

Assessing the Dissociability of Recollection and Familiarity in Recognition Memory

Michael S. Pratte
Vanderbilt University

Jeffrey N. Rouder
University of Missouri

Recognition memory is often modeled as constituting 2 separate processes, recollection and familiarity, rather than as constituting a single process mediated by a generic latent strength. One way of stating evidence for the more complex 2-process model is to show dissociations with select manipulations, in which one manipulation affects recollection more than the second and the second affects familiarity more than the first. One of the best paradigms for assessing such dissociations is the confidence-ratings paradigm, because within it criterial and mnemonic effects may be separately estimated. There is, unfortunately, a relative lack of easily interpretable dissociation experiments in the confidence ratings paradigm, and those that exist do not show clear evidence for dissociations. We report the results of 3 experiments with conventional manipulations designed to maximally dissociate recollection and familiarity. To provide valid assessment, without recourse to aggregation over items or participants, we develop a hierarchical version of Yonelinas' (1994) dual-process model and a novel test of dissociability for state-trace plots. The data do not provide evidence that recollection and familiarity are dissociable. Instead, estimates of recollection and familiarity are positively related across conditions and experiments. On balance, performance is more parsimoniously accounted for by a single mnemonic process that drives both recollection and familiarity estimates.

Keywords: recognition memory, dual-process memory models, memory models, state trace, hierarchical Bayesian analysis

Over the past 30 years psychologists have conceptualized memory as consisting of several distinct systems or processes. If one defines memory broadly, this approach has much value, as it is highly likely that different memory processes are involved in disparate tasks. For instance, it seems plausible that maintaining a grocery list in mind taps different mnemonic processes than learning to ride a bike. Modern memory research focuses on delineating various memory processes, understanding interactions between processes, and understanding the physiological basis for these processes (Schacter & Tulving, 1994).

One popular dichotomy in human memory is the distinction between unconscious, automatic retrieval and conscious, deliberate recollection (Mandler, 1980). This dichotomy has served as the lens through which a large number of mnemonic phenomena are understood. For example, memory loss due to aging or due to hippocampal damage is widely considered to affect conscious recollection while largely sparing automatic retrieval (e.g., Light, Prull, La Voie, & Healy, 2000; Parkin & Leng, 1993; Yonelinas, 2002). Another example of this distinction can be found in social psychology, where stereotype information is largely assumed to

tap automatic retrieval rather than conscious recollection (e.g., Bargh, 1997; Correll, Park, Judd, & Wittenbrink, 2002). In fact, in today's literature it is hard to find mnemonic domains that are not informed by the conscious recollection versus automatic retrieval dichotomy.

Although the multiple-process view is widely accepted in cognitive neuroscience and is certainly applicable across diverse domains, it remains controversial among memory theorists (e.g., Dunn, 2008; Heathcote, 2003; Heathcote, Raymond, & Dunn, 2006; Wixted, 2007). The main controversy is whether verbal learning tasks, such as learning a list of anatomy terms, are necessarily mediated by multiple processes or are mediated by a single process. In verbal learning there is a long history of considering memory as a single univariate construct called *memory strength* (Kintsch, 1967). Accordingly, each item has a single scalar strength associated with it, and the act of studying an item increases its strength. At test, participants recognize items as being studied only if the memory trace has sufficient strength. These strength models, which are conventionally considered single-process models, are quite parsimonious because memory is modeled as affecting a scalar variable rather than multiple systems. It is doubtful that such a simple account can explain the vast variety of tasks commonly subsumed under memory or learning. Advocates of strength-based approaches, however, view these as parsimonious accounts for everyday verbal learning tasks such as trying to memorize a grocery list, recalling who said what in a distant conversation, or mastering a collection of anatomy terms.

Even in the domain of everyday explicit learning, there is a large corpus of claims for the dual-process view based on behavioral results (e.g., Jacoby, 1991; Tulving, 1985; Yonelinas, 1994), brain

This article was published Online First May 7, 2012.

Michael S. Pratte, Department of Psychological Sciences, Vanderbilt University; Jeffrey N. Rouder, Department of Psychological Sciences, University of Missouri.

This work was carried out with support from National Science Foundation Grant SES-1024080.

Correspondence concerning this article should be addressed to Michael S. Pratte, Department of Psychological Sciences, 301 Wilson Hall, Vanderbilt University, Nashville, TN 37240. E-mail: prattms@gmail.com

imaging results (e.g., Rugg & Curran, 2007; Yonelinas, Otten, Shaw, & Rugg, 2005), and neuropsychological results with patients (e.g., Howard, Bessette-Symons, Zhang, & Hoyer, 2006; Vann et al., 2009; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998). A complete critical review of all of these claims is well outside the scope of this paper, and here we concentrate on the question of whether there is behavioral support for the dual-process view.

Our experimental logic is straightforward and conventional—we perform empirical manipulations targeted to affect one memory system more so than the other. For example, changing perceptual features between study and test, such as font, may primarily affect familiarity. Conversely, enhancing the semantic meaning of items at study may primarily affect recollection. To assess how these manipulations affect processes, we evaluate the corresponding parameters of recollection and familiarity with Yonelinas' (1994) dual-process signal detection model. Although this approach is standard and seemingly straightforward, there are a number of methodological and conceptual difficulties that have been unappreciated in previous research. Our approach here is characterized by detailed and careful attention to statistical methodology, and we provide three innovations in analysis that substantially improve inference over previous studies. First, we use the logic of state-trace analysis to assess data patterns that extend across multiple conditions and experiments. Second, we use a hierarchical Bayesian approach to assess psychological process uncontaminated by nuisance variability from participants and items. Third, we implement a novel contrast, which we call the *slope test*, to avoid a multiple-test artifact in making statistical claims with state-trace data. These innovations lead to a more rigorous and principled assessment than previously available. The data we obtain, if conventionally interpreted, seemingly show that changing perceptual features between study and test selectively influences familiarity as distinct from recollection, while enhancing semantic meaning affects both recollection and familiarity. To our knowledge, this selective influence demonstration of perceptual effects on familiarity is the first of its kind in a confidence-ratings receiver operating characteristic (ROC) task. However, when we use our more principled and rigorous methodology, the evidence is far less persuasive and, on balance, is more favorable to a single-process interpretation than a dual-process interpretation.

State-Trace Plots

The main approach taken herein is to study the effects of manipulations that are targeted to affect recollection or familiarity, say semantic and perceptual enhancement, respectively. We refer to these manipulations as the *recollection-enhancement manipulation* and *familiarity-enhancement manipulation* respectively, though giving a manipulation such a name by no means guarantees that it affects the targeted process. To assess whether these manipulations do indeed affect targeted processes, we adopt a state-trace representation of parameter variation (Bamber, 1979; Dunn, 2008). In a state-trace plot, parameters that measure one process are plotted as a function of parameters that measure another, and in Figure 1A, recollection is plotted as a function of familiarity. Consider the case where the recollection- and familiarity-enhancement manipulations are each manipulated through two levels and crossed. We generically refer to the levels on each factor

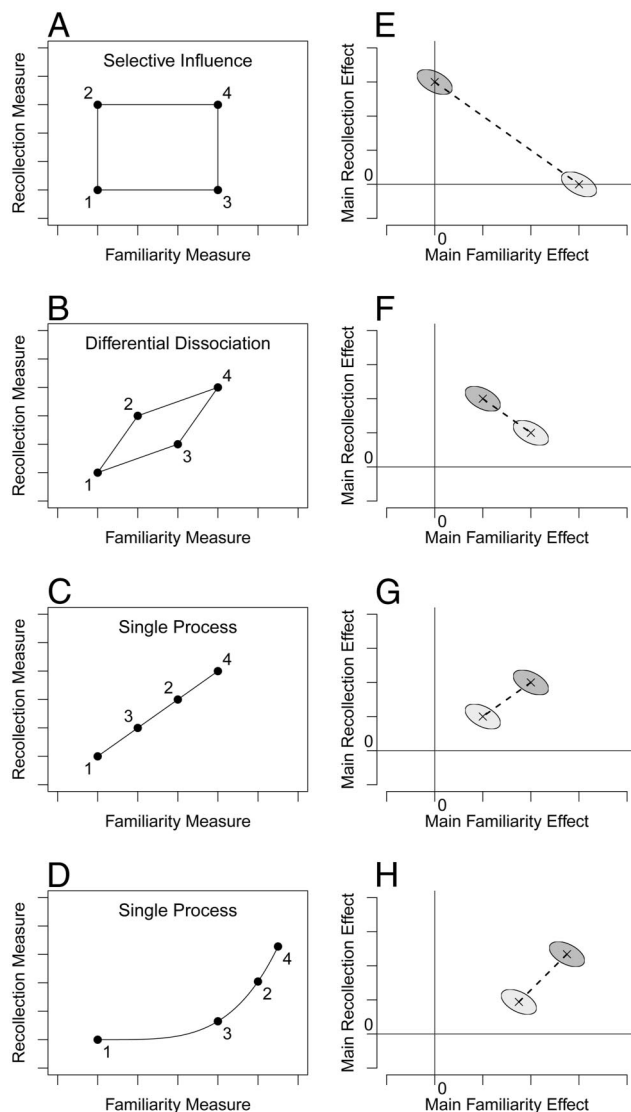


Figure 1. Assessing state-trace plots. The left column shows diagnostic patterns. The points labeled 1 through 4 in these plots are familiarity and recollection estimates from four conditions: (1) low recollection and low familiarity enhancement; (2) high recollection enhancement and low familiarity enhancement; (3) low recollection enhancement and high familiarity enhancement; (4) high recollection and high familiarity enhancement. A: A double selective influence pattern is indicative of two processes. B: A differential dissociation pattern is indicative of two processes. C: A straight-line pattern is indicative of one process. D: A curved pattern is indicative of one process. The right column shows the corresponding recollection and familiarity main effects. The ellipses indicate 95% confidence regions in the joint estimates; the darker and lighter ellipses are for the recollection-enhancement and familiarity-enhancement manipulations, respectively. The dashed lines connect the ellipses, and the slopes of these lines serve as the critical test statistic in the *slope test*. E–H: Main effects from the selective influence, differential influence, and the two single-process patterns, respectively. The sign of the slope is indicative of the number of processes.

as “high” (meaning much enhancement) and “low” (meaning no enhancement). Point 1 in Figure 1 denotes parameter values when the level of both factors is low (i.e., there is neither targeted

recollection nor familiarity enhancement). Point 2 denotes the case where the recollection-enhancement factor is high and the familiarity factor is low; Point 3 denotes the reverse assignment. Finally, Point 4 denotes the case where levels of both manipulations are high.

The pattern in Figure 1A demonstrates a *double selective influence*, in which the recollection-enhancement manipulation enhances recollection and not familiarity, and vice versa. The pattern in Figure 1A, if observed, would provide exceedingly strong evidence for the separability of familiarity and recollection.

Figure 1B shows another pattern that is indicative of two separable processes: a *differential dissociation* (Dunn & Kirsner, 1988). Here, selective influence does not hold, as both manipulations affect both parameters. The defining feature of the pattern is that the recollection-enhancement manipulation affects recollection more than the familiarity-enhancement manipulation, and the familiarity-enhancement manipulation affects familiarity more than the recollection-enhancement manipulation.

Figures 1C and 1D provide examples of state-trace patterns that fail to provide evidence of two processes. In both figures, the parameter values lie on a smoothly increasing line. As memory gets better, only a single dimension, that defined by the line, is needed to describe the pattern of parameter variation. One may think of the position on this line as memory strength in a single-process model. For the displayed patterns, both recollection and familiarity estimates may be expressed as a 1-to-1 function of this underlying memory strength. In general, if points in state-trace space can be accounted for with a smoothly increasing (or decreasing) function, the single-process interpretation is more parsimonious than a two-process interpretation.

The critical comparisons in state-trace analysis are whether the recollection-enhancement manipulation affects recollection more than the familiarity-enhancement manipulation and whether the familiarity-enhancement manipulation affects familiarity more than the recollection-enhancement manipulation. We avoid statements such as "Manipulation X affects recollection more than familiarity" because of fundamental conceptual difficulties in interpretation. Recollection and familiarity effects in the Yonelinas model are measured on different scales with different units and with different ranges. Consequently, the implied subtraction in the direct contrast of recollection and familiarity is not meaningful. As an alternative, some researchers have transformed the parameters before subtraction, for example by comparing percentage change in one parameter subtracted from percentage change in the other (e.g., Yonelinas, 2002). Such transforms, while seemingly natural, still imply a linear relationship between the logarithm of parameters. The advantage of state-trace analysis, in contrast, is that one need not make any assumptions about an appropriate scale for subtraction. Because the critical comparisons are ordinal in nature, the approach we present is not reliant on scaling assumptions. Any monotonic transformations of the variables will not change the ordering of effects on a dimension, and so conclusions from state-trace patterns, and the test we develop herein, are invariant to the parameterization of familiarity and recollection.

Although state-trace plots such as those in Figure 1 suggest certain processing constraints, they are not foolproof. Difficulties in interpretation may result when the quantities plotted on the axes do not measure what is assumed. An example of these difficulties is provided by reanalysis of Tulving's (1985) *remember-know*

paradigm. Many researchers have proposed that remember responses index recollection more so than familiarity and that know responses index familiarity more so than recollection. One finding in this paradigm is a differential dissociation in which manipulations thought to affect recollection affect the proportion of remember responses more so than know responses, and manipulations thought to affect familiarity affect the proportion of know responses more so than remember responses (see Gardiner & Richardson-Klavehn, 2000, for a review). Figure 2A shows a state-trace plot of data from Gardiner (1988) and Gregg and Gardiner (1994). The patterns in this figure imply two processes, and according to dual-process theorists, both of these processes are mnemonic in nature.

Critics of Tulving's interpretation of the remember-know paradigm note that remember and know responses may be viewed as different criterial settings on a common latent-strength axis (Donaldson, 1996; Dunn, 2004, 2008; Hirshman & Henzler, 1998; Wixted & Stretch, 2004). The open points in Figure 2A come from the signal detection model depicted in Figure 2B. The familiarity-enhancement condition reflects primarily criterial changes, while the recollection-enhancement condition reflects primarily sensitivity changes. Hence, the state-trace plot provides an indication that there are in fact two factors in processing, but it cannot distinguish whether these factors are two separate memory processes or simply the distinction between sensitivity and criteria.

Recollection and Familiarity Enhancement in Recognition Memory

One consequence of the above example with the remember-know paradigm is that researchers are on firmer ground if they use paradigms and models that separate mnemonic and criterial effects. Our desire to both separate mnemonic processes from criteria shifts and decompose the mnemonic effects into recollection and familiarity motivates our choice of the confidence-ratings paradigm and the Yonelinas (1994) dual-process model parameter measures. In the confidence-ratings paradigm, each response option is presumed to correspond to a different criterial setting. The Yonelinas model is useful because, when applied to confidence ratings, it provides separate estimates of recollection, familiarity and criteria. State-trace plots of these parameters across selected manipulations are used herein to assess whether recognition memory is mediated by one or two processes.

Recognition memory has been a staple phenomenon in understanding the nature of mnemonic processing, and there is a wealth of literature on the effects of various manipulations. Many studies, however, contain only one factor that is purported to affect one process (often recollection) more than the other. Salient examples of this approach include Yonelinas et al.'s (2002) study of the mnemonic consequences of temporal lobe damage; Curran, DeBuse, Woroch, and Hirshman's (2006) study of the mnemonic consequences of the amnesia-inducing drug Midazolam; and Howard et al.'s (2006) study of the mnemonic consequences of aging. In these studies, researchers measured dual-process parameters and found that the target group (patients, drug-exposed group, older adults) had diminished recollection but largely preserved familiarity compared to appropriate controls. This data pattern could be accounted for by Points 3 and 4 in Figure 1B, Points 2 and 4 in Figure 1C, or Points 3 and 4 in Figure 1D. Because these patterns

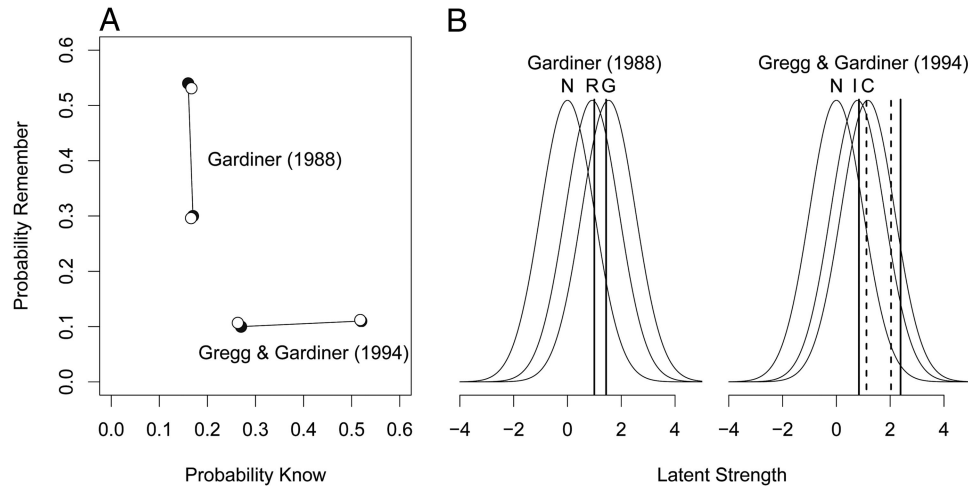


Figure 2. Difficulty in interpreting state-trace plots in the remember-know-new paradigm. **A:** Selective influence from the conjunction of Gardiner's (1988) levels-of-processing manipulation and Gregg and Gardiner's (1994) perceptual match-mismatch manipulation. Open points are predictions from the single-process account. **B:** The equal-variance signal-detection model predicts the differential dissociation nearly exactly. The effect of processing depth in Gardiner (1988) is in sensitivity (N = new, R = read items at study, G = generate related items at study). The region between criteria corresponds to *know* responses; the region below both criteria corresponds to *new* responses; the region above them corresponds to *remember* responses. The effect of perceptual match between study and test in Gregg and Gardiner (1994) marginally increases sensitivity (N = new, I = perceptual mismatch, C = perceptual match) and more substantively affects criteria. Dashed and solid lines denote criteria for the match and mismatch conditions, respectively. An increase in *know* responses for the mismatch condition is accounted for by this difference in criteria.

are based on a single variable (age status, patient status), it is difficult to tell whether they are accounted for better by one or by more than one process. In summary, experiments that report the result of a single manipulation or a single group difference provide ambiguous evidence for addressing whether there is one or more mnemonic processes.

To be fair, the above studies are not meant to be definitive demonstrations of two processes but to serve as partial evidence within a larger picture. Nonetheless, there are difficulties in combining them to form this larger picture. Let's take the Howard et al. (2006) result that there is a significant age-related difference in recollection but not in familiarity. Let's also consider one of the results we report subsequently in Experiment 1 that there is a main effect of perceptual match on familiarity but no corresponding effect on recollection. On the surface, the conjunction of these two findings would seem to provide strong evidence for two processes. Although such evidence should receive consideration, there are two difficulties in interpretation. First, our perceptual match effects and Howard et al.'s age effects may happen at different baselines. Figure 1D shows the potential importance of having controlled baselines. If there is curvature in the relationship between recollection and familiarity and baselines vary, differential effects are expected even if memory is governed by a single process. As in the case of the figure, familiarity effects are more likely for poor-performance baselines, and recollection effects are more likely for high-performance baselines. Fortunately, this difficulty can be mitigated either by using common baselines (as we do in the crossed design) or by basing analysis on state-trace plots.

A second difficulty is that the implied double selective influence is based on the conjunction of four significance tests, two of which

are conceptually difficult statements of failures to reject the null (the preservation of familiarity across age and the preservation of recollection across perceptual match conditions). These test results are critically dependent on having an accurate measure of sample variability of process estimates. As is discussed in the next section, we believe researchers have understated the variability in recollection and familiarity estimates because the rates used in constructing these contrasts are based on data aggregated across items, participants or both. When variability is understated, there is little control of either Type I or Type II error rates. Consequently, it is difficult to interpret both rejections and failure-to-reject outcomes.

The Aggregation Problem

Almost every study in memory is based on aggregated data. In memory tasks, there is not only variability in the cognitive system (e.g., that due to mixing familiarity and recollection processes) but also nuisance variability across disparate people and items. Researchers often account for variability across people by computing participant-specific hit and false alarm rates in analysis. To do so, however, they must aggregate data across items even though some items are better remembered than others. Likewise, in cases where researchers compute item-specific rates, data are aggregated over participants even though some participants have better memories than others. In sum, aggregation almost always happens, and it seems necessary to construct empirical hit and false alarm rates.

We have been critical of aggregation in memory contexts (Morey, Pratte, & Rouders, 2008; Pratte & Rouders, 2011; Pratte, Rouders, & Morey, 2010; Rouders & Lu, 2005; Rouders, Lu, Morey, Sun, & Speckman, 2008) because parameters estimated from ag-

gregated data do not, generally, match the average of parameters estimated from disaggregated data. Our experience with the unequal-variance signal detection model is instructive. In Pratte et al. (2010) we fit a hierarchical version of the unequal-variance signal detection model without recourse to aggregation across either participants or items. When we compared our results to those from data aggregated across participants, items, or both, two important trends stood out.

First, the average of parameters was different than the parameters estimated from aggregated data. Sensitivity, for example, was overestimated by 16% with aggregation. Likewise, the asymmetry in ROCs, as measured by increases in the variance parameter of an unequal-variance signal detection model over a baseline of 1.0, was underestimated by 27%. Because variance is an index of asymmetry, analyses of aggregated data underestimate the degree of asymmetry in ROC curves. In the current context, aggregation will lead to an underestimation of recollection, as recollection is positively related to ROC asymmetry in the dual-process model.

Second and perhaps more important, our analysis indicates that researchers relying on aggregated data may dramatically overstate confidence in parameter estimates. We routinely found that the confidence intervals surrounding estimates from aggregated data were too small, often by alarming factors. For example, confidence intervals on sensitivity estimates from data aggregated across items were underestimated by 15% when compared to the model that takes item and participant variation into account. Confidence intervals were underestimated by 60% when data were aggregated over participants but not items, and the underestimation rises to a staggering 80% when the aggregation is over both items and participants. This conclusion, that aggregated data may dramatically understate variability, is consistent with the seminal concerns of Clark (1973) and is discussed at length in Rouder and Lu (2005). In the current context, aggregated data lead to inflated Type I errors when effects are observed and overconfidence in null effects when they are not. These facts greatly complicate the interpretation of previous results.

A Hierarchical Model Solution

Our approach to assessing cognitive structure without contamination from aggregation is to construct hierarchical nonlinear regression models that explicitly account for participant and item variability (Rouder & Lu, 2005). We have successfully used this approach to model recognition memory (e.g., Morey et al., 2008; Pratte et al., 2010; Rouder, Lu, et al., 2007), as well as in other domains (e.g., Morey, Rouder, & Speckman, 2009; Rouder, Morey, Speckman, & Pratte, 2007; Rouder, Tuerlinckx, Speckman, Lu, & Gomez, 2008). In Pratte and Rouder (2011) we provide a hierarchical version of Yonelinas' dual-process model. In this development, the effects of the mnemonic structure are modeled separately from contaminant variation from people and items. Below we provide a summary and some extensions of the Pratte and Rouder hierarchical model.

Methods of estimation are provided in Pratte and Rouder (2011), and software for fitting the model is available in the R package *hbm*.

The hierarchical model is based on Yonelinas' (1994) dual-process model. This model specifies that responses reflect either successful recollection or, if recollection fails, evidence from fa-

miliarity. Familiarity is modeled as an equal-variance signal-detection process. In the hierarchical version, there are separate recollection and familiarity parameters for each subject-by-item-by-condition combination. In addition, these parameters depend on study-test *lag*, the number of study and test trials that intervene between study and test for a given item. Recollection is denoted by R_{ijkl} and familiarity by d'_{ijkl} , where $i = 1, \dots, I$ indexes participants, $j = 1, \dots, J$ indexes items, $k = 1, \dots, K$ indexes conditions (e.g., levels of a level-of-processing manipulation), and $l = 1, \dots, L$ indexes lag. Of course, there are far too many parameters in the model, and constraint is needed. In particular, because participant-by-item combinations are never replicated in a memory experiment, participant-by-item interactions would be impossible to estimate. We use additive main-effects models to provide constraint and have shown that these additive models are highly appropriate even when participant-by-item interactions are present at reasonable levels (Pratte et al., 2010). Recollection is given as

$$R_{ijkl} = G[\mu_k^{(r)} + \alpha_i^{(r)} + \beta_j^{(r)} + \gamma^{(r)}(L_l - L_0)],$$

where $\mu_k^{(r)}$ is the k th condition effect, $\alpha_i^{(r)}$ is the i th participant effects, and $\beta_j^{(r)}$ is the j th item effect on recollection. Parameter $\gamma^{(r)}$ is the slope of the lag effect for the L_l lag; L_0 is the mean lag across the experiment such that the sum of $(L_l - L_0)$ is zero. The function G indicates that there is a link between the additive components and recollection. Such a link is necessary because recollection is a probability, constrained to the zero-one interval, and the additive components are real valued. We use the probit link (inverse cumulative distribution function of a standard normal) throughout, as this choice is both flexible and convenient.

Further constraint is gained by imposing hierarchical structures on participant and item effects:

$$\alpha_i^{(r)} \sim \text{Normal}(0, \sigma_{\alpha^{(r)}}^2), \quad (1)$$

$$\beta_j^{(r)} \sim \text{Normal}(0, \sigma_{\beta^{(r)}}^2). \quad (2)$$

The variances of these distributions are free parameters that reflect the degree of participant and item variability. These structures will pull large effects toward zero, mitigating the influence of outliers. Hierarchical models generally provide for more accurate estimation, and we have shown that they improve the fit of the dual-process model over letting participant and item effects be unconstrained (Pratte & Rouder, 2011).

Familiarity in the hierarchical model is split into two components: a baseline familiarity, $d^{(n)}$, denoting familiarity of words before study, and a studied-item familiarity, $d^{(s)}$, denoting the familiarity after study. To scale the latent strength axis, one sets the middle criteria to 0, and this setting is made without any loss of generality (see Figure 3). All other criteria are free to vary across participants, reflecting individual preferences for particular responses over others. Baseline familiarity $d_{ijk}^{(n)}$ is modeled as additive components:

$$d_{ijk}^{(n)} = \mu_k^{(n)} + \alpha_i^{(n)} + \beta_j^{(n)},$$

where $\mu_k^{(n)}$, $\alpha_i^{(n)}$, and $\beta_j^{(n)}$ are the condition, participant, and item effects on baseline familiarity, respectively. There is no lag effect, as baseline familiarity effects reflect pre-experimental familiarity. Hierarchical normal structures, analogous to (1) and (2), are placed on participant and item baseline familiarity effects.

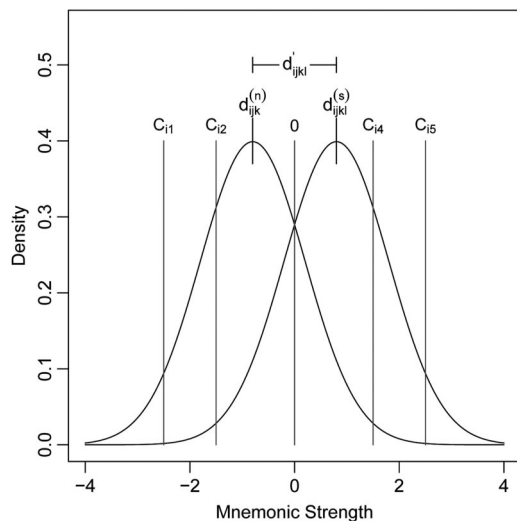


Figure 3. Familiarity component of the hierarchical dual-process model. The middle criterion is fixed to zero, and the means of both the new and the studied-item distributions are free parameters. The difference between these means is denoted by d' .

In Pratte and Rouder (2011), we placed additive components on studied-item familiarity $d^{(s)}$:

$$d_{ijkl}^{(s)} = \mu_k^{(s)} + \alpha_i^{(s)} + \beta_j^{(s)} + \gamma^{(s)}(L_l - L_0),$$

where $\mu_k^{(s)}$, $\alpha_i^{(s)}$, $\beta_j^{(s)}$, and $\gamma^{(s)}$ are condition, participant, item, and lag effects on studied-item familiarity, respectively, and where hierarchical normal structures, analogous to (1) and (2), are placed on participant and item familiarity effects. A flaw with this approach is that it is possible for condition-by-participant-by-item combinations to yield negative effects of study (i.e., that $d_{ijkl}^{(s)} < d_{ijk}^{(n)}$). We have found, however, that a more accurate and parsimonious model may be achieved by placing participant and item effects on d' , the increase in familiarity due to study. Let d'_{ijkl} be the increase in familiarity due to study for the i th subject tested on the j th item in the k th condition with the l th level of lag, such that $d_{ijkl}^{(s)} = d_{ijk}^{(n)} + d'_{ijkl}$. Parameter d'_{ijkl} may be assumed to be nonnegative; study should not make items less memorable. To ensure positivity, we use an exponential link:

$$d'_{ijkl} = \exp[\mu_k' + \alpha_i' + \beta_j' + \gamma'(L_l - L_0)],$$

where μ_k' , α_i' , β_j' , and γ' are the condition, participant, item, and lag effects on sensitivity, respectively. Hierarchical normal structures, analogous to (1) and (2), are placed on participant and item effects on d' .

This new model, with multiplicative effects on d' , may be compared with our previous hierarchical dual-process model using the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002). DIC is similar to the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) in that it explicitly takes into account both model fit and parsimony. The advantage of DIC in this context is that it is suitable for hierarchical Bayesian models, where constraint is expressed through prior specification rather than reduction in parameters. DIC differences are scaled similarly to AIC and BIC

differences, and differences in DIC greater than 10 may be considered strong evidence for one model over the other. For the experiment reported in Pratte and Rouder (2011), this new model outperforms our previous, less constrained model by 103 DIC points. For Experiments 1, 2, and 3 reported below, the new model outperforms our previous one by 42, 122, and 27 DIC points, respectively. We therefore adopt this new model for all analyses, although similar results are obtained using the previous model in Pratte and Rouder (2011). We benchmarked this hierarchical dual-process model on simulated data in the manner described in Pratte and Rouder. The model provides accurate parameter estimates and confidence intervals compared to known true values.

Covariation of Recollection and Familiarity Across People and Items

One of the advantages of hierarchical models is accurate recovery of individual participant and item effects on recollection and familiarity. These effects may be used to explore the nature of processing as follows. If there is only one mnemonic process, then it is reasonable to expect that people good at recollection are necessarily good at familiarity, and items that elicit high recollection elicit high familiarity. Conversely, independence of such effects would be evidence for a two-process system.

Figure 4 shows estimated participant and item effects from a large recognition-memory experiment reported originally in Pratte et al. (2010), reanalyzed with the hierarchical dual-process model presented above. Participant-specific recollection is clearly correlated with participant-specific familiarity. Likewise, item-specific recollection is correlated with item-specific familiarity. These results suggest that differences in people's mnemonic abilities and differences in how difficult items are to remember may be explained by a single underlying process.

Although the strong covariation of participant and item effects indicates a single process, participants and items are sampled rather than manipulated. These correlations may therefore reflect common processing that is not indicative of mnemonic processing. For example, if a particular participant is prone to distraction, then

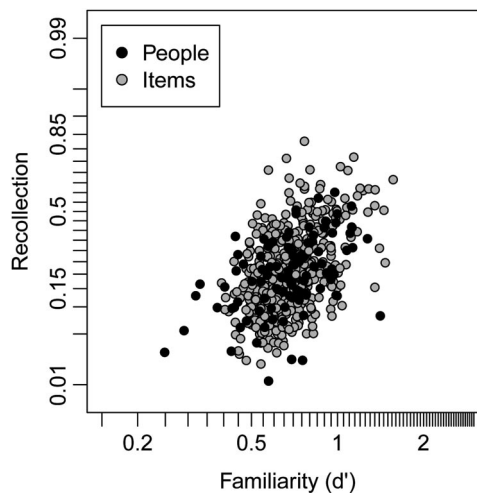


Figure 4. Recollection and familiarity estimates for participants (dark points) and items (light points) from Pratte and Rouder (2011).

this participant's estimate of recollection and familiarity may both be attenuated, resulting in correlations across people. Likewise, it is plausible that items covary in nonmnemonic ways that may affect both recollection and familiarity. Hence, such correlations may not be a definitive marker of mnemonic processing (Jacoby & ShROUT, 1997). Consequently, we focus here on experimental rather than observational effects. Our strategy is to perform a state-trace analysis on recollection and familiarity parameters obtained with the hierarchical model across several manipulations designed to differentially dissociate recollection and familiarity.

Inference About the Influence of Manipulations

At this point, we have developed a hierarchical model to accurately estimate familiarity and recollection in Yonelinas' model, and these estimates are uncontaminated by variation across items and people. We also have adopted a state-trace representation to interpret parameter variation across manipulations. There remains one more task for assessment: constructing contrasts and statistical tests that differentiate state-trace patterns.

The conventional assessment of two processes is based on differential-influence logic: The recollection-enhancement manipulation results in a greater increase in the recollection parameter than does the familiarity-enhancement manipulation, and the familiarity-enhancement manipulation results in a greater increase in the familiarity parameter than does the recollection-enhancement manipulation. Let ΔR_r denote the (main) effect of the recollection-enhancement manipulation on the recollection parameter R :

$$\Delta R_r = [(R_2 - R_1) + (R_4 - R_3)]/2,$$

where the subscripts refer to the conditions denoted in Figure 1. Likewise, let ΔR_f denote the effect of the familiarity-enhancement manipulation on recollection:

$$\Delta R_f = [(R_3 - R_1) + (R_4 - R_2)]/2.$$

A test of whether the recollection enhancement manipulation affects recollection more than the familiarity enhancement manipulation is whether ΔR_r is greater than ΔR_f . Likewise, the following two familiarity-averaged differences may be defined:

$$\Delta F_r = [(F_2 - F_1) + (F_4 - F_3)]/2$$

$$\Delta F_f = [(F_3 - F_1) + (F_4 - F_2)]/2.$$

A test of whether the familiarity enhancement manipulation affects familiarity more than the recollection enhancement manipulation is whether ΔF_f is greater than ΔF_r . The double selective influence and differential dissociation patterns of Figure 1A and 1B, respectively, imply both that $\Delta R_r > \Delta R_f$ and that $\Delta F_f > \Delta F_r$.

Figures 1E through 1H provide a graphical representation of these contrasts. Shown are plots of ΔR as a function of ΔF for recollection-enhancement and familiarity-enhancement conditions. Ellipses are used to indicate sampling noise, which for the dual-process model is negatively correlated.¹ The darker and lighter ellipses represent the recollection-enhancement and familiarity-enhancement manipulations, respectively. The key constraints that (a) the recollection effect must be greater in the recollection-enhancement condition than in the familiarity-

enhancement condition and (b) the familiarity effect must be greater in the familiarity-enhancement condition than in the recollection-enhancement condition may be seen in Figure 1E (for selective influence) and Figure 1F (for differential dissociation). Likewise, this pattern fails for the single-process configurations, as shown in Figures 1G and 1H.

It is more desirable to have a single contrast or test statistic rather than to consider multiple separate tests (see Nieuwenhuis, Forstmann, & Wagenmakers, 2011). A good test statistic for comparing Δ_r and Δ_f across conditions is the slope of the line that connects the centers of the ellipses (shown as a dashed line in Figures 1E–H). This slope is negative when $\Delta R_r > \Delta R_f$ and $\Delta F_r > \Delta F_f$, implying there is evidence of selective influence or a differential dissociation. Otherwise, it is positive. We call the assessment of this slope the slope test. The slope test differentiates the dual-process patterns in Figures 1A and 1B from the single-process patterns in Figures 1C and 1D.

Experiment 1

One of the classic manipulations in memory is levels of processing (LOP; Craik & Lockhart, 1972). In the deep-encoding condition, participants were presented a word at study and reported a semantically related associate. In the shallow-encoding condition, participants reported the number of vowels in studied words. Deep encoding at study not only should improve performance but should primarily affect recollection (e.g., Yonelinas, 2001). A second classic manipulation is perceptual match between study and test. Words are presented in different fonts, color, or modalities (visual vs. auditory). When a word is tested with the same perceptual features as it was studied (called a perceptual match), there is improved performance. According to dual-process theory, perceptual matches should primarily affect familiarity (Gregg & Gardiner, 1994; Yonelinas, 2002). In Experiment 1, study items were presented either in lowercase and colored red or presented in uppercase and colored green. In the perceptual match condition, the case and color were preserved across study and test; in the perceptual mismatch condition, case and color were changed across study and test.

If the dual-process model holds, these manipulations should produce a pattern of estimates much like that in Figures 1A and 1B. Alternatively, if recognition memory is mediated by a single process, estimates of recollection and familiarity should fall along a line, as in Figures 1C and 1D.

Method

Participants. Fifty-two University of Missouri students participated in Experiment 1 in return for credit toward a course requirement.

Stimuli. Stimuli were 480 words drawn from the MRC Psycholinguistic Database (Coltheart, 1981) restricted to contain between four and nine letters and have Kucera–Francis frequency of usage between 1 and 200 occurrences per million words of text

¹ This correlation may be seen by studying the likelihood function of the model. The likelihood contains a ridge where high levels of R are associated with low levels of F and vice versa. The consequence is a trade-off between ΔR and ΔF .

(Kucera & Francis, 1967). Study lists were constructed by randomly sampling 240 of these words, and there was a unique study list for each participant. Test lists were composed of all 480 words. Presentation order was completely randomized at study and at test and across all participants.

Design. The design of the experiment was a two-by-two within-subject factorial design with independent variables of perceptual match (levels of match and mismatch across study and test) and level of processing (levels of deep and shallow encoding). Each of these factors was manipulated in a within-list manner.

Procedure. At study each word was accompanied by one of two sets of instructions. For half of the words, the instructions read "Read the word out loud and enter the number of vowels." For these *shallow* study trials, participants used the numeric keyboard keys to enter the number of vowels in the word on the screen. For the remaining words, the instructions read "Read the word out loud and enter a related word". For these *deep* study instructions, participants were instructed to type the first word that came to mind that was related to the word on the screen. Study instructions varied randomly from trial to trial. Each study word was displayed either in lowercase letters and colored red or in uppercase letters and colored green. Following study, the test phase consisted of the sequential presentation of the 240 studied words and of 240 new words. The color/case of half of the studied words was the same as when the word was studied, and the color/case of the remaining half was different (e.g., study in lowercase red, tested in uppercase green). This perceptual match versus mismatch manipulation was crossed with levels of processing in a balanced latin-squares design. For each item at test, participants made one of the six confidence ratings: *sure new*, *believe new*, *guess new*, *guess studied*, *believe studied*, and *sure studied*. They were asked to use all response options to report their confidence as accurately as possible. Prior to the experiment, participants performed a 5-item practice study session and a 10-item practice test session with a separate list of 10 items. The practice session included all four conditions that were presented in the experiment, though not in equal numbers.

Results

Figure 5A shows ROC curves averaged over participants and items. The obvious characteristics are a sizable sensitivity effect of LOP and a small effect of perceptual match. To assess whether

there is one or more than one process, we fit the hierarchical Bayesian version of the Yonelinas dual-process model and performed the slope test. We also fit the dual-process model more conventionally, which is discussed next. Conventional analysis allows us to compare our results to previous experiments. Moreover, by comparing the hierarchical Bayesian approach to a more conventional approach, we may assess the degree to which the more conventional approach is affected by distortions from aggregation.

Conventional analysis. To compute recollection and familiarity estimates in a conventional fashion, we follow Heathcote et al. (2006). The Yonelinas dual-process model was fit separately to each individual's data by first aggregating data over items to compute hit and false alarm rates and then finding parameter values that maximized the likelihood of these rates. Individuals' recollection estimates served as input to a repeated-measures analysis of variance (ANOVA) to assess the effect of the experimental conditions on recollection. For recollection, there is a main effect of LOP ($M = .16$ vs. $.50$ for the deep and shallow conditions, respectively), $F(1, 51) = 76.4$, $p < .001$; no main effect of perceptual match ($M = .34$ vs. $.32$ for the match and mismatch conditions, respectively), $F(1, 51) = 1.16$, $p \approx .29$; and no interaction, $F(1, 51) = 0.84$, $p \approx .36$. Likewise, to assess the effects on familiarity, individual familiarity estimates were submitted to a separate repeated-measures ANOVA. For familiarity, there is a main effect of LOP ($M = .90$ vs. 1.42 for the deep and shallow conditions, respectively), $F(1, 51) = 50.8$, $p < .001$; a main effect of perceptual match ($M = 1.07$ vs. 1.24 for the mismatch and match conditions, respectively), $F(1, 51) = 10.0$, $p \approx .0026$; and marginal evidence for an interaction, $F(1, 51) = 3.1$, $p \approx .08$. These results lead to two conclusions:

1. Although the effect of perceptual match was small, it still was sufficiently large to lead to a detectable effect in model parameters. In this regard, the manipulation may be viewed as successful.
2. The pattern of effects reveals a selective influence where perceptual match affects familiarity but not recollection. Such selective effects in one of the parameters have conventionally been taken as at least partial evidence for two processes (e.g., Howard et al., 2006).

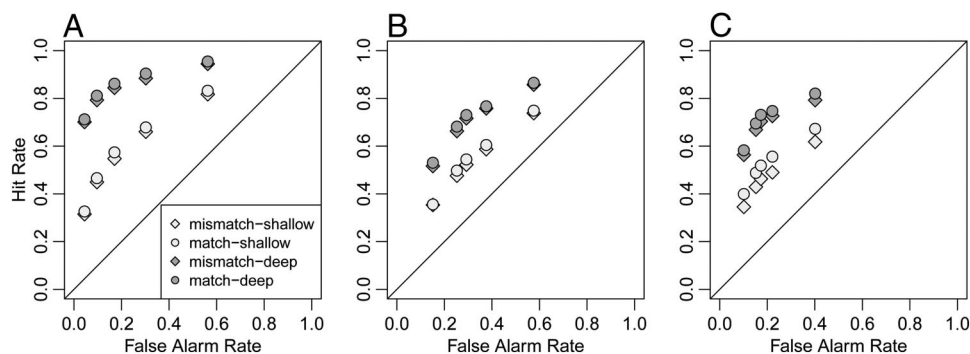


Figure 5. Receiver operating characteristic (ROC) curves across conditions in Experiments 1, 2, and 3. Hit and false alarm rates were constructed by aggregating across both participants and items.

The preceding analysis suffers from five potential flaws: First, parameter estimates are obtained by aggregating across items, resulting in possible underestimation of variability and an associated increase in statistical decision errors above nominal rates. Second, the analysis relies on establishing a null recollection effect of perceptual match, which is outside the scope of conventional significance testing. Third, the result is based on the combination of two tests; there is no control for inflated error rates resulting from multiple comparisons. Fourth, the results do not meet the true differential dissociation or double selective influence criterion—there is a selective effect in only one of the parameters rather than in both. Fifth, the use of two, separate ANOVAs does not account for covariation in estimation between recollection and familiarity. Negative covariation is the norm in these models. All five of these potential flaws are mitigated in the following Bayesian hierarchical model analysis.

Hierarchical model analysis. Figure 6A shows the state-trace plot of overall recollection ($\mu_k^{(r)}$) as a function of overall familiarity ($\mu_k^{(f)}$) for the four conditions.

Figure 6B shows a plot of the main effects, and the main effect on recollection is plotted as a function of the main effect on familiarity. Shown are the joint posterior distributions for the LOP effect (ellipse far from the origin) and for the perceptual match effect (ellipse near the origin). This figure follows the same format as the main-effect plots in Figure 1 (right side). The dashed line shows the mean slope, which is unambiguously positive.

There are several approaches to testing and model selection in Bayesian modeling, and our preferred approach is selection by Bayes factors (Kass & Raftery, 1995; Rouder, Speckman, Sun, Morey, & Iverson, 2009). Unfortunately, algorithms for computing Bayes factors for these hierarchical models are unknown, and development is outside the scope of this paper. Instead, we adopt a straightforward strategy of computing the analog of a p value from posterior distributions for target contrasts. Figure 7 shows an example. Shown is the posterior distribution of the effect of perceptual match on familiarity, which is given by $\Delta F_f = [(\mu'_3 + \mu'_1) - (\mu_4 + \mu_2)]/2$ and may be computed directly from the samples from the posterior distributions provided by the Bayesian analysis (see Gelman, Carlin, Stern, & Rubin, 2004). The value of zero is in the left tail, with 97.7% of the mass above this value and the remaining 2.3% of the mass below it. We construct a two-tail analog of the p value, which we denote p^* , as follows: We consider the probability of observing zero or below (2.3%, in the left shaded tail) plus the mirror image in the right tail (the upper 2.3%); hence, for this posterior, $p^* = .046$. We report these two-tail p^* values throughout. According to the Bayesian hierarchical model, there is a modest effect of perceptual match on familiarity ($p^* = .046$). There are well-known critiques about the use of p values (Edwards, Lindeman, & Savage, 1963; Rouder et al., 2009; Sellke, Bayarri, & Berger, 2001; Wagenmakers, 2007) that apply here as well. We use this approach as a best available option and with full knowledge of the limitations.

For recollection, the effect of LOP was so large that all of the observed samples from the posterior distribution (out of 5,000 total samples after burn-in and thinning) were greater than zero ($p^* \approx 0$). This effect can be confirmed by noting that the LOP ellipse in Figure 6B is far from the line that denotes $\Delta R = 0$. The perceptual-match main-effect contrast for recollection was not substantial, and the zero value was well within the distribution ($p^* = .68$). This

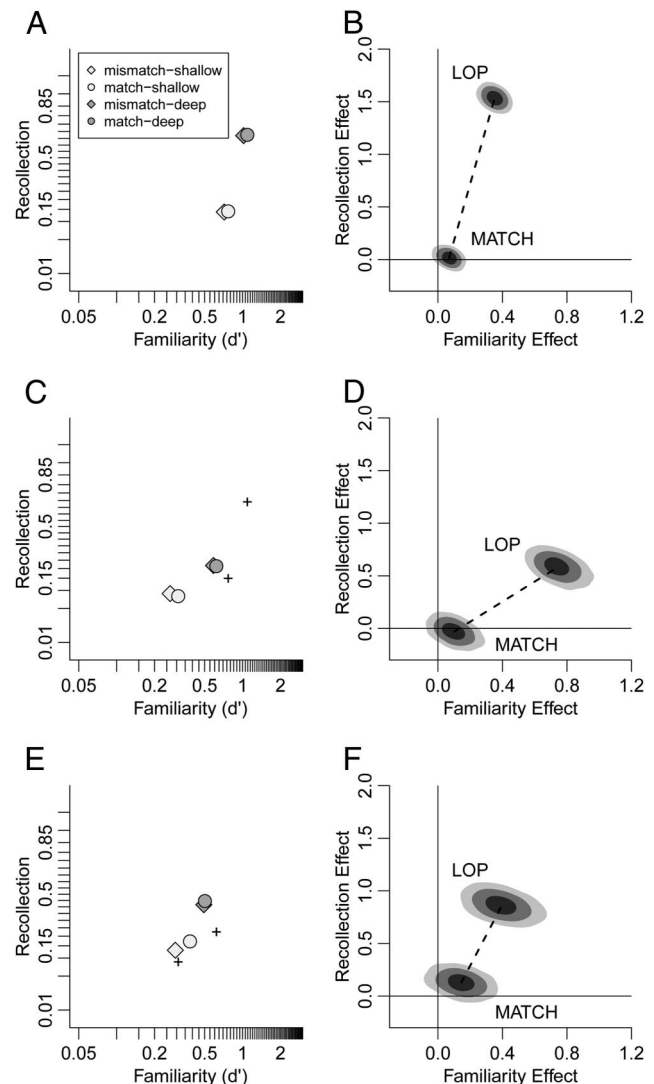


Figure 6. Posterior condition means and main effects. A: Recollection estimates as a function of familiarity estimates of each condition in Experiment 1. B: Posterior distributions of marginal levels-of-processing (LOP) effects (ellipse near origin) and marginal match effects (ellipse far from origin) in recollection and familiarity from Experiment 1. The three shading levels denote regions containing 25, 75, and 95% of the posterior density. Dashed lines connect posterior means of the two effects, and the slopes of these lines are used for the slope test. Note that recollection effects are presented on the probit space and familiarity effects on the log space. C and D: Posterior means and effects from Experiment 2, respectively. Points denoted by crosses in C are estimates from Experiment 1, averaged over match conditions. E and F: Posterior means and effects from Experiment 3, respectively. Points denoted by crosses in E are estimates from Experiment 2, averaged over match conditions.

result can be confirmed by noting that a large minority of the posterior mass of the perceptual match ellipse corresponds to a negative value of the contrast. There was a substantial main effect of LOP on familiarity and a more marginal one of perceptual match ($p^* \approx 0$ and $p^* = .046$, respectively). Notice that the critical p^* value on the familiarity effect, .046, is more

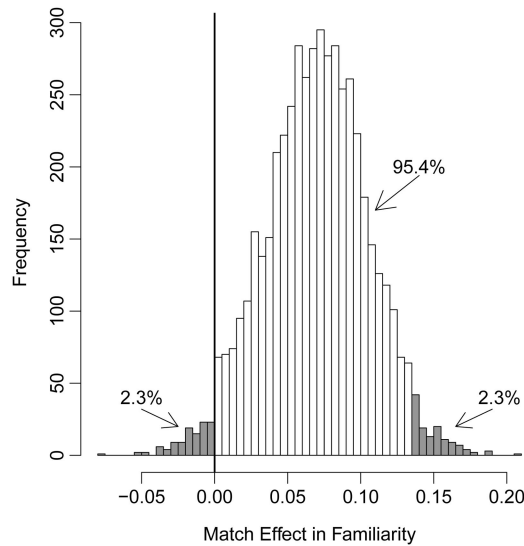


Figure 7. Posterior distribution of the marginal perceptual-match effect in familiarity in Experiment 1. The shaded region on the left denotes mass below zero. The shaded region on the right is the equivalent amount of mass in the right tail of the distribution. The total shaded area defines the Bayesian p value, denoted p^* .

modest in the Bayesian analysis than the $p \approx .0026$ value for the same contrast in the conventional analysis. This difference reflects the fact that the variability of the parameters is understated in the conventional analysis due to aggregation. Overall though, the hierarchical analysis shows that the perceptual match effect was sufficient in size to be reflected modestly in familiarity, a mnemonic parameter.

The slope test is based on samples from posterior distributions in exactly the same fashion. The slope is substantially positive—all of the 5,000 samples from the posterior of slope were greater than zero (i.e., $p^* \approx 0$). Negative slopes are definitely not present; consequently, the data may be accounted for parsimoniously with one process.

Discussion

Conventional analysis of Experiment 1 yielded patterns that would be considered evidence for two processes. In particular, the match effect was manifest in familiarity ($p < .003$) but not in recollection. From a conventional point of view, this result is highly noteworthy. Selective influence of familiarity parameters has been elusive in paradigms that control for decision criteria effects. Finding an instance of this type of selective influence in conjunction with previous recollection effects (e.g., that from aging; Howard et al., 2006) seems to provide a high degree of evidence for two processes. Hence, if we did not have the benefit of the methodologically principled Bayesian state-trace analyses, these results might be interpreted as strong behavioral support for two processes. Yet, the more methodologically principled analysis shows that the data provide a lack of evidence for a two-process model. In particular, the pattern fails the slope test (the slope was positive rather than negative). The analysis as a whole shows that

conventional methodologies may produce artifactual evidence for two processes.

The hierarchical model analysis and associated plots in Figures 6A and 6B reveal that the perceptual match effects are small relative to the LOP effects. This small magnitude makes assessing the patterns in Figure 6B difficult. Our experiment is not underpowered—we tested 52 participants on 480 items each. Indeed, finding mnemonic effects of perceptual match has proven elusive (see Hockley, 2008; Mulligan, Besken, & Peterson, 2010; Murnane & Phelps, 1995). One study that did find large perceptual match effects is that of Boldini, Russo, and Avons (2004), who asked participants to respond before a strict deadline. When participants were allowed ample time to make old-new judgments, Boldini et al. found no perceptual match effects. In contrast, when participants were speeded by the imposition of a response window, large perceptual match effects were observed. The dual-process explanation for this interaction is that recollection takes time to develop. Under conditions of speed stress, participants may not have time to recollect and, consequently, may rely more heavily on familiarity. Accordingly, because participants conceivably rely more heavily on familiarity to perform the task, manipulations of familiarity, such as perceptual match, become more effective.

Boldini et al. (2004) performed an old-new recognition memory experiment without confidence judgments. Because confidence judgments are necessary for fitting the Yonelinas dual-process model, we expanded the paradigm. Experiment 2 is a speeded confidence-ratings recognition memory task. This experiment provides an additional test of the dual-process model. According to the dual-process model, speeding responses should produce an overall decrease in recollection rather than in familiarity. We may check this prediction by comparing recollection and familiarity estimates across Experiment 1, which had no speed stress, and Experiment 2, which had speed stress.

Experiment 2

Experiment 2 was identical to Experiment 1 with the exception that participants were required to make confidence-rating responses within a short time window following presentation of each test word. Differences in method between Experiments 1 and 2 are presented:

Method

Participants. Fifty-three University of Missouri students participated in Experiment 2 in return for credit toward a course requirement.

Procedure. Each test trial began with a fixation cross. Following 500 ms of fixation, a test word appeared. After 500 ms, **** was presented below the test word. Participants were instructed to respond as soon as possible, after the **** appeared. If the response was made before the **** appeared, a low-pitched error tone sounded, followed by the message *Too Fast, Respond after *****, presented for 3 s. If the response was made more than 500 ms after the **** appeared, the same low-pitched error tone sounded, followed by the message *Too Slow, Respond Faster!!!*, also presented for 3 s. Participants thus had high incentive to respond within a window from 500 to 1,000 ms after onset of the test word. For comparison, in Experiment 1 the average response time was 1.97 s and the 25th percentile was 1.05 s.

Prior to the study phase, participants completed two training phases to become acquainted with the response window and test instructions. First, participants were given test trials with *WORD* serving as the stimulus. They were instructed to make any of the six confidence ratings as quickly as possible after the **** appeared. This training phase continued until responses were made within the correct response window 15 times consecutively (most participants had no problem doing so after making a few “too slow” errors). Next, participants completed a practice study-test phase with a 10-item test list, followed by the experiment.

Results

Four participants had overall sensitivity below chance, indicating a failure to follow instructions, and were discarded from further analyses. Furthermore, responses outside the 500 ms to 1,000 ms window, which constituted 13% of the total, were discarded. Figure 5B shows ROC curves for each condition. As can be seen, there is once again a large LOP effect and a negligible perceptual match effect.

Manipulation check. We checked whether the deadline of 1,000 ms was sufficient to change processing by comparing overall performance across Experiment 1 (without the deadline) and Experiment 2. As can be seen in the ROC curves (see Figure 5), there is a large decrement in performance for all four conditions. Speeding lowered the grand average hit rate from .71 to .62, and increased the false alarm rate from .17 to .32. This large effect indicates that a 1,000 ms deadline is sufficient to affect performance. According to the dual-process account, this performance deficit should be differentially located in recollection rather than familiarity, as recollection is the slower process. Additionally, participants used the extreme responses (*sure studied*, *sure new*) more with the deadline than without. In Experiment 1, only 10% of participants used the extreme responses on 90% or more of the trials. In Experiment 2, 29% of the participants used the extreme responses 90% or more of the time. This increase in extreme responses is not due to more accurate responding; in fact, overall accuracy decreases with speeding. Possible interpretations of this increase in extreme responses are provided in the Discussion.

Conventional analysis. As before, conventional analysis was performed by first estimating individuals' parameters by maximum likelihood methods and then submitting these estimates to repeated-measure ANOVAs. There is a main effect of LOP on recollection ($M = .11$ vs. $.19$ for the deep and shallow conditions, respectively), $F(1, 48) = 14.3$, $p < .001$; no main effect of perceptual match on recollection ($M = .16$ vs. $.14$ for the match and mismatch conditions, respectively), $F(1, 48) = 0.52$, $p \approx .47$; and no interaction, $F(1, 48) = 1.3$, $p \approx .26$. There is also a main effect of LOP on familiarity ($M = .43$ vs. $.83$ for the deep and shallow conditions, respectively), $F(1, 48) = 55.3$, $p < .001$; no main effect of perceptual match on familiarity ($M = .61$ vs. $.65$ for the mismatch and match conditions, respectively), $F(1, 48) = 0.65$, $p \approx .42$; and no interaction, $F(1, 48) = 2.6$, $p \approx .12$. Thus, the conventional analysis of these data does not provide evidence of selective influence but rather implies that there was no significant effect of match in either recollection or familiarity.

Hierarchical model analysis. Figure 6C shows the state-trace plot of overall recollection as a function of overall familiarity for the four conditions. Assessing posterior distributions of main

effects reveals that whereas LOP affected both recollection ($p^* \approx 0$) and familiarity ($p^* \approx 0$), perceptual match affected neither recollection ($p^* \approx .62$) nor familiarity ($p^* \approx .12$). Figure 6D shows the main effects in recollection as a function of main effects in familiarity. All of the 5,000 samples from the posterior of slope (dashed line) were greater than zero, indicating that the slopes are definitely not negative.

The dual-process model predicts that the primary effect of speeding responses would be to attenuate recollection rather than familiarity. We checked this prediction by comparing the results of Experiment 2 to those of Experiment 1. Figure 6C shows the results of Experiment 2 (diamonds and circles) and Experiment 1 (pluses show estimates averaged across perceptual mismatch and match conditions). The effect of speeding depends on LOP and for the shallow level is primarily manifest in familiarity rather than recollection. This result is contrary to dual-process predictions and replicates a similar finding by Ratcliff and Starns (2009). The effect of speeding for the deep level is in both recollection and familiarity, and the familiarity effect is not predicted by the dual-process account.

Discussion

On the basis of results of Boldini et al. (2004), we expected the perceptual match effect to be larger under the speeded conditions of Experiment 2 than Experiment 1. However, we found that the match effect was actually smaller in Experiment 2. Thus the Boldini et al. result of increased perceptual match effects under speed stress does not replicate in a confidence-rating paradigm. Given the small size of the perceptual match effects, it is not too surprising that there is no evidence for two processes.

It is reasonable to worry about limitations in speeding confidence recognition-memory responses. This approach is not used often (but see Ratcliff and Starns, 2009); unspeeded confidence ratings and speeded yes-no recognition are more common. One of the results we found was that participants tended to use extreme responses, including inaccurate extreme responses, when confidence ratings are speeded. This finding presents no particular problem for the confidence-as-different-criteria approach used in the signal-detection models: Extreme responses are accounted for by bunching of criteria, so that there is much area under the latent-strength distributions for extreme responses and little area for more intermediary ones. Why this bunching may occur, however, is outside most mnemonic theories. Alternatively, this result may indicate a shifting of strategies or processing. Perhaps participants under time pressure make only coarse discriminations between old and new items and use the extreme responses to indicate these binary choices. Such an alternative is consistent with the increase of extreme responses with an overall decrease in accuracy. This alternative, however, is incompatible with the extant dual-process theory. Accordingly, recollection is associated with slow, accurate, extreme responses and familiarity is associated with faster, more varied, and less accurate responses. Hence, the shift to coarse, inaccurate responses with speeding is not predicted by dual-process theory. Instead, responses should be well distributed, as they are in Experiment 1. The possibility of strategy or processing differences induced by speeding confidence-ratings responses does not threaten the main conclusion of this paper that

there is a lack of evidence for two processes in general and a lack of evidence for the Yonelinas model in particular.

Experiment 3

One of our concerns in Experiment 2 was a fairly low level of overall performance. In Experiment 3, we provide a replication of Experiment 2 with shortened study lists. Decreasing the number of studied items decreases interitem interference, and according to the dual-process model such interference should affect only recollection, as familiarity is automatic and operates independently of other storage processes (Yonelinas & Jacoby, 1994). Experiment 3 therefore not only provides another opportunity to assess the Boldini et al. effect but allows us to explore a portion of the state-trace space with high recollection and low familiarity, should two processes describe the data.

Method

Participants. Seventy-four University of Missouri students participated in Experiment 3 in return for credit toward a course requirement.

Stimuli. Stimuli were 160 words randomly selected from the 480-word pool used in previous experiments.

Procedure. Experiment 3 was identical to Experiment 2 in that LOP and perceptual match-mismatch (case/color) manipulations were used, as was a 500–1,000 ms response window. In Experiment 3, however, participants studied 80 words and were tested on 160 words (compared with 240 and 480 item study and test lists, respectively, in Experiments 1 and 2).

Results

Data from 10 participants were discarded due to an experimenter error. Responses for remaining participants that were made outside of the response window (15%) were also discarded. Figure 5C shows ROC curves for the remaining data for each condition. Once again, there is a large effect of LOP and a small effect of perceptual match. As in Experiment 2, about one third (36%) of the participants used the extreme *sure new* and *sure studied* responses on 90% or more of the trials. The impact of these participants is discussed further in the section on combining results across experiments.

Conventional analysis. There is a main effect of LOP on recollection ($M = .14$ vs. $.26$ for the deep and shallow conditions, respectively), $F(1, 63) = 19.6$, $p < .001$; no main effect of perceptual match on recollection ($M = .21$ vs. $.19$ for the match and mismatch conditions, respectively), $F(1, 63) = 0.52$, $p \approx .24$; and no interaction, $F(1, 63) = 0.56$, $p \approx .45$. There is also a main effect of LOP on familiarity ($M = .68$ vs. 1.09 for the deep and shallow conditions, respectively), $F(1, 63) = 32.0$, $p < .001$; a main effect of perceptual match on familiarity ($M = .81$ vs. $.96$ for the mismatch and match conditions, respectively), $F(1, 63) = 0.65$, $p \approx .003$; and no interaction, $F(1, 63) = 0.33$, $p \approx .57$. These results replicate the dissociation seen in Experiment 1 with the conventional analysis: LOP affected both recollection and familiarity, and perceptual match selectively influences only familiarity.

Hierarchical model analysis. Figure 6E shows the state-trace plots of overall recollection as a function of overall famil-

ilarity for the four conditions. Assessing posterior distributions of main effects reveals that LOP affected both recollection ($p^* \approx 0$) and familiarity ($p^* \approx 0$). There is marginal evidence for main effects of perceptual match in both recollection ($p^* \approx .04$) and familiarity ($p^* \approx .09$). However, both effects are driven entirely by the match effect in the shallow LOP condition. Figure 6F shows main effects in recollection as a function of those in familiarity. All samples from the posterior of slope (dashed line) were greater than zero, implying that these data do not provide evidence for two processes via the slope test.

Discussion

The results of Experiment 3 are similar to those of Experiment 1. The conventional analysis yielded evidence for two processes, providing evidence that perceptual match affected familiarity ($p < .01$) and not recollection ($p > .2$). This analysis, however, is deficient for two reasons. First, the Bayesian analysis, which provides undistorted estimates and accurate confidence intervals on them, indicates small effects of perceptual match in both recollection and familiarity, rather than a selective influence. Second, the slope test, which is developed within the logic of state-trace analysis, is more concordant with a single process than with two processes.

One goal of Experiment 3 was to increase recollection over Experiment 2, in order to observe data that violate the apparent monotonicity of the state-trace plot seen across Experiments 1 and 2. Comparing estimates from Experiment 2 (crosses in Figure 6E) with those in Experiment 3 shows that the list length manipulation did primarily affect recollection as predicted by the dual-processes model. However, as discussed in the next section, we believe that these estimates may be biased due to the fact that a subset of participants used the *sure* responses on the majority of trials.

Combining Results Across Experiments

In Experiments 1 and 3, we found a statistically significant perceptual match effect on familiarity. These experiments are the first we know of that show mnemonic effects of perceptual match that cannot be attributed to criteria shifts. Even though we found these effects primarily in familiarity, state-trace analysis reveals no evidence for two processes. Although the mnemonic effects of perceptual match are statistically significant, they are too small to draw firm conclusions about the nature of processing. We comment on the implications of the small size of this effect in the General Discussion. In this section, we use state-trace analysis to combine the results across all three experiments to provide additional insight about processing.

One of the advantages of state-trace analysis is that if a single-process model holds, the data from disparate experiments should lie on a single curve regardless of the manipulations. All that is needed is some set of manipulations for which recollection and familiarity span a suitable range. When we combine the data across all three experiments, the overall parameter estimates for recollection span a range from $.07$ to $.69$ in probability, and the overall parameter estimates for familiarity span a range from $.2$ to 1.3 in d' units. These spans are sufficiently large to see patterns in the state-trace analysis. Figure 8A shows a combined state-trace plot for Experiments 1 through 3. Recollection is plotted as a function

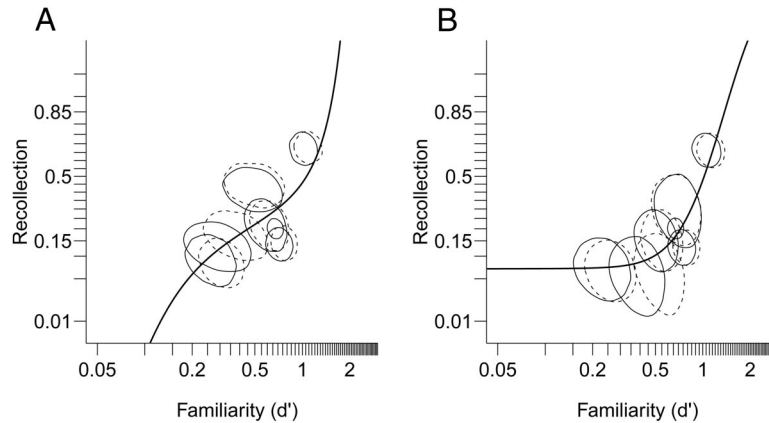


Figure 8. Hierarchical dual-process estimates from Experiments 1, 2, and 3. A: Recollection plotted as a function of familiarity from all conditions in Experiments 1, 2, and 3. Circles denote 95% confidence regions; dashed lines are from perceptual mismatch conditions, and solid lines are from match conditions. The small circle in the middle is from an experiment reported in Pratte et al. (2010) that included 97 participants. B: Estimates after excluding data from participants who used the extreme *sure* responses on 90% or more of trials. Curves are monotonic spline smooths (Ramsay & Silverman, 2005) fit using the **R** package *fda*.

of familiarity for the 12 conditions from the three experiments. The overall trend is a strong positive relationship where recollection and familiarity covary. A hypothetical monotonically increasing curve is plotted that accounts for much of this covariation. Although the shape of this curve will change with transformations of the parameters, for the parameterization used here effects tend to be in familiarity when overall performance is low and in recollection when performance is higher. The critical point is that the differential effects are not characteristic of a differential dissociation pattern and in themselves provide no evidence for two processes.

The positive relationship captured by the line is not perfect, however, and the deviations are worthy of consideration as evidence for two processes. Consider, for example, the two ellipses marked with arrows. Clearly, the relationship between these selected points is not positive: The lower point is marked by lower recollection than the upper point and possibly higher familiarity. We think, however, that there are a few problems with interpreting the negative relationship between the arrowed points as evidence for two processes. First, the points are chosen post hoc, and there needs to be some correction for multiple post hoc comparisons. Second and perhaps most important, the difference may be due to a minority of participants who display an extreme response pattern. The task in Experiments 2 and 3 is speeded confidence ratings. Some participants, perhaps in an attempt to deal with the burden of making a choice among six alternatives in limited time, effectively reduced the task to a two-choice task by using only extreme responses. In fact, 33% of participants used *sure new* and *sure studied* responses on at least 90% of trials across Experiments 2 and 3, whereas only 10% displayed this pattern in Experiment 1.² Unfortunately, when participants limit their responses to the endpoints, the data in each condition may be represented by a single hit and false alarm rate, that is, by a single point in ROC space. Such a point can be accounted for perfectly by recollection alone or familiarity alone, or any combination of each. Hence, when responses are limited to the extremes, the data do not identify recollection or familiarity parameters, and estimates in the Bayes-

ian approach unduly reflect the influences of the prior. We reanalyzed the experiments, discarding participants who produced extreme responses on more than 90% of trials. When those participants who provided limited information on recollection or familiarity are excluded, the negative relationship is largely attenuated (see Figure 8B). Hence, we are hesitant to interpret this negative relationship as a robust characteristic of the data. On balance, the combined state-trace plots provide some evidence that recollection and familiarity tightly covary, as predicted by a single-process model.

The confidence regions in Figure 8B are substantially larger than what are typically reported, and there are two reasons why they provide an accurate characterization of how well recollection and familiarity may be estimated. First, these estimates were obtained by taking both participant and item variability into account. In previous studies data were aggregated over participants, items, or both in order to obtain hit and false alarm rates necessary for model estimation, and as we have shown (Pratte & Rouder, 2011), this practice leads to a systematic underestimation of confidence intervals. Second, the Bayesian analysis allows us to draw confidence *regions* on the joint distribution of recollection and familiarity, rather than marginal confidence intervals on each parameter separately. Recollection and familiarity parameters in the likelihood function of the dual-process model are not independent but have a negative relationship that is more pronounced with smaller sample sizes. Drawing separate confidence intervals on these two parameters fails to account for this nonindependence and thus mischaracterizes how well both recollection and familiarity are known, given the data. It is important to note that the large confidence regions in Figure 8B do not result from small sample sizes. Even after removal of participants who produced extreme

² Ratcliff and Starns (2009) had participants extensively practice the response-deadline task. After practice, their participants were less likely than ours to make excessive extreme responses.

responses, Experiments 1, 2, and 3 included 47, 35, and 41 participants, respectively, and these sample sizes are at least as large as typical recognition memory studies.

General Discussion

In these experiments, we attempted to assess whether there are two mnemonic processes by employing conventional manipulations that should differentially affect processing. Our analysis differs from previous ones in two important ways. First, we avoided the use of aggregated data and instead base our estimates on hierarchical models that simultaneously account for participant and item variability. Second, we avoided using multiple significance tests, affirming the null with significance tests, and considering single factors in isolation. Instead, we used a novel slope test developed within the state-trace framework. We draw three conclusions:

1. The results from Experiments 1, 2, and 3, when viewed separately, provide no evidence for two processes. The state-trace plots may be fit with a single line, and the slope test is never passed.
2. When the results of all three experiments are overlaid on a single state-trace plot (see Figure 8), there is some evidence for a single-process interpretation.
3. When viewed through the lens of conventional analysis rather than the more principled Bayesian analysis, the results of Experiments 1 and 3 show a selective influence of perceptual match on familiarity. This finding has been elusive in paradigms that control for differences in decision criteria, such as confidence ratings, and would be conventionally interpreted as support for two processes.

The contrast between the conventional analysis, in which there is evidence for two processes, and the Bayesian state-trace analysis, in which there is no evidence for two processes, shows that conventional analyses are flawed in substantial and meaningful ways.

The first conclusion, the failure to observe dual-process signatures in the state-trace tests on each experiment separately, undoubtedly reflects the relatively small size of the perceptual match effect. It may be argued that if larger perceptual match effects are observed, the results may pass the slope test. Although sizable match effects have been observed in the remember-know paradigm (e.g., Gregg & Gardiner, 1994), these effects are arguably criterial in nature rather than mnemonic (Dunn, 2004). Our match effects are the largest we know of within the confidence rating paradigm (see Mulligan et al., 2010). Although Boldini et al. (2004) found larger match effects in an old-new task under speeded conditions, we failed to replicate this finding within the confidence rating paradigm in two experiments. Given all the past difficulty in finding mnemonic effects with the confidence-ratings paradigm, it is not likely that our experiments are idiosyncratic failures to replicate. The more likely scenario is that large mnemonic effects of perceptual match, as opposed to criterial effects, simply do not exist.

We would lower our skepticism of the dual-process account if there existed a manipulation that (a) results in large effects in familiarity and not in recollection and (b) does so in a paradigm

that allows for criterion effects such as confidence ratings. We chose to explore the perceptual match manipulation because, based on a review of experiments spanning multiple paradigms and dual-process models, it had the best chance of selectively influencing familiarity (Yonelinas, 2002). Other choices known to affect *know* responses in the remember-know framework, such as changing response criteria from lax to strict (see Yonelinas, 2002), seem even less likely to yield dual-process patterns of mnemonic parameters in a confidence ratings paradigm. In our view, finding this critical familiarity manipulation is a primary challenge to dual-process theory, and we can think of no plausible candidates. The fact that no such manipulation has been identified even after decades of research raises the possibility that it does not exist, perhaps because memory is not mediated by two processes. Stated succinctly, even after much experimentation, familiarity has yet to be brought under sufficient experimental control.

The second conclusion, that the combined state-trace plot across all experimental conditions is well accounted for by a simple curve, is indicative, at least to some degree, of a single mnemonic process. The pattern can be stated succinctly: When performance is low, effects tend to be manifest in the familiarity parameter; when performance is more intermediate, effects tends to be in both parameters. This pattern seemingly holds regardless of the nature of the manipulation. We view this finding as more exploratory than confirmatory. It is provided provisionally, and there are certainly caveats and limitations to the interpretation. First, in addition to LOP and list length, these points were spread out by a response deadline manipulation. We are worried that there may be a strategy shift to using the most extreme responses in confidence ratings under deadline pressure. When participants use only the outside response options, it is impossible to accurately measure recollection and familiarity effects. The resulting estimates in this case are unduly affected by the specification of the prior and the choice of parameterization. Second and perhaps more important, the interpretation of the evidence in Figure 8 is made by inspection without recourse to any test. We have provided confidence regions on points but have yet to develop an omnibus test statistic. Recent work by Prince, Heathcote, and Brown (2012) provides a promising approach for doing so, but further statistical development would be needed to incorporate this approach with the hierarchical model we use here to obtain accurate estimates. We find the single-process interpretation of the combined results in Figure 8 most useful when combined with the slope test failures. Taken together, the agreement among these analytic approaches provides strong evidence against two-process theories.

We have used the Yonelinas dual-process model to derive estimates of recollection and familiarity from confidence ratings. This choice was made because (a) the Yonelinas model is popular and influential in several areas of memory research and (b) we have developed a hierarchical version of this model that accounts for participant and item variability without recourse to averaging. However, Wixted and Mickes (2010) have recently introduced an alternative dual-process model of confidence ratings, in which both familiarity and recollection are continuous variables, as opposed to the discrete recollection process assumed in the Yonelinas model. We strongly suspect that a state-trace analysis using this continuous-recollection model would yield the same results reported here—that recollection and familiarity can be described by a line. However, doing so would require the development of a

continuous-recollection model that accounts for participant and item variability, in order to obtain accurate confidence intervals on parameter estimates that are critical for state-trace analysis. Because this model is based on signal detection theory, the techniques we have used to estimate the hierarchical Yonelinas model may provide a starting point for such development.

The experimental logic we set forth was to assess whether there is any evidence that target manipulations dissociate recollection from familiarity in recognition memory. This logic, the search for factors that affect processes in theoretically clear and transparent ways, is the workhouse and gold standard of experimental science in several fields. Finding such manipulations would provide strong evidence for dual-process theory, and a failure to find them must be viewed as evidence against it. Our evidence is behavioral and experimental, while an increasing amount of evidence for two processes is based on neurophysiological measures and is increasingly observational. Some of this evidence appears to be quite substantial, including Curran's dissociation of different brain waves associated with familiarity and recollection (Curran et al., 2006; Rugg & Curran, 2007) and Yonelinas' demonstration of patients who have deficits in either familiarity or recollection alone (Sauvage, Fortin, Owens, Yonelinas, & Eichenbaum, 2008; Vann et al., 2009; Yonelinas et al., 2002). Our approach here is marked by careful methodological control, so that we make principled inferences without recourse to aggregation and without combining reject and fail-to-reject significance test results. These other approaches seem evidentiary; however, it is not known whether they will stand when placed under the same methodological spotlight. For example, Freeman, Dennis, and Dunn (2010) recently used state-trace plots to assess Curran and colleagues' claim of dissociable EEG signatures of recollection and familiarity. They showed that the variability in these potentials may be reduced to a single dimension. Our results clearly do not address recent neurophysiological findings, and reconciling them with our findings remains for future integration.

In summary, we were unable to find evidence of dissociable influence, as assessed with a state-trace approach and the Yonelinas dual-process model, across reasonable manipulations in a confidence-ratings recognition memory task. This inability stands as evidence against dual-process theories of recognition memory and, perhaps more modestly, as evidence for a single-process interpretation. Although there can be little doubt that there are multiple distinct physiological systems for human learning and memory, in our judgment it remains controversial that more than one of these is involved in explicit, episodic memory.

References

- Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology*, 19, 137–181.
- Bargh, J. (1997). The automaticity of everyday life. In R. S. Wyer (Ed.), *The automaticity of everyday life: Advances in social cognition* (pp. 1–61). Mahwah, NJ: Erlbaum.
- Boldini, A., Russo, R., & Avons, S. E. (2004). One process is not enough! A speed-accuracy tradeoff study of recognition memory. *Psychonomic Bulletin & Review*, 11, 353–361. doi:10.3758/BF03196582
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335–359. doi:10.1016/S0022-5371(73)80014-3
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 33(A), 497–505. doi:10.1080/14640748108400805
- Correll, J., Park, B., Judd, C., & Wittenbrink, C. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83, 1314–1329. doi:10.1037/0022-3514.83.6.1314
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.
- Curran, T., DeBuse, C., Worch, B., & Hirshman, E. (2006). Combined pharmacological and electrophysiological dissociation of familiarity and recollection. *Journal of Neuroscience*, 26, 1979–1985. doi:10.1523/JNEUROSCI.5370-05.2006
- Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, 24, 523–533. doi:10.3758/BF03200940
- 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. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review*, 115, 426–446. doi:10.1037/0033-295X.115.2.426
- Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, 95, 91–101.
- Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193–242. doi:10.1037/h0044139
- Freeman, E., Dennis, S., & Dunn, J. (2010). An examination of the ERP correlates of recognition memory using state-trace analysis. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the Cognitive Science Society* (pp. 97–102). Austin, TX: Cognitive Science Society.
- 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. Craik (Eds.), *The Oxford handbook of memory* (pp. 229–244). New York, NY: Oxford University Press.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London, England: Chapman & Hall.
- Gregg, V. H., & Gardiner, J. M. (1994). Recognition memory and awareness: A large effect of study-test modalities on “know” responses following a highly perceptual orienting task. *European Journal of Cognitive Psychology*, 6, 131–147. doi:10.1080/09541449408520140
- Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1210–1230. doi:10.1037/0278-7393.29.6.1210
- Heathcote, A., Raymond, F., & Dunn, J. (2006). Recognition and familiarity in recognition memory: Evidence from ROC curves. *Journal of Memory and Language*, 55, 495–514. doi:10.1016/j.jml.2006.07.001
- Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, 9, 61–65. doi:10.1111/1467-9280.00011
- Hockley, W. E. (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1412–1429. doi:10.1037/a0013016
- Howard, M. W., Bessette-Symons, B., Zhang, Y., & Hoyer, W. J. (2006). Aging selectively impairs recollection in recognition memory for pictures: Evidence from modeling and receiver operating characteristic curves. *Psychology and Aging*, 21, 96–106. doi:10.1037/0882-7974.21.1.96
- Jacoby, L. L. (1991). A process dissociation framework: Separating auto-

- matic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541. doi:10.1016/0749-596X(91)90025-F
- Jacoby, L. L., & ShROUT, P. E. (1997). Toward a psychometric analysis of violations of independence assumptions in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 505–510. doi:10.1037/0278-7393.23.2.505
- Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795. doi:10.2307/2291091
- Kintsch, W. (1967). Memory and decision aspects of recognition learning. *Psychological Review*, 74, 496–504. doi:10.1037/h0025127
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Light, L. L., Prull, M. W., La Voie, D. J., & Healy, M. R. (2000). Dual-process theories of memory in old age. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 238–300). New York, NY: Oxford.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87, 252–271. doi:10.1037/0033-295X.87.3.252
- Morey, R. D., Pratte, M. S., & Rouder, J. N. (2008). Problematic effects of aggregation in zROC analysis and a hierarchical modeling solution. *Journal of Mathematical Psychology*, 52, 376–388. doi:10.1016/j.jmp.2008.02.001
- Morey, R. D., Rouder, J. N., & Speckman, P. L. (2009). A truncated-probit item response model for estimating psychophysical thresholds. *Psychometrika*, 74, 603–618. doi:10.1007/s11336-009-9122-3
- Mulligan, N. W., Besken, M., & Peterson, D. (2010). Remember-know and source memory instructions can qualitatively change old-new recognition accuracy: The modality-match effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 558–566. doi:10.1037/a0018408
- Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 158–172. doi:10.1037/0278-7393.21.1.158
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, 14, 1105–1107. doi:10.1038/nn.2886
- Parkin, A. J., & Leng, R. C. (1993). *Neurophysiology of amnesic syndrome*. Hillsdale, NJ: Erlbaum.
- Pratte, M. S., & Rouder, J. N. (2011). Hierarchical single- and dual-process models of recognition memory. *Journal of Mathematical Psychology*, 55, 36–46. doi:10.1016/j.jmp.2010.08.007
- 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
- Prince, M., Heathcote, A., & Brown, S. (2012). The design and analysis of state-trace experiments. *Psychological Methods*, 17, 78–99. doi:10.1037/a0025809
- Ramsay, J., & Silverman, B. (2005). *Functional data analysis* (2nd ed.). New York, NY: Springer.
- Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, 116, 59–83. doi:10.1037/a0014086
- Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12, 573–604. doi:10.3758/BF03196750
- 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
- Rouder, J. N., Lu, J., Sun, D., Speckman, P. L., Morey, R. D., & Naveh-Benjamin, M. (2007). Signal detection models with random participant and item effects. *Psychometrika*, 72, 621–642. doi:10.1007/s11336-005-1350-6
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Pratte, M. S. (2007). Detecting chance: A solution to the null sensitivity problem in subliminal priming. *Psychonomic Bulletin & Review*, 14, 597–605. doi:10.3758/BF03196808
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237. doi:10.3758/PBR.16.2.225
- Rouder, J. N., Tuerlinckx, F., Speckman, P. L., Lu, J., & Gomez, P. (2008). A hierarchical approach for fitting curves to response time measurements. *Psychonomic Bulletin & Review*, 15, 1201–1208. doi:10.3758/PBR.15.6.1201
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, 11, 251–257. doi:10.1016/j.tics.2007.04.004
- Sauvage, M. M., Fortin, N. J., Owens, C. B., Yonelinas, A. P., & Eichenbaum, H. (2008). Recognition memory: Opposite effects of hippocampal damage on recollection and familiarity. *Nature Neuroscience*, 11, 16–18. doi:10.1038/nn2016
- Schacter, D., & Tulving, E. (1994). What are the memory systems of 1994? In D. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 1–38). Cambridge, MA: MIT Press.
- Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of *p* values for testing precise null hypotheses. *American Statistician*, 55, 62–71. doi:10.1198/000313001300339950
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 64, 583–639. doi:10.1111/1467-9868.00353
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1–12. doi:10.1037/h0080017
- Vann, S. D., Tsivilis, D., Denby, C. E., Quamme, J. R., Yonelinas, A. P., Aggleton, J. P., . . . Mayes, A. R. (2009). Impaired recollection but spared familiarity in patients with extended hippocampal system damage revealed by three convergent methods. *Proceedings of the National Academy of Sciences, USA*, 106, 5442–5447. doi:10.1073/pnas.0812097106
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problem of *p* values. *Psychonomic Bulletin & Review*, 14, 779–804. doi:10.3758/BF03194105
- 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
- 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. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, 130, 361–379. doi:10.1037/0096-3445.130.3.361
- 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., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology*, 48, 516–535. doi:10.1037/1196-1961.48.4.516
- Yonelinas, A. P., Kroll, N. E. A., Dobbins, I., Lazzara, M., & Knight, R. T.

- (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, 12, 323–339. doi:10.1037/0894-4105.12.3.323
- Yonelinas, A. P., Kroll, N., Quamme, J. R., Lazzar, M. M., Sauvé, M. J., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience*, 5, 1236–1241. doi:10.1038/nn961
- Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, 25, 3002–3008. doi:10.1523/JNEUROSCI.5295-04.2005
- Received November 25, 2011
Revision received February 28, 2012
Accepted March 5, 2012 ■