Cognitive Psychology 79 (2015) 40–67

Contents lists available at ScienceDirect

Cognitive Psychology

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

Old processes, new perspectives: Familiarity is correlated with (not independent of) recollection and is more (not equally) variable for targets than for lures

Rani Moran *, Yonatan Goshen-Gottstein

School of Psychological Sciences, Tel Aviv University, Ramat Aviv, 69978, Israel

ARTICLE INFO

Article history:
Available online 17 April 2015

Keywords:
Recollection
Familiarity
Continuous dual processes model
Stochastic independence
Functional independence
Aggregation bias

ABSTRACT

According to dual-process models of memory, recognition is subserved by two processes: recollection and familiarity. Many variants of these models assume that recollection and familiarity make stochastically independent contributions to performance in recognition tasks and that the variance of the familiarity signal is equal for targets and for lures. Here, we challenge these 'common-currency' assumptions. Using a model-comparison approach, featuring the Continuous Dual Process (CDP; Wixted & Mickes, 2010) model as the protagonist, we show that when these assumptions are relaxed, the model's fits to individual participants' data improve. Furthermore, our analyses reveal that across items, recollection and familiarity show a positive correlation. Interestingly, this across-items correlation was dissociated from an across-participants correlation between the sensitivities of these processes. We also find that the familiarity signal is significantly more variable for targets than for lures. One striking theoretical implication of these findings is that familiarity—rather than recollection, as most models assume—may be the main contributor responsible for one of the most influential findings of recognition memory, that of subunit zROC slopes. Additionally, we show that erroneously adopting the common-currency assumptions, introduces severe biases to estimates of recollection and familiarity.

* Corresponding author at: Department of Psychology, Tel Aviv University, Ramat Aviv, POB 39040, Tel Aviv 69978, Israel.
E-mail addresses: rani.moran@gmail.com (R. Moran), goshengott@gmail.com (Y. Goshen-Gottstein).

http://dx.doi.org/10.1016/j.cogpsych.2015.01.005
0010-0285/

1. Introduction

A family of dual-processes models of recognition memory assumes that recognition is based on two distinct processes, recollection and familiarity. *Recollection* involves the retrieval of contextual details that were associated with target presentation, whereas *familiarity* reflects a memory signal that is not accompanied by a vivid conscious experience of the studied event. Rather, the familiarity signal is based on a subjective feeling of acquaintance that is experienced upon an item's presentation, thus enabling one to distinguish previously experienced probe items from novel ones. The distinction between recollection and familiarity is supported by a large body of evidence (Diana, Reder, Arndt, & Park, 2006; See Brainerd, Gomes, & Moran, 2014, for a bivariate-recollection theory that suggests that in addition to familiarity there are two distinct forms of recollection: target recollection and context recollection). Accordingly, a set of theory-driven techniques for disentangling the mutual contributions of both processes to recognition has been developed, the most popular of which is the remember–know paradigm (Tulving, 1985b; other techniques include the dual-process ROC curves and the process-dissociation paradigm).

In the remember–know paradigm, participants report their subjective state of awareness, associated with each 'old' recognition response. When participants can recall specific details from the item-study episode (e.g., the item-color, the feeling it evoked), they are instructed to respond 'remember'. When such recall is consciously absent, yet the item nevertheless seems familiar, they are instructed to respond 'know'. In some cases, a third eligible 'guess' response is to be chosen, when the judgment of the item as 'old' could be attributed to neither recollection nor familiarity (Gardiner, Ramponi, & Richardson-Klavehn, 1998). The common interpretation of R–K judgments is that remember responses reflect the recollection process, whereas know responses correspond to familiarity in absence of recollection.

The validity of this interpretation was challenged by single-process theories (Donaldson, 1996; Dunn, 2004) according to which, the crucial difference between remember and know judgments pertains to memory strength, with the former reflecting 'strong' memories and the latter reflecting 'weak' memories. Recently however, compelling support has been given to the dual-process interpretation with the demonstration that recollection can be weak and familiarity can be strong. Specifically, Ingram, Mickes, and Wixted (2012) found that for both know and remember responses, old–new confidence was positively correlated with item accuracy, yet for the remember responses alone, confidence was positively correlated with source accuracy. Moreover, Ingram et al. found that state-trace analyses suggested a dual process interpretation of the data (for similar results, see Wixted & Mickes, 2010; See Newell & Dunn, 2008, for a review of the state-trace methodology). The authors concluded that stronger know responses are supported by higher levels of familiarity, whereas stronger remember responses are supported by higher levels of recollection.

In the current paper, we examined two common theoretical assumptions underlying many of the dual-process recognition models, which we label the "common-currency assumptions". First, that recollection and familiarity are independent processes. Second, that familiarity for targets and lures can be adequately represented by an equal-variance signal-detection model. Using a model-comparison approach, we tested these assumptions and demonstrated the hazardous implications, which erroneous reliance on these assumptions may yield.

1.1. Are recollection and familiarity independent?

1.1.1. Stochastic independence of processes

If recognition is indeed driven by two distinct processes, then studying the relationship between these processes can deepen our understanding of human memory. In particular, a pivotal question one should ask is to what extent are their separate contributions independent? To this end, it is important to distinguish between two different concepts of process independence: functional independence and stochastic independence (c.f., Mandler, 1959; Tulving, 1985a). *Functional independence* pertains to the ability of each process to operate separately even if the other is prevented from operating. This form of independence is demonstrated when one process varies as a function of an independent variable and the other does not. *Stochastic independence* of processes, on the other hand, occurs when the probability distribution over outputs of each process is invariant with respect to the output of the other process (or equivalently, when the *conditional* probability distribution of the output of one process, is invariant with respect to the 'conditioning values' of the other process). The occurrence of stochastic independence can be measured within experimental conditions, thus not requiring systematically manipulating an independent variable, as dictated by the functional-independence methodology.

In the current paper, we focused on the question of stochastic independence. Interestingly, most if not all extant dual-process theories of recognition, assume that the experiences of recollection and familiarity are stochastically independent (e.g., Continuous Dual Process, CDP, Wixted & Mickes, 2010; Sum-Difference Theory of Remembering and Knowing, STREAK, Rotello, Macmillan, & Reeder, 2004; Dual-Process-Signal-Detection, DPSD, Yonelinas, 1994; Variable-recollection dual-process, VRDP, Onyper, Zhang, & Howard, 2010). Henceforth, we label this assumption the Stochastic Independence Assumption (SIA).

In the following sections, we describe a hierarchy of the different levels at which the SIA may apply. The main thrust of the current paper lies in presenting a modeling approach, which reveals that this assumption is violated at the level, which is commonly assumed by dual process models. Finally, we demonstrate that this violation yields strong biases in the estimation of the processes underlying recognition and that the biases may be eliminated via our proposed modeling approach.

1.1.2. Possible levels of stochastic independence

The SIA may have different interpretations, as a function of the level at which it is assumed. Because these interpretations have critical implications with regard to data analysis and data modeling, we now explicate the different assumptive levels.

1.1.2.1. The participant ∗ item level. One interpretation of the SIA is that familiarity and recollection are independent at the *participant ∗ item* level. This means that when a specific participant is tested on a specific word, the distribution over the possible outcomes of the recollection process is invariant with respect to the outcome of the familiarity processes and vice versa. At this level, the participant and the item are 'fixed' so the process (recollection or familiarity) outcome distributions are insensitive to sources of variability across either the items or the participants that are sampled for testing. Rather, the distributions correspond only to sources of variability that may vary across *trials* (i.e., *trial effects*) for the same participant and item, such as the level of the participant's alertness. The SIA at the participant ∗ item level implies that trial effects, which introduce (across-trial) variability in recollection and familiarity, are independent across these two processes.

A hypothetical way to examine the validity of this assumption would be by conducting repeated, independent measurements of recollection and familiarity for the same person and the same item and probing whether the outcomes of the two processes are stochastically independent (see Fig. 1, Panel 1). However, such independent measurements cannot in practice be executed, because each measurement would likely affect subsequent measurements, for example, by altering the memory engram. Although it is impossible to empirically test the SIA at the participant ∗ item level,[1] some dual-process theorists have claimed that the SIA in their models, applies to this very level (Jacoby & Shrout, 1997; Rouder, Lu, Morey, Sun, & Speckman, 2008).

1.1.2.2. The participant across items level. The '*participant across items*' (henceforth 'across items') constitutes a different level of the SIA. Here it is assumed that for a given participant, the conditional distribution of one process is invariant with respect to the outcome of the other process when process outcomes are measured across *different items* (rather than for any fixed item). In other words, recollection and familiarity are assumed to be stochastically independent even when the outcome

[1] The repeated measurement approach is successfully used in different domains of cognitive psychology. For example, in psychophysics, repeated measurements underlie the method of constant stimuli for (absolute of difference) threshold estimation. In such applications, it is assumed that such measurement are independent at the participant ∗ item level.

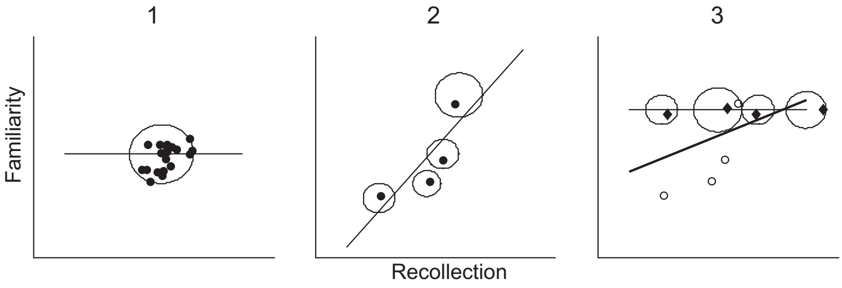

Fig. 1. An illustration of the different assumptive levels of the SIA. Panel 1 depicts 20 hypothetical independent repeated paired measurements of recollection (abscissa) and familiarity (ordinate) for one participant and one item. The ellipse represents a contour of the bivariate Gaussian distribution of recollection and familiarity across different trials. For this particular participant and item (i.e., at the participant ∗ item level), the SIA is not violated as evident from the contour (A contour of a distribution depicts points of equal probability density. For bivariate Gaussian distributions, the contours are elliptical. The components are stochastically independent if and only if the ellipse axes are parallel to the component axes. Note also that non-correlation is a general mathematical consequence of stochastic independence. For bivariate Gaussian distributions, stochastic independence between the components is equivalent to non-correlation.) and from the horizontal regression (of familiarity on recollection) line. (In practice, the regression slope will not be 0, due to sampling error. For sake of simplicity, we plot the average slope that is expected under the hypothetical distribution depicted by the contours.) Panel 2 depicts four distributions, corresponding to four different items (each satisfying the SIA at the participant ∗ item level), for a single participant. Each item is tested only once (filled circles) and these measurements are affected by both trial and item effects. Items that tend to evoke higher level of recollection also tend to generate higher levels of familiarity. Thus, a positive correlation between recollection and familiarity ensues across items and is evident in the positively sloped regression line. Panel 3 reproduces the four measurements from the middle panel (here denoted with empty circles) but includes additional measurements from a second participant, tested on the same four items (filled diamonds), with the distributions depicted by the ellipses. For the second participant, the SIA is not violated across items, as evident in the thin horizontal regression line, despite item-heterogeneity. The thick line in Panel 3 depicts the regression across both participants and items, that is, across all eight measurements. The SIA is violated at the across-items-and-participants level, as evident in the positive regression slope.

distributions (of these processes) are sensitive not only to trial effects, but also to *item effects*, i.e., to variability (in recollection and familiarity) across the population of items, which are sampled for testing. Notably, item effects on recollection and familiarity can either be correlated or non-correlated across items. The difference between the item ∗ participant and the across-items assumptive level is illustrated in Panels 2 and 3 of Fig. 1. Notably, when items are heterogeneous and when the SIA maintains at the participant ∗ item level, it may (Panel 2) or may not (Panel 3) be violated at the across items level.

Although non-repeated testing of participants with multiple items is typical of recognition tests, recollection and familiarity cannot be directly observed for individual items. Importantly, current *indirect* measures of familiarity are based on the SIA and hence as Norman & O'Reilly (2013, p. 612) state: 'there is no way to test this assumption using behavioral data alone because of chicken-and-egg problems (i.e., one needs to measure familiarity to assess its independence from recall, but one needs to assume independence to measure familiarity).' A primary goal of this article was to develop an indirect measure to meet the challenge of gauging the stochastic-dependence (or lack thereof) between recollection and familiarity across items.

1.1.2.3. The item across participants level. The '*item across participants*' (henceforth 'across participants') level of the SIA is relevant whenever different participants are tested on a same item (e.g., Curran & Hintzman, 1997). Here, it is assumed that for a given item, the conditional outcome distribution of one process is invariant with respect to the outcome of the other process when process outcomes are measured across *different participants*. That is, recollection and familiarity are stochastically independent even when (process) outcome distributions are sensitive to '*participant effects*' i.e., variability between participants in recollection and familiarity that a given item elicits.

1.1.2.4. The across both participants and items level. The SIA holds at the 'across both participants and items' level, if the conditional distribution of one process is invariant with respect to the outcome of the other process when process outcomes are measured across *both different participants and different items*. Hence, recollection and familiarity are stochastically independent even when (process) outcome distributions are sensitive not only to variability inherent in trial-effects, but also to participant effects, item effects and participant-item interaction effects (Fig. 1, Panel 3).

1.1.2.5. A hierarchy of assumptive levels. The various levels of the SIA form a hierarchy with respect to the sources of variability that affect the process-outcome distributions at each level. The participant $*$ item level occupies the lowest rank in the hierarchy as it focuses on process-outcome distributions that are sensitive only to variability in trial effects, for a fixed participant and item. At the higher, 'across participants' or 'across items', levels, process outcome distributions are sensitive not only to variability in trial effects, but also to variability in participant and item effects, respectively. Finally, the highest 'across both participants and items level' accommodates all sources of variability that may affect process outcomes, including trial, participant and item effects (as well as interactions between these effects). Since at each level of the hierarchy process distributions are sensitive to all the sources that affect lower-levels, violations of the SIA tend to 'climb up' the hierarchy. In other words, low-level violations contribute to higher-level SIA violations. Thus, for example, if the outputs of the recollection and familiarity processes are already correlated for some participant $*$ item combinations, then this correlation will contribute to, say, a correlation between recollection and familiarity when different items are pooled for a given participant.[2] In this sense, the higher the level of the SIA, the 'stronger' is the assumption.

1.1.3. The assumptive SIA level of dual-process models: Theory and practice

What are the assumptive levels for dual-process recognition models? Notably, the SIA assumptive level is typically not specified in descriptions of theoretical models or of mnemonic measurement-procedures. Here, we will infer from model-based applications to empirical data, that the various 'across' levels assumptions are prevalent in the dual-processes literature (but see Jacoby & Shrout, 1997; Pratte & Rouder, 2011; Rouder et al., 2008, who explicitly endorse the participant $*$ item SIA level).

To illustrate, consider the remember–know task. Yonelinas and Jacoby (1995) assumed that a target receives a 'Remember' judgment whenever it is recollected and a 'Know' judgment in the *conjunctive* event that the target is familiarized and it is not recollected. Moreover, due to the SIA, this conjunctive probability is equal to the multiplication of the probabilities that an item evokes familiarity and that an item does not evoke recollection:

$$\begin{aligned} P(\text{'Know'}) &= P(\text{item evokes familiarity AND item does not evoke recollection}) \\ &= P(\text{item evokes familiarity}) * (1 - P(\text{item evokes recollection})), \end{aligned} \tag{1}$$

In short, $K = F(1 - R)$ or $F = \frac{K}{1-R}$.

In practice, P('Know') and P(item evokes recollection) are estimated based on the *proportions of items* that receive know (K) and remember (R) judgments, respectively. Critically, such proportions are affected not only by sources of variability inherent to different trials (i.e., trial effects) but also by sources of variability inherent to different items (i.e., item effects; By extension, when the proportions are calculated from data that is pooled across both items and participants, the proportions are also affected by participant and by participant-item interaction effects). Therefore, for Eq. (1) to be mathematically valid, it is not sufficient to assume an absence of a correlation between recollection and familiarity due to trial effects (i.e., an SIA at the participant $*$ item level). One must also assume the absence of a correlation between these processes due to item effects (i.e., an SIA at the across-items level).

[2] Note however that, different sources of variability may exert opposite influences on the correlation between recollection and familiarity. Thus, 'lower level' contributions to the correlation can be offset by other 'higher level' variables that affect the correlation. Still, because higher-level correlations reflect the aggregated influence of all sources of variability that affect recollection and familiarity at that level, as a rule, they will be sensitive to any violation of independence at lower levels.

Importantly, Eq. (1) is invalid for estimations using proportions computed across items, if the SIA is restricted to the participant $*$ item level. To illustrate, assume that a participant is tested for recognition, with the item-population comprising two different types of items, equally frequent in the item-population, each (separately) satisfying the SIA at the participant $*$ item level. The first type comprises 'easy' items, which independently evoke recollection and familiarity, with probabilities $R_{easy} = .9$ and $F_{easy} = .8$. The second type comprises 'difficult' items, which independently evoke recollection and familiarity with probability $R_{hard} = .3$ and $F_{hard} = .1$. What are the expected values of R, F and K when pooling across items? These values are obtained by averaging the corresponding values across the two item types. Specifically, $R = \frac{R_{easy}+R_{hard}}{2} = \frac{.9+.3}{2} = .6$ and $F = \frac{F_{easy}+F_{hard}}{2} = \frac{.8+.1}{2} = .45$. To obtain the mean value of K we first calculate K_{easy} and K_{hard} separately. Notably, Eq. (1) holds for each item type separately (because although the assumption of the SIA was restricted to the participant $*$ item level, it carries over to 'across *homogeneous* items', for which item effects are non-variable). Thus, the expected proportions of K responses for each of the sub-populations are: $K_{easy} = F_{easy}(1 - R_{easy}) = .8 * (1 - .9) = .08$ and $K_{hard} = F_{hard}(1 - R_{hard}) = .1 * (1 - .3) = .07$. Averaging across both item types, we obtain $K = \frac{K_{easy}+K_{hard}}{2} = \frac{.08+.07}{2} = .075$. Strikingly, Eq. (1) is violated at the across items level as $F(1 - R) = .45 * (1 - .6) = .18 \neq K = .075$. The reason for this violation is that when 'easy' and 'difficult' items are mixed, easy targets are more likely than difficult ones to elicit both recollection and familiarity. Thus, item effects are positively correlated for recollection and familiarity, rather than stochastically independent, even if the SIA holds at the participant $*$ item level.

When the SIA is violated across items, using the mathematically invalid Eq. (1) to analyze data that is pooled across items may yield severe estimation biases. In real-life situations, R and F are unknown and are estimated based on the proportions of remember and know responses. Continuing our example, suppose that one conducted an experiment that yielded, ignoring sampling biases, the expected proportions of remember and know responses: $R = .6$ and $K = .075$. Plugging these values into Eq. (1) to estimate F yields, $\widehat{F} = \frac{K}{1-R} = \frac{075}{.6} = .125$ a gross underestimation of the 'true' value $F = .45$. The reason for this bias is that due to the positive recollection-familiarity correlation (henceforth, RFC) across items, the conditional probability that an item evokes familiarity is larger when the item is recollected than when it is not recollected. Thus, the proportion of familiar out of recollected items would be *larger* than the proportion of familiar out of non-recollected items. Because the $\frac{K}{1-R}$ estimate of F is based on the proportion of familiar out of non-recollected items, it yields an underestimation.

Furthermore, such downward biases tend to increase as recollection becomes more prominent. One potentially harmful consequence of such estimation biases is that they may lead to unwarranted theoretical conclusions. Consider for example a manipulation that functionally dissociates R from F to the effect of increasing the former but not the latter. If F is underestimated, then this dissociation may be artifactual, in that F too may have increased by the manipulation, but its increase may have been masked by the larger F-underestimation, when R increased.

The above issues pertain not only to the remember–know paradigm but more generally, to the prevalent practice of fitting an SIA-endorsing model to an individual participant's empirical data or to 'group data' pooled across both participants and items. To the best of our knowledge, two models constitute the only exceptions to this rule: the hierarchical process-dissociation model (Rouder et al., 2008) and the hierarchical DPSD model (Pratte & Rouder, 2011), both of which dispense with the necessity to pool data altogether and restrict the SIA to the participant $*$ item level (we discuss the hierarchical approach in Section 4.1.5.).

1.1.3.1. Aggregation biases. The dual-process literature has not always been sensitive to the idea that de facto estimation and modeling practices imply a SIA interpretation at the across-items level. For example, Jacoby and Shrout (1997), in their process-dissociation procedure (PDP; Jacoby, Toth, & Yonelinas, 1993), assumed stochastic independence between mnemonic processes, without specifying the assumptive level. In response, Curran and Hintzman (1995) criticized this assumption by presenting evidence for violations of the SIA across items (see also Curran & Hintzman, 1997; Hintzman & Curran, 1997; Jacoby, Begg, & Toth, 1997). In reply to the criticism, Jacoby and Shrout endorsed the SIA only at the participant $*$ item level but not at the across-items level, while acknowledging that in principle, a violation of the SIA across items may bias parameter-estimators. Notably, Jacoby and

Shrout denied the idea that such biases originate from violations of the implicit assumption of SIA across-item. Instead, they attributed these biases to the use of participant $*$ item level equations to analyze data that are pooled across items, hence—'aggregation biases'. In other words, aggregated biases are attributed to the practice of pooling data per se. We prefer the across-item SIA-violation interpretation, because if the SIA is restricted to the participant $*$ item level, then there is a theoretical gap between the level in which a model's mathematical equations are valid (participant $*$ item), and the level in which, in practice, the data are analyzed (pooling across items).

Two facts remain undisputed, however, regardless of the use of terminology. First, a violation of the SIA at the across items level, does not imply violations at the weaker participant $*$ item level. Second, if a model espouses SIA merely at the participant $*$ item level, then whenever pooled data are analyzed, a discrepancy emerges between the level at which the models equation are mathematically valid (participant $*$ item) and the model-application level (across items). This may result in systematic biases in estimated parameters. Without corrective measures, such biases can play havoc, distorting parameter estimation and the ensuing theoretical conclusions.

In the current study, we challenged the SIA across items. Proponents of the aggregation bias interpretation should reinterpret our goal as an investigation of whether models can benefit, in terms of mitigating aggregation biases, by incorporating a putative across-item stochastic dependence. The current work presents a corrective procedure, which is instrumental for measuring mnemonic processes void of aggregation biases.

1.1.4. Reasons to doubt the SIA

So far, we have argued that the SIA at the across items level is implicitly assumed by de facto measurement and model fitting practices. Are there reasons to believe that recollection and familiarity are in fact stochastically dependent, rather than independent, across items? We believe that the answer is affirmative, for both theoretical and empirical considerations. At a theoretical level, suppose that items vary systematically in their ability to evoke recollection and familiarity (as the 'easy' and 'difficult' items in the example above). For example, a particular participant may find it easier to encode concrete relative to abstract words or short vs. long words. During test, these better-encoded items would evoke higher levels of both recollection and familiarity compared with poorly-encoded items, yielding a positive RFC across items.

In addition, there is relevant empirical data. Both Curran and Hintzman (1995) and Rouder et al. (2008) presented evidence in support of stochastic dependence between recollection and familiarity across items in word-stem completion experiments that were based on the PDP. Note, however that this paradigm is different from standard (old/new) recognition paradigms. In particular, we do not know to what extent the 'consciously controlled recollection' and the 'automatic' processes, which are presumed to be operative in the word-stem completion paradigm, correspond to the recollection and familiarity protagonists of standard recognition paradigms.

More recently, the SIA was challenged by Bodner and colleagues (Brown & Bodner, 2011; Tousignant & Bodner, 2012), who found a positive RFC when participants rated their experiences of both recollection and familiarity on different scales, using an independent rating task (Higham & Vokey, 2004). Additionally, these authors found that some of the recollection-familiarity dissociations, which were revealed when binary judgments were used, disappeared when an independent rating task was used instead. Thus, previous 'binary judgment-based' dissociations may have been spurious, an outcome of a questionable SIA. Qua evidence for stochastic process-dependence, however, both Brown and Bodner's (2011) and Tousignant and Bodner's (2012) findings of a positive RFC are subject to the concern of rating cross-contamination. Specifically, participants may find it odd to rate an item as high on the recollection scale and simultaneously as low on the familiarity scale.

In the current paper, we studied the nature of recollection-familiarity dependence in a paradigm wherein remember–know judgments are mutually exclusive, in particular, free from the cross-contamination concern. That is, we asked whether independent converging evidence would be found for a positive RFC across items or whether the evidence would instead support the default assumption of dual-process models that recollection and familiarity are stochastically independent. Throughout the paper, we were interested in evaluating the RFC separately for targets and lures. Notably, when lures and targets are considered together, a positive RFC may reflect nothing more than a probe-type mediation effect if, for example, targets evoke both higher recollection and higher familiarity than lures.

With the aid of computational models, the correlation between recollection and familiarity may be estimated even in a task that requires categorically exclusive metacognitive judgments (Hirshman, 1998) by comparing a 'correlation model' that allows for the existence of an RFC, with a model that assumes process independence. A recent model, the continuous dual process model (CDP; Wixted & Mickes, 2010), is precisely tailored for this purpose. As we describe below (Section 1.3), the fundamental structural property of CDP, which permits examination of the question of an RFC, is that in each trial, recognition judgments are based on *both* familiarity and recollection (indeed, on their sum) rather than on one *or* the other.

This property is not shared by all dual-process models. For example, in DPSD and VRDP, lures lack any recollective trace and therefore an RFC for lures is undefined. Moreover, when recollection occurs for targets, the familiarity signal is not consulted and therefore exerts no influence on performance. For example, according to DPSD, when a target is recollected, the participant issues an 'old' response with the highest confidence and a remember judgment, irrespective of the magnitude of the familiarity signal. Thus, these models can only identify the distribution of familiarity for non-recollected items. Consequently, one cannot probe whether the distribution of familiarity differs for recollected vs. non-recollected targets—the focal question of an investigation of an RFC. Thus, through the empirical prism of the remember–know recognition paradigm, stochastic independence in the DPSD and the VRDP[3] is a foundational assumption, not an assumption that can be directly tested by fitting these models to empirical data. In conclusion, the CDP model offers a unique model-based way to test for converging evidence for the idea that recollection and familiarity are correlated.

1.2. Is Familiarity for targets and lures equally variable?

Another common assumption of extant dual-process models of recognition (e.g., CDP, DPSD and VRDP) is that familiarity is equally variable for targets and lures.[4] Examination of this assumption was the second focus of our model-comparison approach.

Unlike familiarity, which is commonly assumed to comprise an equal-variance distribution for targets and lures, dual-process models, posit that recollection is either non-existent for lures (DPSD, VRDP) or otherwise, is existent but is less variable than recollection for targets (CDP). Importantly, these assumptions are intimately related to one of the most robust and influential findings of recognition memory: the subunit (typically ~0.80) slopes of the old–new z-ROC curve (e.g., Ratcliff, Sheu, & Gronlund, 1992). Given the familiarity–variability egalitarianism these models postulate, it is only by virtue of higher target recollection variability that they are able to account for this finding. Importantly, a putative equality of target-lure familiarity variability would be violated if the study episode contributed variables gains to item-familiarity. According to the encoding variability hypothesis (Hintzman, 1986; Ratcliff, McKoon, & Tindall, 1994; Ratcliff et al., 1992; Wixted, 2007; but see a recent debate in which it has been argued that currently, there is no empirical evidence for encoding variability, Jang, Mickes, & Wixted, 2012; Koen & Yonelinas, 2010, 2013; Starns, Rotello, & Ratcliff, 2012), different items are encoded with different levels of efficiency. If study-gains in familiarity are sensitive to encoding variability then the target familiarity distribution would be higher in both mean and variance compared with the lure distribution. In this case, the familiarity equal-variability assumption would turn out to be erroneous. Falsely assuming equal variance, therefore, would detract from systematic investigations as to possible sources that mediate higher target variability (such as encoding variability). Furthermore, like the stochastic independence assumption described above, we will show below (Section 3.4) that this false assumption too induces measurement biases. Fortunately, in addition to the issue of an RFC, a model comparisons procedure, rooted in the CDP model, provides a suitable framework for testing this assumption.

[3] We note that VRDP was not designed as a model of remember–know judgments. Here, we mean that a putative recollection-familiarity correlation will not influence the old–new and confidence judgments, according to this model.

[4] While this assumption is critical in the DPSD (for reasons that will be discussed in Section 4.2.), it is not critical in CDP. Yet, prior applications of CDP adopted this assumption as a matter of convenience.

1.3. A brief description of CDP

CDP is a purely dual signal-detection based model (see Fig. 2), which assumes that both recollection and familiarity are continuous processes (Wixted & Mickes, 2010). Each process is associated with a 'noise' and a 'signal' distribution corresponding to lures and target respectively. More formally, it is assumed that the pair of recollection and familiarity signals (R_p, F_p) is distributed according to a bivariate normal distribution $(R_p, F_p) \sim N\left(\left(\mu_{R,p}, \mu_{F,p}\right), \begin{pmatrix} \sigma^2_{R,p} & \rho_p\sigma_{R,p}\sigma_{F,p} \\ \rho_p\sigma_{R,p}\sigma_{F,p} & \sigma^2_{F,p} \end{pmatrix}\right)$, where the subscript p corresponds to the probe type (T-targets, L-lures), $\mu_{R,p}$ and, $\mu_{F,p}$ are the means of the recollection and familiarity signals respectively, $\sigma_{R,p}$ and $\sigma_{F,p}$ are the standard deviations of these signals and ρ_p is the correlation between these signals. For lures it is assumed that $\mu_{F,L} = \mu_{R,L} = 0$ and that $\sigma_{R,L} = \sigma_{F,L} = 1$.

According to CDP, old/new recognition decisions are based on the sum of the recollection and familiarity signals for a particular test item $M_p = R_p + F_p \sim N(\mu_{F,p} + \mu_{R,p}, \sigma^2_{R,p} + \sigma^2_{F,p} + 2\rho_p\sigma_{R,p}\sigma_{F,p})$, where M denotes the aggregate mnemonic evidence (*memory strength*). The confidence criteria are placed along the combined mnemonic evidence axis. Additionally, the model assumes that when asked to make remember/know/guess judgments, participants can interrogate separately the constituent R, F signals. Accordingly, metacognitive remember and know criteria, denoted c_R and c_K, respectively, are placed on the recollection and familiarity axes, respectively (see Fig. 2).

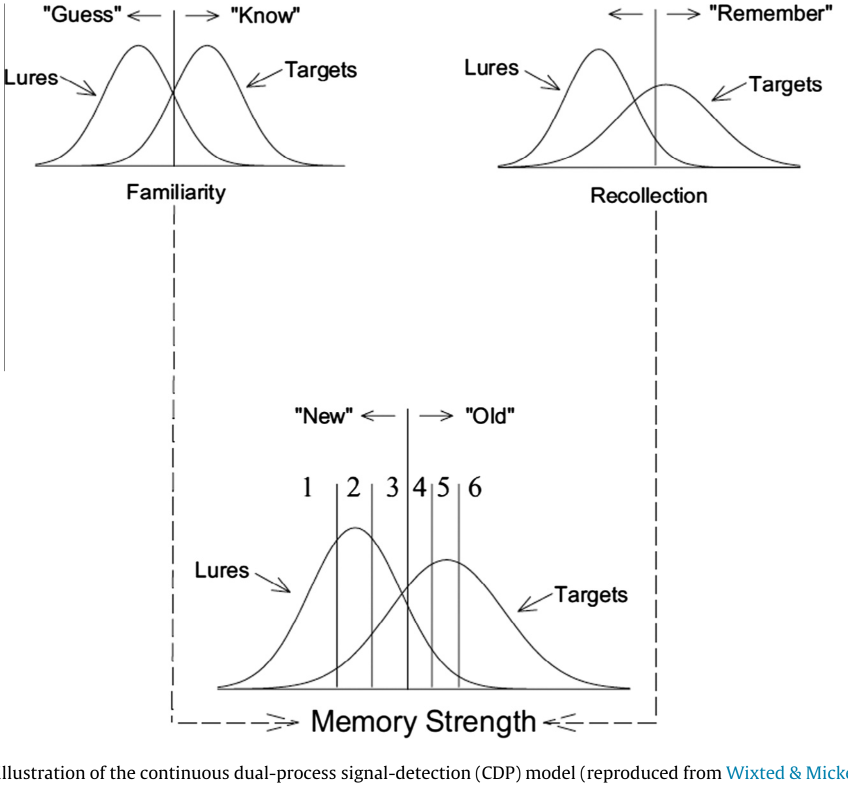

Fig. 2. An illustration of the continuous dual-process signal-detection (CDP) model (reproduced from Wixted & Mickes, 2010, p. 1033). For old/new decisions, the recollection and familiarity signals are assumed to be summed, so the model reduces to the standard unequal-variance signal-detection account. However, when the participant is asked to make a remember/know/guess judgment, memory is queried for recollection, and the participant makes a remember judgment if the recollection signal exceeds a decision criterion. If recollection fails to exceed that criterion, memory is queried for familiarity, and the participant makes a know judgment if familiarity exceeds a criterion (otherwise, a guess judgment is made).

In support of CDP, Wixted and Mickes (2010) demonstrated that it provides adequate fits for sample recognition data. Moreover, CDP can account for the striking finding (described above; Ingram et al., 2012) that low-confidence remember judgments are associated with lower old–new accuracy but with higher source accuracy than high-confidence know judgments, a finding which strongly challenges strength based single-process models.

Prior applications of CDP (Wixted & Mickes, 2010) made the two 'common-currency' assumptions, that familiarity is equal in variance for targets and for lures, $\sigma_{F,T} = \sigma_{F,L} = 1$, and that familiarity and recollection are stochastically independent across items (in particular, these processes were assumed to be non-correlated: $\rho_L = \rho_T = 0$). We now turn to our model-comparison study.

2. Model comparison study

2.1. Empirical data

We fitted several CDP model variants (described below) to the data of Exp. 1 and 2 of Wixted and Mickes (2010). Both experiments consisted of a recognition task with 20 confidence levels (1–10 corresponding to 'new' decisions, 11–20- to 'old' choices). Old judgments were followed with a 'meta-cognitive' remember/know/guess judgment. Additionally, memory for source details was tested. Each participant was tested on 128 targets and 128 lures. The difference between the experiments is that whereas in Exp. 1, source memory was tested for each item following the metacognitive judgment, in Exp. 2, the source-memory test was issued only after metacognitive judgments were given for all 'old' probes.

2.2. Model fitting method

2.2.1. Data preparation

A full CDP model for the above data sets consists of a large parameter set, with a massive 19 parameters dedicated to the confidence criteria alone (in addition to the other parameters). Unfortunately, a large number of model parameters, limits the reliability with which they can be estimated.[5] Furthermore, we were interested in fitting the model to individual participants, for many of whom some of the confidence categories were empty. Given that the confidence criterion parameters are of no focal interest to us, we pooled the responses of confidence levels 1–5, 6–10, 11–15, and 16–20 into four composite confidence categories '1', '2', '3', '4', respectively. The models we used thus required only three confidence criteria parameters. After pooling, the data consisted of eight response categories ('1', '2', '3-R', '3-K', '3-G', '4-R', '4-K', '4-G'). Thus, the data provided 2 (probe type) $*$ (8–1) = 14 free empirical observations (per participant) to test model fits.

2.2.2. Model inventory

We fitted the data of each participant with several CDP model variants, forming a hierarchy of submodels (see Table 1 for a summary of the model parameters). In all models, it was assumed that $\mu_{F,L} = \mu_{R,L} = 0$ and that $\sigma_{R,L} = \sigma_{F,L} = 1$. The two extreme members in this family are the simplest 'common-currency', 8 free-parameters model, in which $\sigma_{F,T} = 1$, $\rho_L = \rho_T = 0$ (henceforth denoted also MODEL8), and the most complex model, which includes 11 free parameters, in which $\sigma_{F,T}$, ρ_L and ρ_T are non-constrained free parameters—the 'non-constrained' model (henceforth also denoted by MODEL11).

Next, we designed another model variant, the 'single-correlation and unconstrained familiarity variance' model (henceforth also denoted by MODEL10). This model was motivated by Tousignant and Bodner's (2012) finding that the correlations between recollection and familiarity ratings across items were similar for both targets and lures ($\sim$.60). Whereas the non-constrained MODEL11 allows

[5] Having a large number of parameters poses several restrictions. When the number of free parameters increases, so does the risk of settling on a local rather than on a global minimum. In this event, the best fitting parameters are not recovered. But even if the globally best parameters are identified, one is still taxed by the general truism of parameter estimation procedures: The higher the number of parameters estimated from the data, the lower the reliability of each parameter.

Table 1
Parameters of the CDP model variants.

Parameter	Meaning
c_2, c_3, c_4	Confidence criteria
c_R, c_K	Remember–know–guess criteria
$\mu_{R,T}$, $\mu_{F,T}$	Means of the recollection and familiarity target distributions
$\sigma_{R,T}$	Standard deviation of the recollection target distribution
$\sigma_{F,T}$	Standard deviation of the familiarity target distribution
ρ_T, ρ_L	Recollection-familiarity Correlation for targets and for lures

Note. The free parameters of the CDP-model variants. The top part of the table presents the eight free parameters, of the common-currency (MODEL8) model (comparable to the variant used in Wixted & Mickes, 2010). The bottom part lists the three additional free parameters of the 'non-constrained' (MODEL11) model.

for different RFCs for targets and for lures, MODEL10 is obtained from the non-constrained model by assuming that $\rho_T = \rho_L$, hence consisting of ten free parameters. To anticipate our results, we found consistent evidence for the superiority of MODEL10 over MODEL11. Therefore, we assumed in subsequent analyses that the correlation between familiarity and recollection is best described by a single common parameter for targets and lures.

The sub-model hierarchy contains two intermediate 9-free parameter models between the common-currency model and the 'single-correlation and unconstrained familiarity variance' model. The first, the 'single-correlation' model, in which $\rho = \rho_T = \rho_L$ is a free parameter but target-familiarity is constrained to a variance of 1, $\sigma_{F,T} = 1$ (henceforth also denoted by MODEL9Corr). The second, the unconstrained familiarity variance model, in which the target familiarity standard deviation $\sigma_{F,T}$ is a free parameter but the RFC is constrained to 0, $\rho_T = \rho_L = 0$ (henceforth also denoted by MODEL9Fam). A comparison across this family of models, summarized in Table 2, should reveal which (if any) of the 'common-currency' assumptions should be modified.

Each of the models was fitted to the data of each participant in Exp. 1 and 2 of Wixted and Mickes (2010). When individual differences among participants are substantial, fitting the model to the entire group-average performance may yield distorted parametric estimates and conclusions (c.f. Estes & Maddox, 2005; But see Cohen, Sanborn, & Shiffrin, 2008). The identifiability of each of the models was confirmed by repeating the fits many times with different starting points (for further details, see Appendix A.3).

2.2.3. Objective function

Model fitting was based on a maximal-likelihood (ML) estimation procedure, as follows. Given a profile of free model parameters, one can compute the model-predicted probabilities of responses in each of the eight response categories for each probe type $p_{i,T}, p_{i,L}, 1 \leqslant i \leqslant 8$ (See Appendix A.1). Given these predictions the likelihood of the data is:

$$L = \prod_{i=1}^{8} \prod_{j=T,L} p_{i,j}^{N_{i,j}}, \tag{2}$$

Table 2
The model inventory.

Model	Constraints
Non-constrained (MODEL11)	–
Single-correlation and unconstrained familiarity variance (MODEL10)	$\rho_T = \rho_L$
Single-correlation (MODEL9Corr)	$\rho_T = \rho_L$, $\sigma_{F,T} = 1$
Unconstrained familiarity variance (MODEL9Fam)	$\rho_T = \rho_L = 0$
Common currency (MODEL8)	$\rho_T = \rho_L = 0$, $\sigma_{F,T} = 1$

Note. The 'Constraints' column indicates which constraints were imposed on the non-constrained model (MODEL11) to obtain the given sub-models. The number of degrees of freedom for each model is obtained by subtracting the number of free model parameters from fourteen, the number of free data points.

where $N_{i,T}, N_{i,L}, 1 \leqslant i \leqslant 8$, are the empirical frequencies of responses in each category According to ML estimation, the 'best fitting' parameters were defined as the combination of free model parameters that maximized the likelihood of the data.

2.3. Model comparison method

2.3.1. Information criteria

Model comparisons were performed with the Akaike information criterion (AIC; Akaike, 1973) and the Bayesian information criterion (BIC; Schwarz, 1978). Both criteria implement a tradeoff between 'goodness of fit', gauged by $-2\ln(\mathrm{L})$ for the best fitting parameters, and model parsimony, as measured by the number of free parameters. The model with the minimal criterion value is preferred. Both criteria implement different penalties on the number of free parameters k:

$$\mathrm{BIC} = -2\ln(\mathrm{L}) + \log(N) * k, \tag{3}$$

$$\mathrm{AIC} = -2\ln(L) + 2 * k, \tag{4}$$

where N is the total number of observations per participant ($N \sim 256$ in the current data). Notably, AIC is more liberal than BIC with respect to incorporating additional model complexity (as long as $\log(N) > 2$).

We also calculated the AIC, BIC values for the entire group. In this approach, the entire set of individual fits is perceived as a single fit for the entire dataset, when each participant has his or her own free parameters and different participants are independent of each other. Thus, Eqs. (3) and (4) were used but with group values for $\ln(L)$, k and N, which were obtained by summing the corresponding values across individuals.

2.3.2. Bootstrap generalized likelihood ratio tests

To anticipate our results, we found that AIC and BIC were in conflict. Whereas according to AIC, MODEL10 was the best single model, BIC indicated that MODEL8 had the upper hand. We thus focused on these two models alone and conducted a follow up, Bootstrap-Generalized-Likelihood-Ratio-Test (BGLRT) analysis. This method relies on powerful statistical machinery for comparing a model with its sub-model namely, Generalized Likelihood Ratio Tests (GLRT; c.f. Riefer & Batchelder, 1988), which we now describe.

Suppose that L_a and L_b are maximal likelihood values for 'model 1' with a free parameters, and for its sub-model, 'model 0', with only b free parameters ($a > b$). GLRT assumes as H_0 that the more parsimonious 'model 0' is the constituent (i.e., data-generating) model vs. the alternative H_1, that the more complex 'model 1' is the constituent model. This hypothesis is tested based on the statistic $\widetilde{L} = 2\ln(L_a/L_b)$ which is asymptotically distributed $\chi^2(a-b)$ (c.f. Riefer & Batchelder, 1988). Extremely high values of this statistic indicate that the improvement of the goodness of fit, which is provided by the complex model justifies rejecting model 0.

Here, we tested MODEL8 vs. MODEL10. However, one major obstacle had to be overcome. Since individual data were fit, which are not 'asymptotic' in size, we could not safely rely on the asymptotical distribution assumption that $\widetilde{L} = 2\ln(L_{10}/L_8) \sim \chi^2(10-8=2)$. Instead, in BGLRT, we used bootstrap methods to calculate directly the H_0 distribution of $\widetilde{L}$ (for each participant). The BGLRT is similar in spirit to the parametric bootstrap cross-fitting method (see Wagenmakers, Ratcliff, Gomez, & Iverson, 2004, for a definition of this method and Jang, Wixted, & Huber, 2011, for a productive application in the context of recognition memory) but is tailored for the purpose of comparing a model with a sub-model, rather than any two models. The BGLRT proceeds according to the following four-step sequence for each participant (further technical details are provided in Appendix A.2).

First, we took a non-parametric bootstrap sample, equal in size to the original empirical sample, by sampling with replacement trials from the empirical data. This resampling procedure is conducted separately for each probe type (targets and lures). Second, we fit MODEL8—the more parsimonious model—to this bootstrap sample to obtain the ML parameters (henceforth 'generative parameters'). Third, based on these generative parameters, we generated a parametric bootstrap sample. To this end, we calculated MODEL8's predicted probability distribution across the eight response categories (for each probe type; see Appendix A.1). Each parametric bootstrap sample was subsequently created by generating independent observations according to these predicted distributions. Fourth, we fitted both MODEL8 and MODEL10 to the parametric bootstrap sample, and calculated $\widetilde{L}$ based on the ML values of these fits. Importantly, now the statistic $\widetilde{L}$ corresponds to 'synthetic data' (more accurately, to a parametric bootstrap sample) that were generated by MODEL8, according to H_0.

We repeated this four-step sequence a large number of ($M = 1000$) times, thus generating distribution for the test statistic under the H_0 assumption that MODEL8 is the constituent model. Finally, we calculated $\widetilde{L}_{empiric}$, for the empirical data. The p-value for $\widetilde{L}_{empiric}$ is the proportion of bootstrap $\widetilde{L}$ values that are larger or equal to $\widetilde{L}_{empiric}$. Similarly, we conducted a BGLRT for the group data by relying on the statistic $\widetilde{L}_{group} = \sum_i \widetilde{L}_i$ where i sums over the individual participants.

The BGLRT has several appealing properties. First, BGLRT integrates the model selection problem with the statistical framework of hypothesis testing. Second, unlike AIC and BIC, which constitute general comparison methods among *any* two models for the same data, generalized likelihood ratio tests capitalize on the additional hierarchical structure when specifically comparing a model with its submodel. Third, in AIC and BIC, model–complexity penalties are based only on the number of free model-parameters. However, when comparing models, there are additional important model-properties to consider such as flexibility, which depends on a model's functional form (c.f. Myung, 2000; Pitt & Myung, 2002; Pitt, Myung, & Zhang, 2002). The BGLRT method, accommodates for such factors because, rather than applying a general 'one size fit all' penalty term, it tailors the $\widetilde{L}$ distribution (under H_0) to the specific models at question. Furthermore, by computing a separate bootstrap distribution for each participant, the method tailors the $\widetilde{L}$ distribution to the specific parametric range that is relevant to the data in question. In conclusion, the BGLRT has several advantages over the use of information criteria. We thus relied on this method as an arbitrator when AIC and BIC were in contention.

3. Results

3.1. Model comparison

We first compared the non-constrained (MODEL11) and the single-correlation and unconstrained familiarity variance model (MODEL10). This comparison was motivated by Tousignant and Bodner's (2012) finding of nearly identical across-item correlations between rating of recollection and familiarity, for targets and lures. We found that even according to AIC, the more liberal of the two information criteria, MODEL11 was superior for only two participants of Exp. 1 ($n = 30$) and for only three participants of Exp. 2 ($n = 32$). The group level AIC supported the same conclusions (AIC = 20240 vs. 20197 for MODEL11 vs. MODEL10 in Exp. 1 and AIC = 23508 vs. 23465 in Exp. 2). Therefore, we did not consider further the non-constrained model and focused only on the four model variants with a single correlation parameter $\rho = \rho_T = \rho_L$.

Next, we compared the remaining four models. The top and bottom parts of Table 3 display the results of this comparison, for Exp. 1 and 2, respectively. Remarkably, the intermediate models MODEL9Corr and MODEL9Fam were not favored by any of the criteria. More importantly, the results are striking in that they feature a contention between the two information criteria. According to AIC, and for both experiments, MODEL10 yielded the best fits according to a majority of participants as well as for the entire group (evident in the group level information criterion). In contrast, the more conservative BIC conveyed a different story according to which for both experiments, MODEL8 was favored. Unfortunately, therefore, conclusions pertaining to the best overall model are ambiguous, in that they dependent on the information criteria. The above notwithstanding, even according to the conservative BIC, the common-currency model is the best model for fewer than half of the participants.[6]

[6] Comparing only MODEL8 and MODEL10, AIC preferred MODEL10 for significant majorities of 21 of the 30 (Exp.1) and 27 of the 32 (Exp. 2) participants. BIC, preferred MODEL 8 for a significant majority of 20 (Exp. 1) or 18 (Exp. 2, the majority is non-significant) participants.

Table 3
Model comparison results for the four models with a single correlation parameter for Exp. 1 and 2 (Wixted & Mickes, 2010).

	MODEL10	MODEL9Corr	MODEL9Fam	MODEL8
AIC Count	**12**	7	4	7
AIC Group	**20197**	20249	20292	20345
BIC Count	4	6	6	**14**
BIC Group	22281	22119	22156	**22012**
AIC Count	**19**	10	1	2
AIC Group	**23465**	23528	23643	23737
BIC Count	5	11	2	**14**
BIC Group	25710	25542	25651	**25532**

Note. AIC and BIC values are rounded to the nearest integers. The 'Count' rows list the number of participants for which each model variant had the best (AIC or BIC) value amongst the four models. 'Group' rows list the group-level information criterion value. Bold entries indicate the model that provided the best fit for the maximal number of participants or that yielded the minimal information criterion. The top and bottom parts correspond to Exp. 1 and 2, respectively.

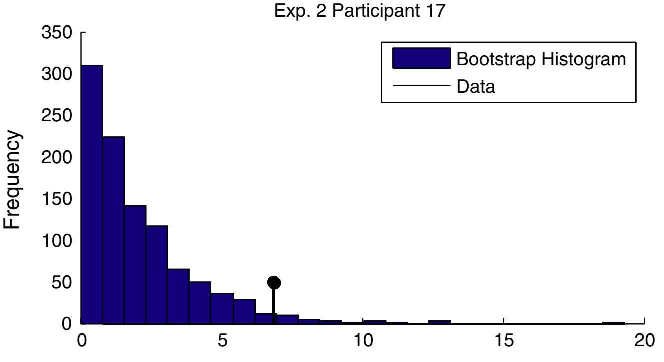

Fig. 3. The results of the BGLRT analysis for Participant 17 in Exp. 2. The histogram depicts the $\widetilde{L} = 2\ln(L_{10}/L_8)$ statistic across 1000 bootstrap samples, and the black circle corresponds to the empirical $\widehat{L}$. Only 26 of the 1000 histogram points were larger than $\widehat{L}_{empiric}$. Hence the *p*-value for rejecting MODEL8 for this participant is $p = .026$.

Since the information criteria failed to provide a unanimously consistent answer with respect to the best model, we conducted an arbitrating BGLRT analysis (see Section 2.3.2). The BGLRT analyses rejected MODEL8, in favor of MODEL10 for significant majorities of 20 of the 30 (Exp.1) and 25 of the 32 (Exp. 2) participants, as well as for the entire group ($p < .01$ for both experiments). Fig. 3 illustrates the BGLRT analysis for a single participant.

Interestingly, for 29 of the 30 participants in Exp. 1, as well as for 30 of the 32 participants in Exp. 2, the results of the BGLRT lined up consistently with the AIC results to the effect that BGLRT rejected MODEL8 if and only if AIC favored MODEL10. Furthermore, across both experiments, there were 24 participants for which AIC and BIC disagreed that is, AIC preferred MODEL10 but BIC favored MODEL8. The BGLRT results were consistent with the AIC results for 21 out of these 24 participants. The agreement between BGLRT and AIC and their contention with BIC is an informative finding in its own right. It is, in fact, consistent with a recent demonstration that BIC may be a too strict comparison criterion for recognition models and AIC may be more adequate (Jang, Wixted, & Huber, 2009). Nonetheless, we urge for a cautious interpretation of these 'AIC superiority' findings, as they may be application-dependent rather than general.

3.2. Model fitting results

In the current section, we present the model fitting results for the 'best' single correlation and target familiarity model (MODEL10) and for the common-currency model (MODEL8). Comparing the estimated parameters for both models reveals the difference in the ways that both models account for the same empirical data. Consequentially, we will demonstrate which of the conclusions that are based on the common-currency model, should be revised when MODEL10 is considered.

Fig. 4 displays the empirical and the model predicted response-proportions for the two different probe types and the four confidence categories (1–4). Overall, the visual impression is that the fits of both MODEL8 and MODEL 10 are remarkably good. Both models capture the confidence distributions for both probe types quite nicely, without any discernible advantage for MODEL10. Importantly, however, predicted confidence proportions in both models are based solely on the composite 'recollection + familiarity' signal, postulated by CDP to mediate confidence judgments, and not on the separate constituent recollection and familiarity components. Qua models of recognition confidence, both MODEL8 and MODEL10 reduce to the standard univariate unequal variance signal detection model (see Fig. 2). Because (up to a scaling factor) MODEL8 has sufficient flexibility to generate any pair of target-lure strength distributions (and confidence criteria), by adjusting the target recollection parameters $\sigma_{R,T}$ and $\mu_{R,T}$, it is not surprising that MODEL8 was able to yield predicted confidence proportions that were similar to those produced by MODEL10. The slight differences between the models, which are still observed in the figure, are attributed to the fact that the models are not fitted solely to confidence responses but to remember/know/guess judgments as well.

Thus, we expected that the advantage of MODEL10 over MODEL8 would better reveal itself when the metacognitive R/K/G judgments are inspected. These judgments are generated by interrogating the separate recollection and familiarity signals and their ensuing distributions, conditional on a particular value (from a range of possible values) of the composite recollection + familiarity signal (which corresponds to the confidence judgment that had already been given). MODEL10, with its additional flexibility in controlling the target-familiarity variance and the correlation between the recollection and familiarity components, should do a better job in capturing the shapes of these conditional distributions, yielding improved model fits. In confirmation, Fig. 5 reveals a clear advantage for MODEL10 over MODEL8. Thus, the additional free parameters $\sigma_{F,T}$ and ρ enhance MODEL10's ability to account for the distributions of metacognitive judgments within confidence levels.

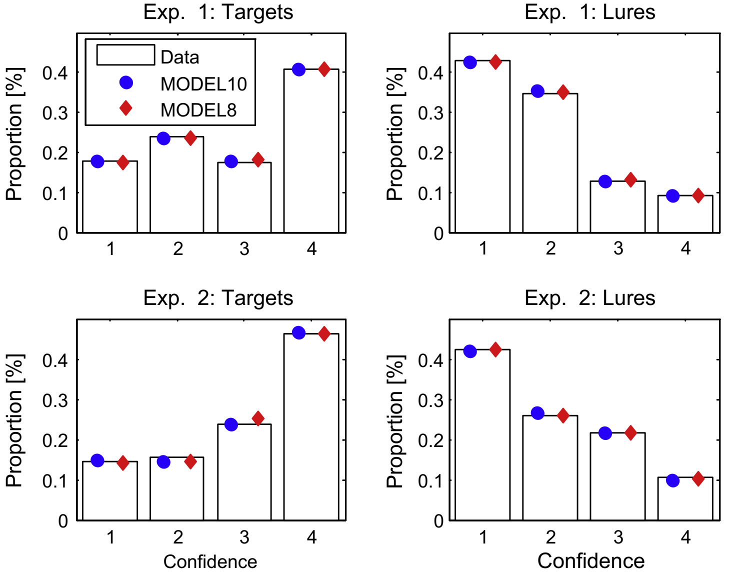

Fig. 4. Response proportions for each of the four confidence categories in Exp. 1, top panels, and 2, bottom panels (Wixted & Mickes, 2010). The left and right panels correspond to targets and lures, respectively. Each panel displays the empirical data (bars) and the model predictions (MODEL10—circles, MODEL8—diamonds). Both the empirical and predicted proportions are presented for the entire group, obtained by averaging empirical or predicted proportions across individuals.

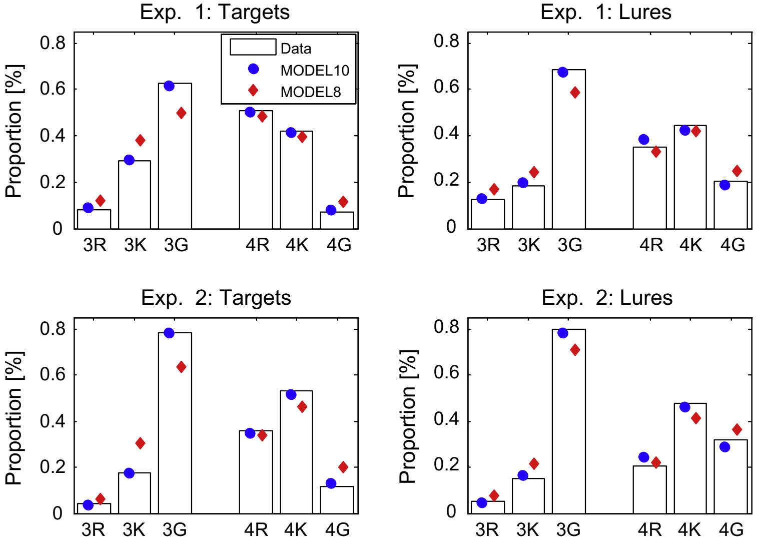

Fig. 5. Response proportions for remember/know/guess judgments for the 'old' confidence categories, 3 and 4, in Exp. 1 and 2 (Wixted & Mickes, 2010). The arrangement of the panels is similar to that of Fig. 4.

Table 4
ML parameters, averaged across participants, for MODEL10 and MODEL8 for Exp. 1 and 2 (Wixted & Mickes, 2010).

Model	Experiment 1		Experiment 2	
	MODEL10	MODEL8	MODEL10	MODEL8
$\mu_{F,T}$	1.10 (0.18)	1.20 (0.15)	1.16 (0.13)	1.13 (0.10)
$\mu_{R,T}$	0.98 (0.14)	0.65 (0.16)	1.14 (0.10)	0.70 (0.09)
$\sigma_{F,T}$	1.72 (0.36)	1.00 (0.00)	1.58 (0.13)	1.00 (0.00)
$\sigma_{R,T}$	1.20 (0.12)	1.64 (0.15)	1.02 (0.10)	1.39 (0.12)
ρ	0.43 (0.09)	0.00 (0.00)	0.58 (0.06)	0.00 (0.00)
c_2	−0.60 (0.33)	−0.44 (0.26)	−0.40 (0.26)	−0.33 (0.20)
c_3	1.50 (0.22)	1.25 (0.18)	0.94 (0.18)	0.76 (0.14)
c_4	2.59 (0.23)	2.20 (0.17)	2.42 (0.15)	1.93 (0.10)
c_R	1.82 (0.16)	1.78 (0.15)	2.11 (0.11)	1.99 (0.11)
c_K	1.54 (0.15)	1.48 (0.13)	1.44 (0.15)	1.38 (0.12)

Note. Brackets show standard errors. For MODEL10, none of the parameters differed significantly across experiments ($p < .05$) according to a t-test. For MODEL8, only c_3 differed across experiment ($p = .034$), but this difference was not significant following application of a Bonferroni correction for multiple comparisons.

Table 4 displays the best fitting parameters for both MODEL8 and MODEL10, averaged across participants. Fig. 6, presents 'model atlases' that are based on the average parameters for each model. Readers might find it helpful to consult this figure when reading the results below.

3.3. Recollection-familiarity correlation

Considering first the RFC across items, MODEL10 provides consistent support for a strong positive correlation ($\rho = .43$, $t(29) = 4.72$, $p < .001$ for Exp. 1; $\rho = .58$, $t(31) = 9.20$, $p < .001$ for Exp. 2). The mean correlation across both experiments was $\rho = .51$. These strong MODEL10 RFCs manifest in the tilt of the ellipse-distribution contours (see top panels of Fig. 6). In contrast, for MODEL8 (see bottom panels)

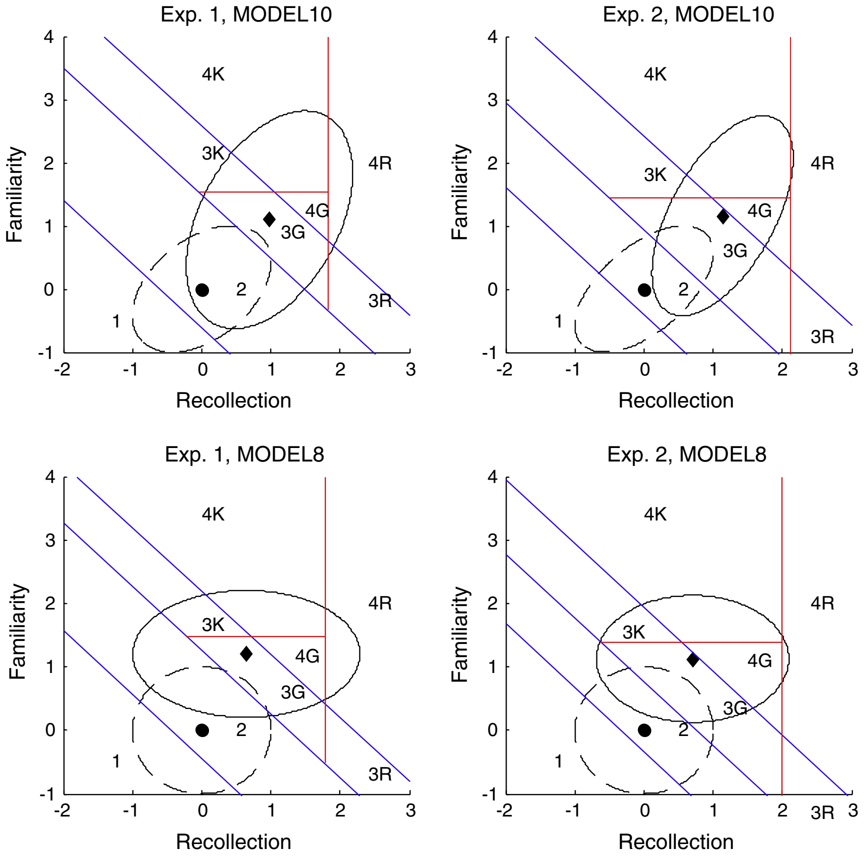

Fig. 6. Model atlases for Exp. 1, left panels, and Exp. 2, right panels (Wixted & Mickes, 2010). The top and bottom panels correspond to the single correlation and unconstrained familiarity variance model (MODEL10) and to the common-currency model (MODEL8), respectively. Each atlas (panel) displays mutual process (recollection and familiarity) distributions for targets and lures (contours in ellipses, dashed for lures, solid for targets), three confidence criteria (blue lines), the 'remember' criterion (vertical red line) and the 'know' criterion (horizontal red line). The set of criterion-lines divide the atlas into eight regions ('states') corresponding to the eight response categories (denoted by the labels 1, 2, 3R, 3K, 3G, 4R, 4K, 4G). Ellipse centers are denoted by a filled circle (lures) or diamond (targets). These atlases are based on the average parameters listed in Table 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the ellipse axes are perpendicular to the main recollection and familiarity axes, a descriptive hallmark of independent processes.

Interestingly, the positive correlations obtained in MODEL10 are similar in magnitude to the correlation reported by Tousignant and Bodner (2012; $r \sim 0.60$).[7] This may imply that if Tousignant and Bodner's recollection and familiarity judgments were influenced by cross contamination it was only to a moderate extent. Note, however, that Tousignant and Bodner used different material than ours so hypothetically, an actual lower correlation might have been inflated to the measured values by substantial cross-contamination. Importantly, the current findings support a positive correlation, in a task which is void of concerns regarding cross rating contamination.

[7] Brown and Bodner (2011) report higher correlations of ~.80. Nevertheless, these correlations were calculated across all items, including both targets and lures. Hence, these correlations may have been subject to a probe-type mediation effect, whereby targets evoke both higher recollection and familiarity relative to lures.

3.4. Target familiarity variability

In both experiments, the recovered target-familiarity variability parameters were larger than 1, the lure variability (for Exp. 1, $\sigma_{F,T} = 1.72$, $t(29) = 2.01$,one tailed, $p < .05$; for Exp. 2, $\sigma_{F,T} = 1.58$, $t(31) = 4.36$, $p < .001$).[8] The larger variance for the target familiarity distribution is evident in the top panels of Fig. 6, where the target ellipses are wider with respect to the familiarity axes than the lure-ellipses. In the bottom, MODEL8 panels, on the other hand, the target and lure ellipses are equally wide with respect to familiarity depicting the equal-variance assumption.

3.5. Target recollection variability

The findings with respect to the RFC and target familiarity variability bear implications with respect to the issue of measurement biases. An important statistic that constrains target-variance in recognition models is the slope of the zROC curve. In CDP, the slope of z-ROC curve is the ratio of the lure and target standard deviations of the aggregated signal (familiarity + recollection): $\frac{\sqrt{2+2\rho}}{\sqrt{\sigma_{R,T}^2+\sigma_{F,T}^2+2\rho_T\sigma_{R,T}\sigma_{F,T}}}$. Note that $\sigma_{R,T}$, $\sigma_{F,T}$ and ρ all tradeoff in generating the zROC slope. Comparing MODEL8 to MODEL10, $\sigma_{F,T}$ and ρ obtain different values. Thus, it would be surprising if $\sigma_{R,T}$ remained invariant across models. The estimated, $\sigma_{R,T}$ parameters confirm this reasoning as they are larger for the MODEL8 than for MODEL10, ($t(29) = 3.8$, $p < .001$, for Exp. 1; $t(31) = 4.38$, $p < .001$, for Exp. 2). This demonstrates that if the common currency assumptions are wrong they would yield biases in estimating recollection-variability for targets.

Remarkably, as compared to MODEL8, in MODEL10, $\sigma_{R,T}$ is reduced (to the effect that they fail to be significantly above 1, the standard deviation of lure recollection.[9] Furthermore, in both experiments, according to MODEL10, the familiarity variability for targets is larger (and significantly so for Exp. 2) than the recollection variability (for Exp. 1, $t(29) = 1.31$, $p > .05$; for Exp. 2, $t(31) = 3.07$, $p < .01$). This can be seen in the top panels of Fig. 6, wherein the target contours are much wider with respect to the familiarity axis than the recollection axis. Furthermore, the target and lure contours seem to be roughly equal in wideness with respect to recollection. In the bottom MODEL8 panels, for comparison, the lure and target contours are equally wide with respect to familiarity (egalitarianism), but the target contours are much wider recollection-wise.

This unexpected finding suggests a surprising theoretical possibility. Recall that in most dual-models of recognition memory, wherein it is assumed that familiarity is equally variable for targets and lures, the subunit z-ROC slopes are attributed to differences between targets and lures with respect to recollection. The current findings imply that familiarity, rather than recollection, may be primarily responsible for the higher overall variability for targets relative to lures and consequently for the subunit z-ROC slopes.

3.6. Target-lure sensitivity

In this section, we show that in erroneously adopting the common-currency assumptions, one can introduce severe biases in estimates of sensitivity for the two processes. Table 4 shows that the mean-recollection parameter for targets, $\mu_{R,T}$, was larger for MODEL10 than for MODEL8, ($t(29) = 2.61$, $p < .05$ for Exp. 1; $t(31) = 4.52$, $p < .001$ for Exp. 2). Recall that target recollection variability $\sigma_{R,T}$ was lower for MODEL10. Thus, MODEL10 should exhibit higher levels of target-lure sensitivity with respect to recollection. To test this, we calculated for each participant, process and model the target-lure sensitivity (d_a; Macmillan & Creelman, 2004):

[8] Because the significance of this finding was marginal in Exp. 1 (the two sided test was non-significant), we further probed this effect with a meta-analysis across both experiments, weighting each experiment's respective statistic by the inverse of the variance of the statistic, assuming a random-effects model (Shadish & Haddock, 1994). This meta-analysis revealed that across experiments, $\sigma_{F,T} = 1.6$, SE = 0.13, which was significantly above 1, $z = 4.66$, $p < .001$.

[9] A meta-analysis (see Footnote 8 for details) revealed that across experiments, $\sigma_{R,T} = 1.1$, SE = .09, which did not differ significantly from 1, $z = 1.10$, $p = .27$.

$$d_a^P = \frac{\mu_{P,T} - \mu_{P,L}}{\sqrt{0.5 * (\sigma_{P,T}^2 + \sigma_{P,L}^2)}}, \quad P = R, F, \tag{5}$$

As Table 5 shows, the mean recollection-sensitivity estimates were about twice as large in MODEL10 that in MODEL8 ($t(29) = 4.55$, $p < .001$ for Exp. 1; $t(31) = 5.72$, $p < .001$ for Exp. 2).

As for familiarity, the mean-target familiarity $\mu_{F,T}$ was roughly equivalent in both model variants but variability was higher (>1) in MODEL10. Thus MODEL10 reflects a reduced level of target-lure familiarity-based sensitivity. Indeed, Table 5 reveals mean sensitivity estimates that are lower by ~30% for MODEL10 than for MODEL8 ($t(29) = 2.11$, $p < .05$, for Exp. 1; $t(31) = 2.09$, $p < .05$, for Exp. 2). In conclusion, using the constrained model seems to severely bias estimates of sensitivity.

3.7. Parameter recovery simulation

Inferential statistics based on the estimated MODEL10 parameters, yielded strong evidence supporting a consistent positive RFC and an increased variance of target familiarity across participants. Next, we examined the possibility that these conclusions result from biases in our parameter estimation procedure. Although, in general, ML estimators are asymptotically unbiased, for relatively small samples, as is the case for each individual in our data-sets, there may be systematic biases in estimated parameters. The validity of our conclusions would benefit by ruling out the possibility that they were based on biased estimators.

Thus, we conducted a parameter recovery simulation. For each participant (in each experiment), we used his or her data-estimated MODEL10 parameters (henceforth 'original parameters') to create 200 parametric bootstrap samples. To this end, we calculated MODEL10's predicted probability distribution across the eight response categories (for each probe type). Each of the parametric bootstrap samples was subsequently created by generating independent observations according to this predicted distribution for each probe type (the size of each sample equaled the size of the empirical sample). Next, we fitted each of these bootstrap samples with MODEL10, to obtain recovered parameters. We then calculated the differences between the original parameters and the recovered parameter sets. Averaging these differences across the 200 bootstrap samples, we obtained estimates for the biases in the recovered parameters for each individual. Table 6 lists these biases, averaged across participants. Note that positive (negative) table entries indicate that the generating parameters are larger (smaller) than the recovered parameters. This, in turn means that MODEL10 tends to under (over) estimate the corresponding parameter.

Critically, Table 6 reveals that in neither experiment was there a significant bias in the variability of target-familiarity ($\sigma_{F,T}$). As for the RFC (ρ) there do seem to be significant biases in both experiments (of ~.08). Importantly, these biases suggest that if anything, the MODEL10 correlations reported in Table 4 are *underestimates* so the actual correlations may be larger. If we add the (under) estimation bias, 0.08 to our estimated RFCs we obtain, across Exp. 1 and 2, a mean RFC of .59, which is astonishingly similar to the .6 correlations reported by Tousignant and Bodner (2012). In conclusion, the parameter-recovery simulation findings argue against the possibility that our findings were spuriously based on parametric estimation biases.

Table 5
Target-Lure sensitivity for recollection and familiarity according to each model, averaged across participants for Exp. 1 and 2 (Wixted & Mickes, 2010).

	d_a^R		d_a^F	
	MODEL10	MODEL8	MODEL10	MODEL8
Exp. 1	1.39 (0.21)	0.71 (0.13)	1.22 (0.22)	1.7 (0.31)
Exp. 2	1.63 (0.15)	0.81 (0.14)	1.25 (0.22)	1.59 (0.28)

Note. Brackets show standard errors.

Table 6
Average biases in the recovered parameters across participants in Exp. 1 and 2.

Parameter	Exp. 1	Exp. 2
$\mu_{F,T}$	0.06 (0.04)	−0.02 (0.02)
$\mu_{R,T}$	0.02 (0.03)	0.04 (0.01)*
$\sigma_{F,T}$	−0.03 (0.05)	0.01 (0.01)
$\sigma_{R,T}$	0.02 (0.06)	−0.04 (0.02)
ρ	0.08 (0.02)*	0.08 (0.01)*
c_2	0.03 (0.02)	0.01 (0.01)
c_3	0.09 (0.03)*	0.04 (0.01)*
c_4	0.10 (0.03)*	0.06 (0.01)*
c_R	0.04 (0.01)*	0.05 (0.01)*
c_K	0.05 (0.03)	0.00 (0.01)

Note. Brackets show standard errors.
* Parameters for which the biases differ significantly from 0, according to a *t*-test ($p < .05$).

3.8. Are recollection and familiarity also correlated across participants?

The finding of a positive RFC across items, begs the question if an analogous correlation holds between the efficiencies of the recollection and familiarity processes across participants. To test this, we calculated Pearson's correlation between d_a^R and d_a^F for MODEL10 (see Eq. (5)), across participants. Whereas Exp. 1 revealed a non-significant negative correlation ($r = -.34$, $p = .065$), Exp. 2 revealed a non-significant positive correlation ($r = .34$, $p = .06$). Pooling participants across experiments (d_a^R and d_a^F did not differ significantly across experiments), the correlation was slightly negative but remained non-significant ($r = -.13$, $p = .3$). In summary, the positive RFC across items was dissociated from a correlation between the efficiency of both processes across participants.

Interestingly, in the word-stem completion task, Rouder et al. (2008) found a similar dissociation. For items, recollection correlated positively with automatic activation but participants with higher recollection enjoyed no increased tendency toward automatic activation.

4. General discussion

In the current paper, we conducted a model comparison study based on the CDP model. The model comparison results challenged two of the common-currency assumptions of recognition memory. First, in violation of the assumption that recollection and familiarity are stochastically independent across items, we found a strong positive correlation (~.5) between the magnitudes of these signals. Furthermore, this correlation was equivalent for targets and for lures. Second, we found that across targets, the variance of the familiarity signal is larger than across lures, in violation of the assumption of equal variability.

These findings bear several important implications for the study of recognition memory. First, having supported the existence of a positive RFC, there is now a need to explore its underlying psychological causes. Second, our finding with respect to the variance of target familiarity suggests the surprising possibility that an extensive and rich body of literature may have misattributed the cause underpinning one of the most influential findings of recognition memory, subunit z-ROC slopes, to recollection, when in fact it is familiarity, which is the main culprit. Indeed, according to our parameter estimators, the variability in familiarity is a larger contributor to the higher overall target variability (compared to lures), causing the subunit zROC slopes.

Additionally, we demonstrated that wrongly postulating the common- currency assumptions may severely bias estimators of the processes' contributions to recognition, as evident in target-lure sensitivity (d_a) values. Indeed, when these assumptions were incorporated into the common-currency model (MODEL8), recollection estimates of sensitivity shrunk by about 50% relative to estimates based on a model free of these assumptions (MODEL10). This underestimation was compensated by grossly overestimating (∼30%) familirity sensitivity.

4.1. Interpreting the correlation between recollection and familiarity: from cross-talk to parallel influences

4.1.1. (Functional) Independence is not (stochastic) independence

Despite the positive across-items correlation between recollection and familiarity, the two processes may still be functionally independent, that is, it may be possible to doubly dissociate them under different experimental conditions. Indeed, in his seminal review of 30 years of research on recollection and familiarity, Yonelinas (2002) summarized a plethora of evidence demonstrating functional dissociations between recollection and familiarity. Yet, the existence of such dissociations is in no way diagnostic with respect to the presence or absence of a correlation between recollection and familiarity. Thus, demonstrations of functional independence do not warrant the application of process estimation procedures, which assume stochastic independence.

To illustrate the difference between these forms of independence, consider two accomplished wine tasters, Rob and Fiona, who rate wines without coordinating with each another. Assume that Rob and Fiona share a preference for mature over young wines. Across wines, this shared preference manifests in a positive correlation between their ratings, because they tend to rate mature wines higher than young wines. Nonetheless, it is possible to (doubly) dissociate between their ratings. For example, if Rob prefers a warmer, whereas Fiona prefers a cooler, room temperature then turning on the air condition will tend to increase Fiona's ratings while simultaneously decreasing Rob's rating. Notably, following the 'air-condition manipulation' their ratings remain positively correlated as they still favor maturity. Thus, despite the positive correlation across wines, the ratings of Rob and Fiona can be functionally independent. As eloquently stated by Tulving (1985a, p. 394): "It is important to realize that there is no necessary logical connection between these two kinds of independence. We could say that 'independence is not independence'".

4.1.2. Crosstalk

We now turn to interpret the underlying cause for the positive RFC. One possible interpretation for this correlation is that there is crosstalk between the two processes to the effect that they influence each other directly, or that they are both influenced by a common third source/process. Findings in the neuro-scientific literature may point to a potential anatomical locus for a crosstalk between recollection and familiarity. Specifically, familiarity has been hypothesized to depend on the functioning of extra-hippocampal medial temporal lobe (MTL) regions, with bi-directional connections to the hippocampus proper, on which recollection has been hypothesized to depend (e.g., Eichenbaum, Yonelinas, & Ranganath, 2007; Norman & O'Reilly, 2003, but see Smith, Wixted, & Squire, 2011).

Interestingly, the Source of Activation Confusion model (SAC; Diana et al., 2006; Oates, Reder, Cook, & Faunce, 2015; Reder et al., 2000) is compatible with the notion of a direct influence between familiarity and recollection. In a nutshell, SAC postulates a spreading activation network of connections between various types of nodes. The experience of familiarity is based on activation of 'concept' nodes or spuriously on 'specific context' nodes, whereas the experience of recollection is based on activation of 'episode' nodes. Importantly, in SAC the activation levels of familiarity-governing unit serves as input, feeding a recollection-governing unit with the property that the magnitude of this input is proportional to the activation of the feeding unit. This SAC mechanism may contribute to a positive correlation between the activation levels of familiarity and recollection units, reflected in a positive RFC.

4.1.3. Parallel influences

Despite the theoretical allure of 'direct influence' or cross-talk accounts, however, the evidence from the correlation across items that we have found is too weak to support such a conclusion. The observed correlation may simply be a product of variables exerting parallel influences on both processes across items. One source of 'parallel influences' may reside in variables that generally affect the efficiency of encoding e.g., levels-of-encoding (Craik & Lockhart, 1972). An influential conceptualization of encoding distinguishes between item specific information and contextual, associative information (Murdock, 1974), associated with familiarity and recollection, respectively. If variables exert a similar influence on our ability to encode both item and associative information, then a poorly encoded memory trace will tend to represent little of both kinds of information, whereas a well-encoded trace will represent a lot of both.

The notion of levels-of-encoding may also account for our finding of equal RFC for targets and lures. Specifically, we conceptualize both lures and targets to be retrieval cues that activate memory traces (c.f., Dennis & Humphreys, 2001), for targets, activated traces which were encoded during the study phase, whereas for lures, activated traces which were encoded pre-experimentally. For both targets and lures therefore, activating a well-encoded trace, relative to a poorly-encoded trace, will tend to retrieve higher amounts of both item and associative information. Variability in encoding efficiency, in turn, may be a direct consequence of inevitable variability in 'item-difficulty' (e.g., in the Wixted & Mickes, 2010, experiments that were analyzed here, items differed in word-abstractness and word-length).

4.1.4. Levels of stochastic independence and their constraint on theory

The above conclusions can be summarized using the 'levels of SIA' terminology. Whereas parallel-influence accounts attribute violations of stochastic independence across items to correlated *item*-effects (between recollection and familiarity), cross-talk accounts introduce a correlation in *trial* effects and hence violate the SIA at the lower, participant $*$ item level. Our across-items correlation findings support only the weaker conclusion that the across-items level (but not the participant $*$ item level) of the SIA is violated. Despite its consistency with our data, the operation of a crosstalk mechanism would thus be too strong a conclusion.

In the current study, we focused on individual-participant data (pooling across items). However, whenever participants are tested on the same items, data for a given item may likewise be pooled across participants. Importantly, our model is readily extendable to this pooling-scheme with a single interpretive difference that ρ now corresponds to an RFC across participants.

Rather than pooling data across participants, here we probed the across-participants correlation between recollection and familiarity based on the estimated individual sensitivities (d_a; Eq. (5)) for both mnemonic processes. Remarkably, we found no consistent RFC across participants. This finding is consistent with the interpretation that the efficiencies of recollection and familiarity are stochastically independent across participants but are subject to parallel influences (or crosstalk) within individual minds. Future studies should examine whether similar findings are obtained by fitting the model to individual-item data pooled across participants.

4.1.5. Overcoming aggregation biases

As mentioned in the introduction, some readers may advocate the alternative interpretation that the assumptive SIA level is the participant $*$ item level and that an RFC across items exerts a bias *due to aggregation* per se (rather than due to the violation of the SIA across items). To reiterate, we advocate the across-item interpretation, because if the SIA is restricted to the participant $*$ item level, then there is a theoretical gap between the level in which a model's mathematical equations are valid (participant $*$ item), and the level in which the data are analyzed (pooling across items). However, this difference in interpretation should not detract from the importance of the current approach even for proponents of the aggregation bias interpretation. Indeed, the incorporation of the RFC parameter ρ into our model constitutes a corrective procedure, which serves the goal of avoiding aggregation biases.

Interestingly, in recent years, at least two general-purpose methods have been developed to avoid the problem of aggregation bias that occurs due to pooling across (heterogeneous) populations of participants and/or items. The first comprises the latent-class multinomial processing tree models (Klauer, 2006), which addresses participant-heterogeneity. The second is an independent random effects signal-detection model (Rouder et al., 2007), which addresses heterogeneity in both participants and items in signal detection tasks. Additionally, in the context of dual-process memory models, Rouder et al. (2008) developed a hierarchical process-dissociation model, which addresses the problem of aggregation-biases that may emerge in applications of the standard PDP to the word-stem completion task due to across-items (and participants) process correlations. This model disposes of the necessity to pool data (across items, participants or both) by postulating a generalized linear model (GLM) structure in which participants and items combine additively qua random effects on the recollection and the automatic processes. Notably, participants' and/or items' random effects on these two processes are allowed to correlate.

There are several fundamental differences between the Rouder et al. (2008) approach and ours (besides the fact that they relate to different paradigms). First, the hierarchical process-dissociation model utilizes a Bayesian framework for parameter estimation (in particular, a prior distribution for the parameters of the model must be provided), whereas our approach is 'classical'. Second, data-analysis based on the hierarchical model dispenses with the necessity to pool data (either across participants or across items). In contrast, we analyze aggregated data with a model that explicates the effect of an RFC across items. Third, and most importantly, the hierarchical process-dissociation model maintains the SIA at the participant $*$ item level and may thus fail if the SIA is violated at that level. Our model, on the other hand, dispenses with the SIA altogether as the ρ parameter is sensitive to all sources of correlation contributing to the across item level, including trial effects.

4.2. Unequal Variability of the familiarity distribution

An important conclusion of our analyses is that for targets, the variability of the familiarity signal is larger than for lures. This conclusion is somewhat surprising in light of two empirical findings, which are commonly cited in support of the notion of equivalent target-lure variability of the familiarity signal. First, systematic investigations of amnesiacs revealed minimal recollection in these patients with spared, intact familiarity (Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998). Importantly, for these recollection-impaired patients a unit zROC slope was revealed. Second, Yonelinas (2001) showed that following selective removal of remember responses for healthy participants, the subunit zROC slope increased to 1. Both of these results were interpreted as reflecting the sole operation of familiarity, which apparently conformed to an equal-variance distribution.

Although this pair of findings is indeed consistent with the DPSD notion that familiarity is an equal-variance signal-detection process, an alternative interpretation is that they are results of memory traces, which are weak (cf., Glanzer, Kim, Hilford, & Adams, 1999) either because of the damage incurred by amnesia (Squire, Wixted, & Clark, 2007), or as a result of the selective removal of the strongest memory traces (indeed, in Yonelinas, 2001, remember responses were almost invariably rated with the highest confidence). According to this weak-memory interpretation, studied items enjoy only minor gains to their strength, in magnitude and more importantly, in variability. As these gains affect the strength variability only to a minor extent, the zROC slope, which reflect the relative lure-target strength-variability, is higher relative to strong memories for which study gains increase the strength variability more substantially. In conclusion, this pair of findings can be reconciled with our conclusion of non-equal target-lure variability of the familiarity signal for items comprising a non-restricted range of memory strengths.

4.3. CDP vs. alternate models of recognition

The conclusions of the current study hinge on the validity of the CDP as a model of human performance, as they are based on comparisons of CDP variants and on CDP parameter-estimates. Our study, however, bears general important implications for the study of recognition memory, including for alternative models. Consider first the debate between the families of single-process and dual-process models of memory. According to single-process models, recognition is governed by a unique 'memory strength' (mnemonic evidence) process rather than by two distinct processes. To the extent that dual-process models, such as CDP, gain in their ability to account for empirical data as the common-currency assumptions are forfeited, so does their stance improve vis. a vis. single-process models. By the same token, improving the CDP also elevates its position vis. a vis. alternative dual-process models.

Future modelers should examine whether alternative dual-process models also benefit from relaxing the common- currency assumptions. As explained in the introduction, in DPSD (Yonelinas, 2002) and VRDP (Onyper et al., 2010; See Footnote 3), however, the SIA appears to be irrefutable. We argue that the current findings, which converge with prior conclusions (Brown & Bodner, 2011; Curran & Hintzman, 1995; Rouder et al., 2008; Tousignant & Bodner, 2012), demonstrate the importance of testing this assumption. Indeed, since in these models familiarity affects performance only when target-recollection fails, estimators of familiarity are influenced only by a subpopulation of the targets—the non-recollected targets. Importantly, in the presence of a positive RFC, familiarity would be stronger for recollected than for non-recollected targets. Hence, familiarity estimates that are insensitive to familiarity for recollected targets will necessarily be biased downwards (see also our example from Section 1.1.3).

We thus put it forward to the proponents of the SIA, to devise methods for subjecting it to a valid empirical test. One such method was devised recently by Pratte and Rouder (2011), who constructed a *hierarchical*-DPSD model. This model enables the estimation of item effects on recollection and familiarity and allows one to probe whether these effects are correlated (across items). Interestingly, this model revealed an intricate relationship between the two processes in the form of a negative across items correlation between recollection and *baseline familiarity*: the lower the familiarity of an item prior to study, the larger the probability that it will be recollected following study.

Turning to familiarity–variability, the DPSD and VRDP models can accommodate unequal target-lure variability of the familiarity signal. Future research could explore the possibility that these models benefit from variance inequality. In summary, we argue that it is essential that the common currency assumptions be tested rather than taken as foundational assumptions. The current paper consists of a 'proof of concept' for both the possibility and the importance of conducting such tests.

Acknowledgments

We thank John Wixted and Laura Mickes for sharing their empirical data. We also thank John Wixted for helpful discussions, and Caren Rotello, Michael Gilead, Charles Brainerd, Carlos Gomes, Andrew Heathcote, Eric-Jan Wagenmakers, John Dunn and Adam Osth for thoughtful comments on earlier versions of this MS.

Appendix A. Fitting the CDP model variants to the empirical data

A.1. Deriving model predictions

In the current section we derive the CDP model predictions for the probabilities of each of the response types (1, 2, 3R/K/G, 4R/K/G in our case), given a set of model parameters (see Table 1). This derivation is presented for the most general model variant, the non-constrained model (MODEL11) but it can readily be applied to all the simpler variants (MODELS10, 9C, 9F, 8) by imposing the appropriate constraints on parameter values (see Table 2). This derivation generalizes the one presented in the appendix of Wixted and Mickes (2010), which dealt only with the common-currency version and did not include 'Guess' responses.

Throughout the section p will denote a probe type, i.e. a target (T) or a lure (L). The equations for target and lures are identical, each involving its own parameters. Additionally, $\varphi(x;\mu,\sigma^2)$ will denote the Gaussian density with mean μ and variance σ^2 and $\Phi(x;\mu,\sigma^2)$ will similarly denote the Gaussian cumulative density function.

Recall that in CDP recollection and familiarity are mutually distributes as $(R_p,F_p)\sim N\left((\mu_{R,p},\mu_{F,p}),\begin{pmatrix}\sigma^2_{R,p} & \rho_p\sigma_{R,p}\sigma_{F,p}\\ \rho_p\sigma_{R,p}\sigma_{F,p} & \sigma^2_{F,p}\end{pmatrix}\right)$ and that their sum is thus distributed $M_p=R_p+F_p\sim N(\mu_{F,p}+\mu_{R,p},\sigma^2_{R,p}+\sigma^2_{F,p}+2\rho_p\sigma_{R,p}\sigma_{F,p})$. Another useful property of the bivariate Gaussian distribution is that conditional on one signal (R or F), the other is still Gaussian. For example:

$$F_p|R_p=r\sim N\left(\mu_{F,p}+\frac{\sigma_{F,p}}{\sigma_{R,p}}\rho_p(r-\mu_{R,p}),(1-\rho_p^2)\sigma^2_{F,p}\right), \quad \text{(A1)}$$

The proportion of responses with confidence below $i = 2, 3, 4$ is given by:

$$P(M_p \leqslant c_i) = \Phi(c_i; \mu_{F,p} + \mu_{R,p}, \sigma^2_{R,p} + \sigma^2_{F,p} + 2\rho_p\sigma_{R,p}\sigma_{F,p}), \tag{A2}$$

From Eq. (A2), one can calculate the probabilities for the different confidence responses by:

$$\begin{aligned} p_1 &= P(M_p \leqslant c_2) \\ p_i &= P(M_p \leqslant c_{i+1}) - \sum_{j=1}^{i-1} p_j \quad i = 2,3 \\ p_4 &= 1 - \sum_{j=1}^{3} p_j, \end{aligned} \tag{A3}$$

For the new, '1' or '2' confidence responses, which are not followed by an R/K/G judgment, these calculations provide the response probabilities. However, for the 'old' confidence responses, '3' and '4', these responses should be divided between the R|K|G judgments. Let $i = 3, 4$ denote and 'old' confidence level. For remember judgments the probability for confidence i or above, combined with a remember judgment is given by:

$$\begin{aligned} p_{\geqslant i,R} &= P(R_p + F_p > c_i \text{ and } R_p > c_R) = \int_{c_R}^{\infty} \varphi(r; \mu_{R,p}, \sigma^2_{R,p}) P(F_p > c_i - r | R_p = r) dr \\ &= \int_{c_R}^{\infty} \varphi\left(r; \mu_{R,p}, \sigma^2_{R,p}\right) \left(1 - \Phi\left(c_i - r; \mu_{F,p} + \frac{\sigma_{F,p}}{\sigma_{R,p}} \rho_p (r - \mu_{R,p}), (1 - \rho_p^2)\sigma^2_{F,p}\right)\right) dr, \end{aligned} \tag{A4}$$

Proceeding to know judgments, the probability of a combination of confidence i or above and a know judgment is given by:

$$\begin{aligned} p_{\geqslant i,K} &= P(R_p + F_p > c_i \text{ and } R_p \leqslant c_R \text{ and } F_p > c_K) \\ &= \int_{-\infty}^{c_R} \varphi(r; \mu_{R,p}, \sigma^2_{R,p}) P(F_p > c_i - r \text{ and } F_p > c_K | R_p = r) dr \\ &= \int_{-\infty}^{c_R} \varphi(r; \mu_{R,p}, \sigma^2_{R,p}) P(F_p > \max(c_i - r, c_K) | R_p = r) dr \\ &= \int_{-\infty}^{c_R} \varphi(r; \mu_{R,p}, \sigma_{R,p}) \left(1 - \Phi\left(\max(c_i - r, c_K); \mu_{F,p} + \frac{\sigma_{F,p}}{\sigma_{R,p}} \rho_p (r - \mu_{R,p}), (1 - \rho_p^2)\sigma^2_{F,p}\right)\right) dr, \end{aligned} \tag{A5}$$

Note that $p_{\geqslant i,R}$ and $p_{\geqslant i,K}$ are cumulative probabilities. As before (Eq. (A3)), we obtain the marginal probabilities, i.e. the probability of a combination of confidence i and a judgment $j = R, K$ by differentiating:

$$\begin{aligned} p_{4,J} &= p_{\geqslant 4,J} \\ p_{3,J} &= p_{\geqslant 3,J} - p_{4,J}, \end{aligned} \tag{A6}$$

Finally, a guess judgment is issued if and only if neither a remember nor a know judgment is given. Thus, the probability for a combination of confidence i and a guess judgment is given by

$$p_{i,G} = p_i - p_{i,R} - p_{i,K}, \tag{A7}$$

Together, Eqs. (A3) (p_1, p_2), (A6) ($p_{3,R}$, $p_{4,R}$, $p_{3,K}$, $p_{4,K}$) and (A7) ($p_{3,G}$, $p_{4,G}$) provide the probabilities for the eight different response categories. In the following section we abuse notation and simply denote these eight probabilities by $p_1, p_2, \ldots, p_8$.

A.2. Fitting the model variants

Given a profile of free model parameters, one can compute, using the formulas above, the model-predicted probabilities of responses in each of the eight response category for each probe type $p_{i,T}, p_{i,L}, 1 \leqslant i \leqslant 8$. Given these predictions and the empirical frequencies of responses in each category

$N_{i,T}, N_{i,L}, 1 \leqslant i \leqslant 8$, one can calculate the likelihood of the data as $L = \prod_{i=1}^{8}\prod_{j=T,L} p_{i,j}^{N_{i,j}}$. In ML estimation, the 'best fitting' parameters were defined as the combination of free model parameters that maximized the likelihood of the data.[10]

We began by fitting MODEL8 to the individual data. The search for the ML parameters was conducted with a combination of a genetic algorithm and the iterative Nelder and Mead (1965) Simplex method, (implanted by the routines "ga", "fminsearch" available in Mathwork's MATLAB). The genetic algorithm was used to construct a starting point for the simplex routine (i.e. the output of the genetic algorithm was fed as a starting point to the simplex algorithm).

Next, we used the best-fitting 8-tupple (of MODEL8) to fit the more complex model variants (MODELS 9Corr, 9Fam, 10, 11). First, we augmented the best fitting 8-tupple with entries for the 'novel' free parameters. These novel entries were 0, for a correlation parameter, or 1 for a target familiarity standard deviation parameter. The augmented vector served as a starting point for the simplex routine for the extended model. For example, when fitting the non-constrained model (MODEL11), we took the best fitting 8-tupple (for MODEL8) and extended it to an 11-tupple by adding two 0 entries (for the recollection-familiarity correlation for lures and for targets) and a single 1 entry (for the target familiarity standard deviation). We then used this 11-tupple as a starting point for a simplex search in the 11-dimensional parameter space. Similarly, when fitting model 9Corr, we extended the 8-tupple with a single zero entry (for the single correlation parameter) and used the 9-tupple as a starting point for a simplex search in the nine-dimensional parameters space.

In the BGLRT analysis (MODEL8 and MODEL10) and in the parameter recovery simulation (MODEL10) we fitted the models using the same principles utilizing the simplex algorithm. In these analyses (in addition to fitting the models to empirical data) we fit the models to parametric or to non-parametric bootstrap samples. The parameters obtained from the model-fits to the empirical data served as a starting point for fitting non-parametric bootstrap samples. The generative parameters of a parametric bootstrap samples served as a starting point for fitting this sample. Thus, in BGLRT, we begin by taking a non-parametric bootstrap sample for the empirical data and fit MODEL8 to this sample. In this fit the ML parameter for the empirical data serve as a starting point. Next the ML parameters of the (non-parametric) bootstrap sample serve as generative parameters for parametric bootstrap samples. When fitting this parametric bootstrap sample with MODEL8, these generative parameters served as a starting point. After the best-fitting 8-tupple was obtained it was used to fit MODEL10 to the same (parametric bootstrap) sample (by extending it with a zero-correlation entry and an unit standard deviation entry and using this 10-tupple as a starting point for MODEL10). Similarly, in the parameter recovery simulation, we used the best fitting MODEL10 parameters to the empirical data to generated parametric bootstrap samples. We then use these same parameters as a starting point to fit MODEL10 to the parametric bootstrap sample.

A.3. Model identifiability

The parameters of a model are *identifiable* if there is a one to one mapping between model parameters and behavioral predictions. When parameters are non-identifiable, interpreting their estimators is problematic, because different parameter combinations yield identical predictions. To verify the identifiability of the models in our inventory (Table 2) we reran our fitting procedures many times with different starting points (generated by a genetic algorithm. as described above). These reruns always yielded the same parameters when the globally ML value was achieved.

[10] The calculation of model predictions $p_{i,j}$ involves the numeric calculation of integrals over infinite ranges. When the results of these integrals are very close to 0, these numerical computations are relatively inaccurate. For some of the participants, some of the response categories were empty, which drove the simplex search to this 'problematic' range of very small integrals. To facilitate our fitting routine, so it avoids such 'problematic ranges', we added 'one artificial sample' and divided it equally across the eight response categories. Thus, when calculating ML the empirical counts $N_{i,T}, N_{i,L}$ were 'adjusted' by adding 1/8 to each of them. Note that the empirical data contains $\sim$128 samples for each probe type so this single sample corresponds to less than 1% of the data. This 'adjustment' was also applied to all the fits (of bootstrap samples) in our BGLRT and parameter recovery simulation. Thus, these analyses control for the effect of this adjustment. Particularly, any (slight) biases to the estimated model parameters caused by this 'adjustment' will manifest in the parameter recovery simulation.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *International symposium on information theory* (2nd ed., pp. 267–281). Tsahkadsor, Armenian SSR.

Brainerd, C. J., Gomes, C. F. A., & Moran, R. (2014). The two recollections. *Psychological Review, 121*(4), 563–599.

Brown, A. A., & Bodner, G. E. (2011). Re-examining dissociations between remembering and knowing: Binary judgments vs. independent ratings. *Journal of Memory and Language, 65*(2), 98–108.

Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review, 15*(4), 692–712.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684.

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

Curran, T., & Hintzman, D. L. (1997). Consequences and causes of correlations in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 496–504.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*(2), 452–478.

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review, 13*(1), 1–21.

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

Dunn, J. C. (2004). Remember–know: A matter of confidence. *Psychological Review, 111*, 524–542.

Eichenbaum, H., Yonelinas, A. R., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30*, 123–152.

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review, 12*(3), 403–408.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition, 7*(1), 126.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 500.

Higham, P. A., & Vokey, J. R. (2004). Illusory recollection and dual-process models of recognition memory. *Quarterly Journal of Experimental Psychology, 57A*, 714–744.

Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411–428.

Hintzman, D. L., & Curran, T. (1997). More than one way to violate independence: Reply to Jacoby and Shrout (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 511–513.

Hirshman, E. (1998). On the logic of testing the independence assumption in the process-dissociation procedure. *Memory & Cognition, 26*(5), 857–859.

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 325–339.

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

Jacoby, L. L., & Shrout, P. E. (1997). Toward a psychometric analysis of violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 505–510.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122*, 139–154.

Jang, Y., Mickes, L., & Wixted, J. T. (2012). Three tests and three corrections: Comment on Koen and Yonelinas (2010). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 513–523.

Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *Journal of Experimental Psychology: General, 138*(2), 291–306.

Jang, Y., Wixted, J. T., & Huber, D. E. (2011). The diagnosticity of individual data for model selection: Comparing signal-detection models of recognition memory. *Psychonomic Bulletin & Review, 18*(4), 751–757.

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika, 71*(1), 7–31.

Koen, J. D., & Yonelinas, A. P. (2010). Memory variability is due to the contribution of recollection and familiarity, not to encoding variability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1536–1542.

Koen, J. D., & Yonelinas, A. P. (2013). Still no evidence for the encoding variability hypothesis: A reply to Jang, Mickes, and Wixted (2012) and Starns, Rotello, and Ratcliff (2012). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 304–312.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology Press.

Mandler, G. (1959). Stimulus variables and subject variables: A caution. *Psychological Review, 66*, 145–149.

Murdock, B. B. (1974). *Human memory: Theory and data*. Hillsdale, NJ: Erlbaum.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44*(1), 190–204.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal, 7*(4), 308–313.

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

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110*(4), 611–646.

Oates, J. M., Reder, L. M., Cook, S., & Faunce, P. (2015). Spurious recollection from a dual-process framework: The role of partial matching. In J. G. W. Raaijmakers, A. H. Criss, R. L. Goldstone, R. M. Nosofsky, and M. Steyvers (Eds.), *Cognitive modeling in perception and memory: A festschrift for Richard M. Shiffrin*. New York: Psychology Press.

Onyper, S. V., Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General, 139*(2), 341–364.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in cognitive sciences, 6*(10), 421–425.
Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*, 472–491.
Pratte, M. S., & Rouder, J. N. (2011). Hierarchical single-and dual-process models of recognition memory. *Journal of Mathematical Psychology, 55*(1), 36–46.
Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data form recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 763–785.
Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99*(3), 518–535.
Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 294–320.
Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95*(3), 318–339.
Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal-detection model. *Psychological Review, 111*(3), 588–616.
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.
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*(4), 621–642.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.
Shadish, W. R., & Haddock, K. C. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York, NY: Russell Sage Foundation.
Smith, C. N., Wixted, J. T., & Squire, L. R. (2011). The hippocampus supports both recollection and familiarity when memories are strong. *The Journal of Neuroscience, 31*(44), 15693–15702.
Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience, 8*(11), 872–883.
Starns, J. J., Rotello, C. M., & Ratcliff, R. (2012). Mixing strong and weak targets provides no evidence against the unequal variance explanation of zROC slope: A comment on Koen and Yonelinas (2010). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 793–801.
Tousignant, C., & Bodner, G. E. (2012). Test context affects recollection and familiarity ratings: Implications for measuring recognition experiences. *Consciousness & Cognition, 21*, 994–1000.
Tulving, E. (1985a). How many memory systems are there? *American Psychologist, 40*(4), 385–398.
Tulving, E. (1985b). Memory and consciousness. *Canadian Psychologist, 26*, 1–12.
Wagenmakers, E. J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology, 48*(1), 28–50.
Wixted, J. T. (2007). Dual-process and signal detection theory of recognition memory. *Psychological Review, 114*, 152–176.
Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*(4), 1025–1054.
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.
Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130*, 361–379. http://dx.doi.org/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*(3), 441–517.
Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34*, 622–643.
Yonelinas, A. P., Kroll, N. E., 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*(3), 323–339.