1$ when Z = N; $j = 2$ when Z = G; $j = 3$ when Z = K; $j = 4$ when Z = R; $C_0 = -\infty$, $C_1 = C_G$, $C_2 = C_K$, $C_3 = C_R$, and $C_4 = \infty$.

To calculate $\mu_{Jr|\mathrm{RT},I}$ and $\sigma^2_{Jr|\mathrm{RT}}$, we made use of the fact that if X and Y follow a bivariate normal distribution, then

$$Y|X \sim N(\mu_{Y|X}, \sigma^2_{Y|X})$$

with

$$\mu_{Y|X} = \mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

and

$$\sigma^2_{Y|X} = (1 - \rho^2_{XY})\sigma^2_Y. \tag{A3}$$

Substituting parts of Equations 4–6 into Equation A3, we obtain

$$\mu_{Jr|\mathrm{RT},I} = \mu_{\mathrm{r}|I} - \frac{ws\sigma^2_f(RT - b + s\mu_{\mathrm{p}|I})}{s^2\sigma^2_f + \sigma^2_{\mathrm{p}}}$$

and

$$\sigma^2_{Jr|\mathrm{RT}} = \sigma^2_f + \sigma^2_{\mathrm{r}} - \frac{w^2s^2\sigma^4_f}{s^2\sigma^2_f + \sigma^2_{\mathrm{p}}}, \tag{A4}$$

where $\mu_{\mathrm{r|new}} = 0$ when I = new, and $\mu_{\mathrm{r|old}} \geq 0$ when I = old; $\mu_{\mathrm{p|new}} = 0$ when I = new and $\mu_{\mathrm{p|old}} \geq 0$ when I = old. In the SS (and dual-process signal detection [DPSD1]) model, $\mu_{\mathrm{r|old}} = \mu_{\mathrm{p|old}} = \mu_{\mathrm{old}}$, and $w = 1$. In the MS1 model, $w = 0$; in the MS2 model, $0 < w < 1$.

DPSD1 Model

The likelihood functions for the DPSD1 model are similar to those above, except that an additional recollection process is included: Each item is either recollected or not, and if it is not recollected, then the judgment is based upon familiarity (f). Priming is driven by f, as in the model above. In the DPSD1 model, recollection is assumed to occur for old items that are judged old with the maximal level of confidence, or receive an R response in a remember–know task. Recollection occurs with a fixed probability, $P(\mathrm{Rc|I})$. According to dual-process theory of recognition, recollection can occur only if an item is old (Yonelinas, 1994). Similarly, in Experiment 2, we assumed that $P(\mathrm{Rc|new}) = 0$ and $P(\mathrm{Rc|old})$ was a free parameter. However, in Experiment 3, $P(\mathrm{Rc|new})$ was a free parameter to allow the likelihood for new items with R responses to be determined (see Footnote 12).

After modifying Equation A2 for Experiment 2, the relevant likelihood functions are

$$\begin{aligned} L(Z = 6, RT|I) &= [1 - P(Rc|I)][1 - \Phi(C_5|\mu_{Jr|\mathrm{RT},I}, \sigma^2_{Jr|\mathrm{RT}})] \\ &\times \phi(RT|b - s\mu_{\mathrm{p}|I}, \sigma^2_{\mathrm{RT}}) + P(Rc|I)\phi(RT|b - s\mu_{\mathrm{p}|I}, \sigma^2_{\mathrm{RT}}) \end{aligned}$$

for confidence ratings of 6, where $P(\mathrm{Rc|old})$ is a free parameter and $P(\mathrm{Rc|new}) = 0$, and

$$L(Z, RT|I) = [1 - P(Rc|I)][\Phi(C_j|\mu_{Jr|\mathrm{RT},I}, \sigma^2_{Jr|\mathrm{RT}}) - \Phi(C_{j-1}|\mu_{Jr|\mathrm{RT},I}, \sigma^2_{Jr|\mathrm{RT}})] \times \phi(RT|b - s\mu_{\mathrm{p}|I}, \sigma^2_{\mathrm{RT}}), \tag{A5}$$

for confidence ratings 1–5 where $Z = j = 1, \ldots, 5$ and $C_0 = -\infty$.

In Experiment 3, according to the DPSD1 model, R judgments are assumed to be made only if an item is recollected. The relevant likelihood functions are therefore

$$L(RT, R|I) = P(Rc|I)\phi(RT|b - s\mu_{\mathrm{p}|I}, \sigma^2_{\mathrm{RT}}),$$

and

$$L(RT, Z|I) = [1 - P(Rc|I)][\Phi(C_j|\mu_{Jr|\mathrm{RT},I}, \sigma^2_{Jr|\mathrm{RT}}) - \Phi(C_{j-1}|\mu_{Jr|\mathrm{RT},I}, \sigma^2_{Jr|\mathrm{RT}})] \times \phi(RT|b - s\mu_{\mathrm{p}|I}, \sigma^2_{\mathrm{RT}}), \tag{A6}$$

where j = 1–3 and $C_0 = -\infty$, $C_1 = C_G$, $C_2 = C_K$, $C_3 = \infty$. When Z = N, $j = 1$; when Z = G, $j = 2$; and when Z = K, $j = 3$. $P(\mathrm{Rc|old})$ and $P(\mathrm{Rc|new})$ are free parameters.

(Appendices continue)

General Fitting Procedure

For each participant's data, the relevant function from Equations A1–A6 was used to determine the likelihood of every valid trial, given some parameter values. The log-likelihood was summed across all trials and converted to a negative value to be used by a function minimization algorithm (L-BFGS), as implemented in R, a programming language and software environment for statistical computing (R Development Core Team, 2008). Different starting values of the parameters to be estimated were used for the minimization routine in order to maximize the chance of finding the global minimum for the negative log-likelihood for each model (equal to maximizing the log-likelihood). (To get some idea of the variability in the maximum likelihood estimates of the parameter values, the minimization routine was run three times for each data set. There was very little variability in the estimates of the parameters across runs; for example, estimates of $\mu_{r|old}$ and $\mu_{p|old}$ only tended to differ to the third decimal place across runs of each experiment.)

Appendix B

Model Recovery

It is common practice to use the Akaike information criterion (AIC; Akaike, 1973) and Bayesian information criterion (BIC; Schwarz, 1978) to select between models on the basis of how well they fit a data set. In deciding whether to base model selection on the AIC and/or BIC, it is important to ascertain whether a given model would fit a data set better than other models according to these measures, had the model actually generated the data set in the first place. The significance of this is illustrated in the following example. Consider the case where Model A is the true model that generated the data. Another Model B may actually give a better AIC and BIC to this data set than Model A, and may do so, for example, because it is the more flexible model. On the basis of the AIC and BIC, Model B would be incorrectly selected over Model A as the generative model, even though it did not in fact generate the data. In this case, the validity of selecting between models on the basis of the AIC or BIC would be undermined. To determine the validity of the AIC and BIC for selecting between models in our study, we conducted model recovery simulations. In model recovery, artificial data are generated from Model A, and then Models A and B are fitted to those data. The question is whether Model A is correctly identified as the model that generated the data.

For these model recovery simulations, we used a method similar to that of Jang, Wixted, and Huber (2009). For each model, we used the parameter estimates of each participant to simulate artificial test trial data for that participant, where the number of test trials was the same as the number of test trials in the method of the experiment being simulated. These artificial data were then fit by each model (in the same way as the models were fit to real data). The log-likelihood was then summed across participants, and the AIC and BIC were determined for each model. The models were then ranked according to the goodness of the AIC and BIC. A model is recovered if it provides a lower AIC or BIC to its own data than the other models. This method of model evaluation is the same as the method we used to evaluate each model's fit to the real data (i.e., the model with the lowest AIC in Table 2 is preferred). We note at this point that other, more sophisticated methods of model recovery exist (see, e.g., the parametric bootstrap cross-fitting method; Wagenmakers, Ratcliff, Gomez, & Iverson, 2004; see also Navarro, Pitt, & Myung, 2004), but it is not clear how to apply these methods to situations where there are more than two models.

Recall that the multiple-systems-2 (MS2) model collapses to the multiple-systems-1 (MS1) model when $w = 0$, and collapses to the single-system (SS) model when $w = 1$ and $\mu_{r|old} = \mu_{p|old}$ (see Equations 3–5). Recovery of the MS2 model is impossible with these parameter values because the simpler models will always be identified. Accordingly, to ensure that the parameter values used in the model recovery simulations for the MS2 model were most characteristic of the MS2 model (i.e., that $0 < w < 1$ and $\mu_{r|old} \neq \mu_{p|old}$), we excluded participants from the recovery simulations for the MS2 model whose estimate of the w parameter for the real data was $w = 0$ or $w = 1$. In Experiment 1, 19 individuals were excluded from this analysis (nine had $w = 0$, and 10 had $w = 1$). In Experiment 2, 11 individuals were excluded from this analysis (two had $w = 0$, and nine had $w = 1$). In Experiment 3, 11 individuals were excluded (two had $w = 0$, and nine had $w = 1$). In the Amnesia Modeling Study, three individuals from the control (CON) group were excluded (two had $w = 1$, and one had $w = 0$).

The recovery simulation results for Experiments 1–3 are shown in Tables B1–B4 below. In each table a rank of 1 indicates that the model had the best AIC or BIC. As shown in Tables B1 and B2, all models were recoverable with the AIC in each experiment. One exception to this is in Experiment 2, where the MS2 model was not recoverable with the AIC (in this case, the SS model had a better AIC and therefore mimicked the MS2 model). It is worth noting that all models could be recovered by the AIC when we repeated the simulations and increased the number of old–new stimuli in each experiment to 300. As shown in Tables B3 and B4, however, recovery when using the BIC statistic was not as consistent, and the MS2 and dual-process signal detection (DPSD1) models were not recovered with the BIC in any experiment. Recovery of these models with the BIC was only slightly improved by increasing the number of old–new stimuli to 300 (the DPSD1 model could be recovered in Experiment 3).

(*Appendices continue*)

The recovery simulation results for the Amnesia Modeling Study are shown in Tables B5 and B6. With the exception of the MS2 model for the CON data (which was mimicked by the SS model) and the MS1 model for the focal hippocampal lesion (HIP) group data (which was mimicked by the SS model), the models were recovered. When the number of old–new stimuli was increased to 300, all models could be recovered. As with the model recovery simulations for Experiments 1–3 above, the MS1 and MS2 models were not always recovered when the BIC was used.

The recovery simulation results for the fits to the data when pooled across participants and experiments (as reported in Tables 2 and 5) are shown in Tables B7 and B8. There were 17,110 data points (Experiment 1 = 8,934, Experiment 2 = 4,656, Experiment 3 = 3,520). All models were recovered with the AIC. The MS2 model was not recovered when the BIC was used.

The results of these model recovery simulations suggest that in the vast majority of cases, the AIC, but not the BIC, can be used to recover the true generative model in our experiments. This provides support for our use of the AIC as a measure for selecting between models.

Table B1
Experiment 1: Model Recovery: Rank Order of Simultaneous Fit to Simulated Data and Akaike Information Criterion (AIC) Value

True model	SS	MS1	MS2
SS			
Rank	1	3	2
AIC	154674	154851	154753
MS1			
Rank	3	1	2
AIC	154540	154453	154508
MS2			
Rank	2	3	1
AIC	63141	63165	63128

Note. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model.

Table B2
Experiments 2 and 3: Model Recovery: Rank Order of Simultaneous Fit to Simulated Data and Akaike Information Criterion (AIC) Value

	Experiment 2				Experiment 3			
True model	SS	MS1	MS2	DPSD1	SS	MS1	MS2	DPSD1
SS								
Rank	1	4	2	3	1	3	2	4
AIC	86618	86693	86644	86690	65132	65187	65149	66542
MS1								
Rank	3	1	2	4	3	1	2	4
AIC	86721	86693	86713	86907	65143	65109	65135	65966
MS2								
Rank	1	3	2	4	3	2	1	4
AIC	26352	26360	26355	26373	25620	25611	25611	25650
DPSD1								
Rank	2	4	3	1	4	2	3	1
AIC	86597	86671	86622	86556	65336	65270	65284	65227

Note. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model.

(*Appendices continue*)

Table B3
Experiment 1: Model Recovery: Rank Order of Simultaneous Fit to Simulated Data and Bayesian Information Criterion (BIC) Value

True model	SS	MS1	MS2
SS			
Rank	1	2	3
BIC	156527	156910	157018
MS1			
Rank	1	2	3
BIC	156393	156512	156773
MS2			
Rank	1	2	3
BIC	63789	63885	63920

Note. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model.

Table B4
Experiments 2 and 3: Model Recovery: Rank Order of Simultaneous Fit to Simulated Data and Bayesian Information Criterion (BIC) Value

	Experiment 2				Experiment 3			
True model	SS	MS1	MS2	DPSD1	SS	MS1	MS2	DPSD1
SS								
Rank	1	2	3	4	1	2	3	4
BIC	87550	87625	87681	87727	65918	65973	66048	67440
MS1								
Rank	2	1	3	4	2	1	3	4
BIC	87654	87625	87749	87943	65929	65895	66034	66865
MS2								
Rank	1	2	3	4	2	1	3	4
BIC	26591	26599	26620	26638	25880	25871	25908	25946
DPSD1								
Rank	1	3	4	2	2	1	4	3
BIC	87530	87604	87659	87592	66122	66056	66183	66126

Note. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model; DPSD1 = dual-process signal detection model.

Table B5
Conroy et al. (2005) Data: Model Recovery: Rank Order of Simultaneous Fit to Simulated Data and Akaike Information Criterion (AIC) Value

	CON			HIP			MTL		
True model	SS	MS1	MS2	SS	MS1	MS2	SS	MS1	MS2
SS									
Rank	1	3	2	1	2	3	1	3	2
AIC	11565	11578	11567	4450	4452	4454	2966	2986	2969
MS1									
Rank	2	1	3	1	2	3	2	1	3
AIC	11535	11530	11544	4408	4412	4413	2956	2954	2958
MS2									
Rank	1	2	3	2	3	1	2	3	1
AIC	7186	7187	7193	4448	4455	4445	2930	2939	2926

Note. CON = control group; HIP = focal hippocampal lesion group; MTL = medial temporal lobe group; SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model.

(*Appendices continue*)

Table B6
Conroy et al. (2005) Data: Model Recovery: Rank Order of Simultaneous Fit to Simulated Data and Bayesian Information Criterion (BIC) Value

	CON			HIP			MTL		
True model	SS	MS1	MS2	SS	MS1	MS2	SS	MS1	MS2
SS									
Rank	1	2	3	1	2	3	1	3	2
BIC	11744	11757	11781	4502	4504	4517	2997	3017	3006
MS1									
Rank	2	1	3	1	2	3	2	1	3
BIC	11713	11709	11758	4460	4464	4475	2987	2985	2995
MS2									
Rank	1	2	3	1	2	3	1	3	2
BIC	7285	7287	7312	4500	4508	4508	2961	2970	2963

Note. CON = control group; HIP = focal hippocampal lesion group; MTL = medial temporal lobe group; SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model.

Table B7
Pooled Data: Model Recovery: Rank Order of Simultaneous Fit to Simulated Data and Akaike Information Criterion (AIC) Value

True model	SS	MS1	MS2
SS			
Rank	1	3	2
AIC	313034	313150	313036
MS1			
Rank	3	1	2
AIC	312859	312750	312753
MS2			
Rank	2	3	1
AIC	312895	312956	312893

Note. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model.

Table B8
Pooled Data: Model Recovery: Rank Order of Simultaneous Fit to Simulated Data and Bayesian Information Criterion (BIC) Value

True model	SS	MS1	MS2
SS			
Rank	1	3	2
BIC	313104	313219	313114
MS1			
Rank	3	1	2
BIC	312929	312820	312831
MS2			
Rank	1	3	2
BIC	312965	313026	312971

Note. SS = single-system model; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model.

(*Appendices continue*)

Appendix C

Expected Values

Single-System (SS), Multiple-Systems-1 (MS1), and Multiple-Systems-2 (MS2) Models

Once the estimates of the model's parameter values had been derived (see Appendix A), the expected model results were derived analytically. Details for the SS, MS1, and MS2 models are given in Table C1 and below.

In the SS, MS1, and MS2 models, the expected values of the identification reaction times (RTs) conditional on judgment Z are given by the following function (cf. Arnold, Beaver, Groeneveld, & Meeker, 1993, Equation 13a):

$$\lambda(j, I) = b - s\mu_{\mathrm{p}|I} + \frac{sw\sigma_f^2}{\sigma_{J\mathrm{r}}} \frac{\phi\left(\frac{C_j - \mu_{\mathrm{r}|I}}{\sigma_{J\mathrm{r}}}\right) - \phi\left(\frac{C_{j-1} - \mu_{\mathrm{r}|I}}{\sigma_{J\mathrm{r}}}\right)}{\Phi\left(\frac{C_j - \mu_{\mathrm{r}|I}}{\sigma_{J\mathrm{r}}}\right) - \Phi\left(\frac{C_{j-1} - \mu_{\mathrm{r}|I}}{\sigma_{J\mathrm{r}}}\right)}, \quad \text{(C1)}$$

where $\sigma_{J\mathrm{r}} = \sqrt{\sigma_f^2 + \sigma_\mathrm{r}^2}$. For Experiment 1 and the Amnesia Modeling Study, $j = 1$ when Z = N, and $j = 2$ when Z = O; $C_0 = -\infty$, $C_1 = C$, and $C_2 = \infty$. For Experiment 2, $j = Z = 1, \ldots, 6$; $C_0 = -\infty$; and $C_6 = \infty$. In Experiment 3, $j = 1$ when Z = N, $j = 2$ when Z = G, $j = 3$ when Z = K, $j = 4$ when Z = R, and $C_0 = -\infty$, $C_1 = C_\mathrm{G}$, $C_2 = C_\mathrm{K}$, $C_3 = C_\mathrm{R}$, $C_4 = \infty$.

Thus, for Experiment 1 and the Amnesia Modeling Study, Equation C1 gives the expected identification reaction time (RT) of hits (H; E[RT|H]) when I = old and Z = O; it gives the expected RT of false alarms (F; E[RT|F]) when I = new and Z = O. Similarly, Equation C1 gives the expected RT of misses (M; E[RT|M]) when I = old and Z = N; and gives the expected RT of correct rejections (CR; E[RT|CR]) when I = new and Z = N. In Experiments 2 and 3, the expected RTs for hits, false alarms, misses, and correct rejections are also given via the same procedure as in Experiment 1 and the Amnesia Modeling Study, but replacing $C_1 = C$ with $C_1 = C_3$ (for Experiment 2) or $C_1 = C_\mathrm{G}$ (for Experiment 3).

In the data, because the mean identification RTs for items judged old or new are weighted means, the expected RTs are given by the weighted expected RTs to hits and false alarms (items judged old), or misses and correct rejections (items judged old). Hence

$$E[RT|Z = 0] = \frac{P(H)E[RT|H] + P(F)E[RT|F]}{P(H) + P(F)}$$

and

$$E[RT|Z = N] = \frac{[(1 - P(H)]E[RT|M] + [1 - P(F)]E[RT|CR]}{2 - P(H) - P(F)}. \quad \text{(C2)}$$

The overall fluency effect can be calculated as E[RT|Z = N] − E[RT|Z = O].

Table C1
Recognition and Priming Measures and Their Expected Values Under the Single-System, Multiple-Systems-1, and Multiple-Systems-2 Models

Measure	Model expected value
P(H)	$1 - \Phi(C - \mu_{\mathrm{r|old}})$
P(F)	$1 - \Phi(C)$
d'	$\mu_{\mathrm{r|old}}$
E[RT|new]	b
E[RT|old]	$b - s\mu_{\mathrm{p|old}}$
Overall priming	$s\mu_{\mathrm{p|old}}$

Note. $C = C_3$ in Experiment 2, and $C = C_\mathrm{G}$ in Experiment 3. H = hit; F = false alarm; RT = reaction time.

Dual-Process Signal Detection (DPSD1) Model

Details of the expected values for the DPSD1 model are given in Table C2 and below.

For Experiment 2, Equation C1 can be used to determine the identification RTs for the familiarity-based (f) recognition responses (Ratings 1–5) when $j = Z = 1, \ldots, 5$ and $C_0 = -\infty$. Old items receiving a 6 rating are based on familiarity and recollection: For the familiarity-based 6 responses, Equation C1 can be used when $j = Z = 6$ and $C_6 = \infty$. The expected RT of recollected 6 items is equal to E[RT|old] $= b - s\mu_{\mathrm{p|old}}$. The expected identification RTs for 6 responses is therefore given by the average of the familiarity- and recollection-based expected RTs, weighted by the proportion of expected familiarity- and recollection-based 6 responses, that is,

$$E[RT|Z = 6, old] = \frac{[1 - P(Rc|old)][1 - \Phi(C_5 - \mu_{\mathrm{r|old}})]\lambda(6, old) + P(Rc|old)(b - s\mu_{\mathrm{p|old}})}{[1 - P(Rc|old)][1 - \Phi(C_5 - \mu_{\mathrm{r|old}})] + P(Rc|old)},$$

(*Appendices continue*)

Table C2
Recognition and Priming Measures and Their Expected Values Under the Dual-Process Signal Detection Model

Measure	Expected model value
P(H)	$[1 - P(\text{Rc}\mid\text{old})][1 - \Phi(C - \mu_{\text{r}\mid\text{old}})] + P(\text{Rc}\mid\text{old})$
P(F)	$[1 - P(\text{Rc}\mid\text{new})][1 - \Phi(C)] + P(\text{Rc}\mid\text{new})$
d'	$z(P(\text{H})) - z(P(\text{F}))$
E[RT|new]	b
E[RT|old]	$b - s\mu_{\text{p}\mid\text{old}}$
Overall priming	$s\mu_{\text{p}\mid\text{old}}$

Note. $C = C_3$ in Experiment 2, and $C = C_G$ in Experiment 3. In Experiment 2, $P(\text{Rc}\mid\text{new}) = 0$. The quantile function for the standard normal distribution is represented by z. H = hit, F = false alarm.

where λ is the function defined in Equation C1 and $j = Z = 6$, $C_5 = C_5$, $C_6 = \infty$.

Equation C1 can also be used for the expected RTs for N, G, and K responses in Experiment 3 when $j = 1, \ldots, 3$ and $C_0 = -\infty$, $C_1 = C_G$, $C_2 = C_K$, $C_3 = \infty$; and $j = 1$ when $Z =$ N; $j = 2$ when $Z =$ G; and $j = 3$ when $Z =$ K. The expected identification RTs for old and new items receiving R responses are $E[\text{RT}\mid\text{old}] = b - s\mu_{\text{p}\mid\text{old}}$, and $E[\text{RT}\mid\text{new}] = b$, respectively.

In the DPSD1 model, the expected values of the identification RTs for hits (E[RT|H]) and false alarms (E[RT|F]) are given by the following:

$$E[RT\mid Z = O, I] = \frac{[1 - P(Rc\mid I)][1 - \Phi(C_{\text{ON}} - \mu_{\text{r}\mid I})]\lambda(2, I) + P(Rc\mid I)(b - s\mu_{\text{p}\mid I})}{[1 - P(Rc\mid I)][1 - \Phi(C_{\text{ON}} - \mu_{\text{r}\mid I})] + P(Rc\mid I)},$$

where λ is the function defined in Equation C1 and $j = 2$ ($Z =$ O), $C_1 = C_3$ (Experiment 2) or $C_1 = C_G$ (Experiment 3), $C_2 = \infty$, and $I =$ old, new.

The expected identification RTs for misses (E[RT|M]) is given by Equation C1 when $j = 1$ ($Z =$ N) and $I =$ old; the expected identification RTs for correct rejections (E[RT|CR]) is given by Equation C1 when $j = 1$ ($Z =$ N) and $I =$ new; in both cases, $C_0 = -\infty$ and $C_1 = C_3$ (for Experiment 2) or C_G (for Experiment 3). Equation C2 can then be used to calculate the expected overall fluency effect.

Appendix D

Instructions Presented to Participants for Responding With Remember–Know Judgments

Instructions that were presented to participants for responding with remember–know judgments, adapted from Gardiner and Richardson-Klavehn (2000):

> Recognition memory is associated with two different kinds of awareness. Sometimes when you recognize a word on the test list as one from the first stage, recognition will bring back to mind something you remember thinking about when the word appeared then (on the first session list). You recollect something you consciously experienced at that time. In a case like this, select the REMEMBER OLD response (key 1).
>
> But sometimes recognizing a word as one you saw during the first session will not bring back to mind anything you remember about seeing it then.
>
> Instead, the word will seem familiar, so that you feel confident it was the one you saw before, even though you don't recollect anything you experienced when you saw it then. Select the KNOW OLD response (key 2) in a case when recognition is accompanied by strong feelings of familiarity in the absence of any recollective experience.
>
> There will also be times when you do not remember the word, nor does it seem familiar, but you might want to guess that it was one of the words you saw during the first stage. Select the GUESS OLD response (key 3) if your response is really just a guess.
>
> If you think a word hasn't been presented before then select the NEW response (key 4).
>
> DO NOT PRESS RETURN UNTIL YOU HAVE UNDERSTOOD THESE INSTRUCTIONS. CHECK WITH THE EXPERIMENTER IF YOU ARE UNSURE.
>
> Press RETURN to continue.

(*Appendices continue*)

Appendix E

Additional Model Fitting Results

Table E1
Fits to Individual Data

Model	Experiment 1 (N = 32)			Experiment 2 (N = 16)			Experiment 3 (N = 18)		
	p	ln(L)	AIC	p	ln(L)	AIC	p	ln(L)	AIC
s free									
SS	9	−74738	**150053**	9	−41949	**84187**	7	−30203	**60658**
MS1	11	−74737	150178	10	−41987	84293	8	−30231	60750
MS2	12	−74697	150162	11	−41943	84238	9	−30193	60709
s fixed									
SS	8	−74772	**150056**	8	−41983	84221	6	−30228	**60672**
MS1	10	−74737	150114	9	−41989	84266	7	−30231	60714
MS2	11	−74703	150111	10	−41947	**84214**	8	−30198	60684

Note. A bold value indicates the model that fit the data the best according to the Akaike information criterion (AIC); p is the number of parameters per participant; each ln(L) value is the sum of the ln(L) across participants. The AIC can be determined with the formulas given in Table 2, for which the total number of free parameters is given by $p \times N$, where N is the number of participants modeled in each experiment. For the "*s* free" model fits, *s* was free for each participant in each model. (The SS model results when *s* is free are the same as those in Table 2, and are presented here for comparison.) For the "*s* fixed" model fits, the value of *s* was fixed for every participant in every model to the mean *s* across participants in Experiments 1–3, when each participant was fit by the single-system (SS) model (this value was $s = 143.6$). L = maximum likelihood; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model.

Table E2
Fits to Data Aggregated Across Participants

Model	Experiment 1 (N = 32)			Experiment 2 (N = 16)			Experiment 3 (N = 18)		
	p	ln(L)	AIC	p	ln(L)	AIC	p	ln(L)	AIC
s free									
SS	9	−79611	159241	9	−44002	**88021**	7	−31171	**62356**
MS1	11	−79594	**159211**	10	−44015	88051	8	−31202	62421
MS2	12	−79594	159212	11	−44001	88024	9	−31170	62358
s fixed									
SS	9	−79611	159241	8	−44002	**88021**	6	−31174	**62360**
MS1	10	−79594	**159209**	9	−44015	88049	7	−31202	62419
MS2	11	−79594	159210	10	−44001	88022	8	−31173	62362

Note. A bold value indicates that the model that fit the data the best according to the Akaike information criterion (AIC); p denotes the number of free parameters used to model the data. For the "*s* fixed" model fits, *s* was fixed to the estimate of *s* when the single-system (SS model) was fit to the aggregated data of Experiment 1. L = maximum likelihood; MS1 = multiple-systems-1 model; MS2 = multiple-systems-2 model.

Received March 4, 2011
Revision received July 20, 2011
Accepted August 8, 2011 ■