

Similar time course of fast familiarity and slow recollection processes for recognition memory in humans and macaques

Zhemeng Wu, Martina Kavanova, Lydia Hickman, Fiona Lin, and Mark J. Buckley

Department of Experimental Psychology, University of Oxford, Oxford OX1 3SR, United Kingdom

According to dual-process theory, recognition memory performance draws upon two processes, familiarity and recollection. The relative contribution to recognition memory are commonly distinguished in humans by analyzing receiver-operating-characteristics (ROC) curves; analogous methods are more complex and very rare in animals but fast familiarity and slow recollective-like processes (FF/SR) have been detected in nonhuman primates (NHPs) based on analyzing recognition error response time profiles. The relative utility of these methods to investigate familiarity and recollection/reollection-like processes across species is uncertain; indeed, even how comparable the FF/SR measures are across humans and NHPs remains unclear. Therefore, in this study a broadly similar recognition memory task was exploited in both humans and a NHP to investigate the time course of the two recognition processes. We first show that the FF/SR dissociation exists in this task in human participants and then we demonstrate a similar profile in the NHP which suggests that FF/SR processes are comparable across species. We then verified, using ROC-derived indices for each time-bin in the FF/SR profile, that the ROC and FF/SR measures are related. Hence, we argue that the FF/SR approach, procedurally easier in nonhuman animals, can be used as a decent proxy to investigate these two recognition processes in future animal studies, important given that scant data exists as to the neural basis underlying recollection yet many of the most informative techniques primarily exist in animal models.

Recognition memory, a form of declarative memory according to some authors (Squire and Zola-Morgan 1991), allows us to make explicit judgments as to the prior occurrence of stimuli based on previous encounters. The delayed matching-to-sample (DMS) and delayed nonmatching-to-sample (DNMS) tasks of recognition memory have been used extensively in animal models. Both tasks are initiated by requiring nonhuman animals to view the sample picture (or as occurred in many earlier studies, by presenting real objects briefly to animals in a Wisconsin General Test Apparatus), and then after a delay in which no stimuli are presented on the screen, typically two choice pictures are presented in which animals should select the stimulus seen earlier as sample (i.e., the “match”) in DMS task while animals should choose the picture different from the sample (i.e., the “nonmatch”) in DNMS task. By manipulating the length of delay between sample and test images, which is often varied within an experiment to test for delay-dependent deficits that might index memory processes; and/or by increasing the number of items to be remembered, the difficulty and cognitive demands on both tasks increase. As in both tasks, studied images will be represented in the test phase, animals may make use of the relatively familiarity of test items to differentiate studied and unstudied images. As familiarity is often considered as an automatic process (Jacoby 1991), animals may just rely on familiarity of studied images to solve recognition tests. Whereas animals can clearly judge relative familiarity of stimuli, the issue of whether animals can “recollect” discrimination per se has been a more contentious issue.

Dual-process signal detection (DPSD) theory suggests recognition memory may draw upon two processes, commonly referred to

as familiarity and recollection (Yonelinas 2001, 2002; Parks and Yonelinas 2007; Wixted 2007; Yonelinas and Parks 2007; Dede et al. 2014; Scalici et al. 2017). In recollection, if the item/object is judged as being encountered previously, the related contextual information about the item/object can be recalled with more details. For instance, recollection of an item/object may prompt retrieval of information about when and where one was when one encountered the item/object, what one was doing, thinking or feeling, as well as the surrounding environments. In contrast, familiarity is simply a feeling of the item/object has been encountered and generates a feeling of familiarity or knowing but without too much details (Yonelinas 2001, 2002; Parks and Yonelinas 2007; Wixted 2007; Yonelinas and Parks 2007; Dede et al. 2014; Scalici et al. 2017). Accordingly, recollection versus familiarity is often characterized as remembering versus knowing. In the assumptions of the DPSD model, recollection and familiarity are assumed to have different characteristics and contributions to receiver-operating-characteristics (ROC) functions: recollection is considered to be a threshold process which generates high confidence-level memories; while familiarity is thought to be a signal-detection process which generates memories of varying confidence levels (Yonelinas 2001, 2002). The confidence levels required for ROC plotting (across cumulative confidence levels) are measured by asking participants to state their confidence in the accuracy of their judgment that the stimulus has been viewed before (“old judgments”). In ROC curve fitting, two distinct indices can be extracted: a symmetrical and curvilinear component associated with familiarity, and an asymmetrical and linear component related to recollection. By analyzing the shape of ROC curves in the recognition memory task, differential contributions of recollection and

Corresponding author: buckley@psy.ox.ac.uk

Article is online at <http://www.learnmem.org/cgi/doi/10.1101/lm.051342.120>. Freely available online through the *Learning & Memory* Open Access option.

© 2020 Wu et al. This article, published in *Learning & Memory*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

familiarity to the recognition memory task can be separated, formally by deriving the recollection and familiarity indices that result from curve fitting (Yonelinas and Jacoby 1994; Yonelinas 2001, 2002; Eichenbaum et al. 2007; Parks and Yonelinas 2007; Yonelinas and Parks 2007; Yonelinas et al. 2010).

Recent animal studies have also investigated familiarity and recollection-like memory processes by analyzing the ROC curves based on DPSD model (Fortin et al. 2004; Sauvage et al. 2008; Guderian et al. 2011). A fundamental difficulty to overcome in animal models using ROC is that one cannot simply ask for a confidence judgment after the recognition memory judgment. Instead, in these tasks, the animal always earns the most rewards by answering correctly, but their willingness to indicate an item was experienced previously should be affected by an interaction of their confidence, the effort to respond, and the relative rewards for answering “yes” or “no”; for example, if the animal does not remember the test item, and the current payoff situation is that answering “no” has a small potential reward and answering “yes” has a large potential reward, then a confident animal should still answer “no” to the correct answer, but an unconfident animal should take a gamble on the higher reward and answer “yes.” In this way, the different tiers of reward/effort in animals act like the tiers of confidence judgments in humans. The first application of ROC curves to distinguish familiarity and recollection-like processes in animals was in the rodent model wherein Fortin et al. (2004) plotted ROC in rodents by varying the reward and effort required to make old versus new judgments so to allow responses to be assessed across different levels of “confidence.” Fortin et al. (2004) showed that these ROC looked like ROC in normal humans; moreover, in rodents with hippocampal lesions the changed shape of the ROC resembled the changed shape of ROC in human patients with hippocampal lesions (i.e., asymmetrical curvilinear ROC became curved symmetrical ROC). In patients this change may be interpreted as loss of recollection but spared familiarity; hence in rodents with hippocampal lesions, Fortin et al. (2004) showed that they too possibly lost a recollective-like component to their memory but their familiarity remained intact (Fortin et al. 2004). Subsequently, Sauvage et al. (2008) used an associative recognition paradigm and ROC to show recollection is reduced and familiarity increased after hippocampal lesions in rats (Sauvage et al. 2008). These issues are not without controversy and opposing views generated lively debate (Eichenbaum et al. 2005, 2008; Wixted and Squire 2008). Later, Guderian et al. (2011), by biasing macaques to respond to new and old stimuli by altering reward amounts they could obtain for their correct responses, further confirmed, using ROC, that recognition memory in animals draws upon separate familiarity and recollection-like processes in macaques (Guderian et al. 2011).

A key alternative approach to investigating the relative contributions of familiarity and recollection has investigated the time course of familiarity and recollection processes. Time courses of recollection and familiarity have also largely been investigated in human studies. For example, Yonelinas and Jacoby (1994) calculated the peak time of familiarity and recollection in a source memory task and showed that while the peak time of familiarity is faster, between 600 and 800 msec, the peak time of recollection is slower, between 800 and 1000 msec; this gives rise to the concept of faster familiarity and slower recollection (Yonelinas and Jacoby 1994). A faster response to retrieving familiar words (i.e., familiarity process) has likewise been found in comparison to retrieving details of words in a word list (i.e., recollection process) (Hintzman and Curran 1994). Some neurological studies have also confirmed an earlier process of familiarity compared to a later process of recollection in recognition memory. For instance, Curran (2000) found the components of event-related brain potentials (ERPs) related to familiarity appears earlier at 300–400 msec than other later parietal

components associated with recollection at 400–800 msec (Curran 2000).

Dosher and his colleagues applied a speed-accuracy trade-off paradigm to interpret the time course of these two processes, in which human participants must respond within a brief window following a go-signal, which is presented at various delays after the offset of the encoded sample items (Dosher 1984; Dosher and Rosedale 1991). In this paradigm, a plot of false alarm errors (i.e., errors to the lure items) against retrieval time shows a rise of errors across extremely short retrieval time that reaches a summit in the intermediate retrieval time then decreases in the longer retrieval times (Dosher 1984; Gronlund and Ratcliff 1989; Dosher and Rosedale 1991; Hintzman and Curran 1994; Rotello and Heit 2000; Rotello et al. 2000; Brainerd and Reyna 2005; Matzen et al. 2011). The peak time of these false alarm errors has been considered to relate to the onset of familiarity, and the suppression of false alarm errors along with the increases of accuracy during the longer moderate/intermediate retrieval times is considered to reflect the onset of recollection, which is conducive to minimizing false alarms (Brainerd and Reyna 2005). Taken together, there is significant consensus as to “fast familiarity” and “slow recollection” (FF/SR) across the human literature, although the methodology of speed-accuracy trade-off paradigm has been questioned in some other human studies (Brainerd et al. 2014, 2019).

More recently, Basile and Hampton (2013) have reported similar FF/SR profiles in NHPs. Their study is similar to the speed-accuracy trade-off paradigm used in human studies (Dosher 1984; Dosher and Rosedale 1991), wherein the time course of familiarity and recollection are investigated by plotting false alarm errors against retrieval time. Basile and Hampton (2013) introduced a modified recognition memory paradigm for NHPs in which animals were trained to select/touch a standard nonmatch button on the touch-screen if they wished to indicate that the test image did not match the studied image, but were trained to select/touch the test image itself if they wished to indicate that the test item was a match to the studied image. By this approach this paradigm allowed miss and false-alarm errors to be distinguished which better facilitates application of signal detection theory (i.e., equal-variance signal detection model, EVSD, in this case) to the NHP DMS task. By plotting false alarm errors and d' (i.e., discriminability/accuracy) against animals' response time, the authors found that recognition choices in the recognition memory task could be categorized in three different ways depending on response time: (i) short-latency errors induced by false alarms to familiar lures, deemed to indicate the process of familiarity; (ii) medium-latency responses are less likely to be affected by false alarms and are more accurate (i.e., d' reached the peak), suggesting the onset of recollection-like process that could correctly reject familiar lures; and (iii) long-latency and low accuracy responses which are deemed to be guesses (Basile and Hampton 2013).

The EVSD model and UVSD model (i.e., unequal-variance signal detection model, with its difference from EVSD models pertaining to whether the variance of old and new item distributions are equal or not) are based on single process theory, which suggests that the subjective experiences of recollection and familiarity are the expressions of recognition memory processes resulting from memory traces of different strength: recollection is associated with stronger memory traces, while familiarity is associated with weaker traces (Yonelinas 2001, 2002). The DPSD model in contrast is generated based on the dual-process theory, wherein performance on tests of recognition memory draws upon recollection (i.e., a threshold process which generates high confidence-level memories) and familiarity (i.e., a signal-detection process which generates memories of varying confidence-levels). The debate over whether recognition memory reflects a single-process, with assumptions derived from EVSD and UVSD models, or dual-

processes, with assumptions associated with DPSD models, are still ongoing (Wixted and Stretch 2004; Wixted 2007; Malmberg 2008; Koen et al. 2017). So far, any attempts to outright disconfirm either single-process models or dual-process models in normal human participants have failed (Malmberg 2008).

There have been criticisms of applying the DPSD model to differentiate familiarity and recollection-like processes in animals. First, the DPSD model assumes that recollection is a threshold process with high confidence judgments in which human subjects either recollect or do not recollect a stimulus; while familiarity is a signal-detection process with various confidence judgments. Whether animals can “recollect” discriminanda per se is still debated as discussed above, though at least some studies are favorable to the view that some animals exhibit behavior akin to recollection (Fortin et al. 2004; Sauvage et al. 2008; Guderian et al. 2011). Second, animal models have had to develop proxies for humans’ reported confidence judgments. As outlined above, rats were biased to respond to new and old odor by altering the amounts of rewards in different heights of test cups (Fortin et al. 2004; Sauvage et al. 2008); and macaques were biased to respond to new and old stimuli by altering different amounts of rewards they could obtain for their correct responses (Guderian et al. 2011). In the above three animal studies, it is assumed that animals have to be sufficiently confident to overcome any given level of bias, so the bias levels can be used to estimate their confidence. However, training animals to respond to correct new and old stimuli based on such bias levels so as to generate ROC is difficult and time-consuming and we are not aware of other studies that have achieved this, successfully, in whole cohorts of animals to-date, hence while an important and influential methodology it is unlikely to become a frequently used methodology in animal models.

Other criticisms on interpreting animal recognition memory using DPSD models are focused upon the model itself. Malmberg (2008) points out the lack of consistency in the measurements of recollection and familiarity derived from DPSD model even across different human studies (Malmberg 2008). A review from Yonelinas and Parks (2007) raised several issues related to this inconsistency, which should be carefully considered when applying DPSD to animal models. A first issue considers the average performance rate; either performing too well or too poorly will result in the ceiling and floor effects, which will lead to the difficulties in ROC curve fitting in the DPSD model and makes the interpretation of familiarity index based on slope of ROC less certain (Yonelinas and Parks 2007). Second, as with human participants, if the entire scale of confidence judgments is not utilized, then ROC curve fitting is less reliable and parameter estimates extracted from ROC curve will be less meaningful (Yonelinas and Parks 2007). Moreover, in animal models, if animals showed significant preferences to trials with certain biases values (i.e., trials with higher biases values representing larger rewards in studies of Fortin et al. 2004; Guderian et al. 2011) then the number of reliable data points to fit ROC curves may decrease. Additionally, curve fitting results are easily influenced when hits or false alarm rates on the extreme ends of the ROC approach 0.0 or 1.0 (Yonelinas and Parks 2007); negative values of R (recollection index) or F (familiarity index) may be reported but are meaningless considering the nature of recollection and familiarity process. All the above issues are challenges for DPSD model-based interpretation of human ROC and have arguably larger impact upon animal ROC studies of recognition memory.

The FF/SR paradigm on the other hand is relatively easy to implement in both humans and NHPs (Doshier 1984; Doshier and Rosedale 1991; Basile and Hampton 2013). In the NHP version of this paradigm reviewed earlier (Basile and Hampton 2013) only two images were used repeatedly, in each trial, so in each trial one was the studied image to remember while the other one was the lure (or vice versa). Basile and Hampton (2013) argued that

this was appropriate as larger sets are easier to remember (i.e., avoids too few errors for analysis) and also that large sets may be too easily discriminable on the basis of familiarity; however, this was not verified in that study. Certainly, small sets are likely to elicit greater proactive interference, but this is expected to be very limiting indeed for recollection and familiarity per se; the task essentially becomes a recency memory task that weakens the interpretation of Basile and Hampton’s data. Therefore, our study of FF/SR here expands upon Basile and Hampton (2013) by looking at both small sets and larger sets (trial-unique) of stimuli, to overcome the aforementioned limitation and more importantly, seek generalization of FF/SR effects across a range of set-sizes.

In this study, we adopted Basile and Hampton’s basic recognition memory paradigm, but further refined the paradigm to probe intraspecies comparisons across humans and NHP. First, in human subjects, we used the tasks to assess whether response time derived measurements of recognition errors could be used to differentiate fast familiarity and slow recollection processes in humans using similar paradigms to that we used in the NHP. Then, to probe the relationship between recognition error response time (FF/SR) derived and ROC derived estimates of familiarity and recollection indices in humans, we first sorted trials from slowest to fastest in order to assign trials to a series of response time bins (ranging from fast to slow), and then fitted human ROC data to the DPSD model and extracted familiarity and recollection indices independently across the different response time bins. To enable comparison of approaches we then plotted recognition errors (i.e., miss errors and false alarm errors) independently across similar response time bins to those used in the ROC approach. We compared the plots of recognition errors to plots of ROC derived familiarity/recollection indices along a common time course. Following Basile and Hampton (2013) we hypothesized a peak of false alarm errors associated with the short latency bins and we further hypothesized this “fast-familiarity” behavior would correspond with a peak of ROC-derived familiarity index; accordingly we also hypothesized that a peak in recognition discriminability/accuracy (d' prime, based on EVSD model as in Basile and Hampton (2013) would correspond in time-course with the peak of the ROC-derived recollection index in the medium-latency bins. If the aforementioned hypotheses were supported then, this would suggest that the time course analyses of patterns of recognition errors are phenomenologically related to the ROC-derived familiarity and recollection indices. Then, benefiting from using a similar recognition memory paradigm in both humans and macaque (Fig. 1), we were able to compare plots of recognition errors against response time in humans to those in the NHP on a similar task so to investigate the time course of familiarity and recollection across species; accordingly we hypothesized similar “fast familiarity” (peak of false alarm errors) and “slow recollection” (peak of discriminability/accuracy) profiles across response time in both species. If our first hypothesis about a relationship between recognition error-derived and ROC-derived familiarity/recollection indices holds then this significantly increases the validity of the response time approach derived from recognition errors in the NHP for probing their familiarity and recollective-like (i.e., dual-process) contributions to recognition, which is of clear benefit to future behavioral neuroscience research in animals as to the neural basis of recollection given the inherent difficulties associated with fully exploiting ROC approaches in animals as we reviewed earlier.

Results

Experiment I: human behavioral study

Performance data from all 27 participants were included in the analysis; each participant completed 180 trials (12 test trials in

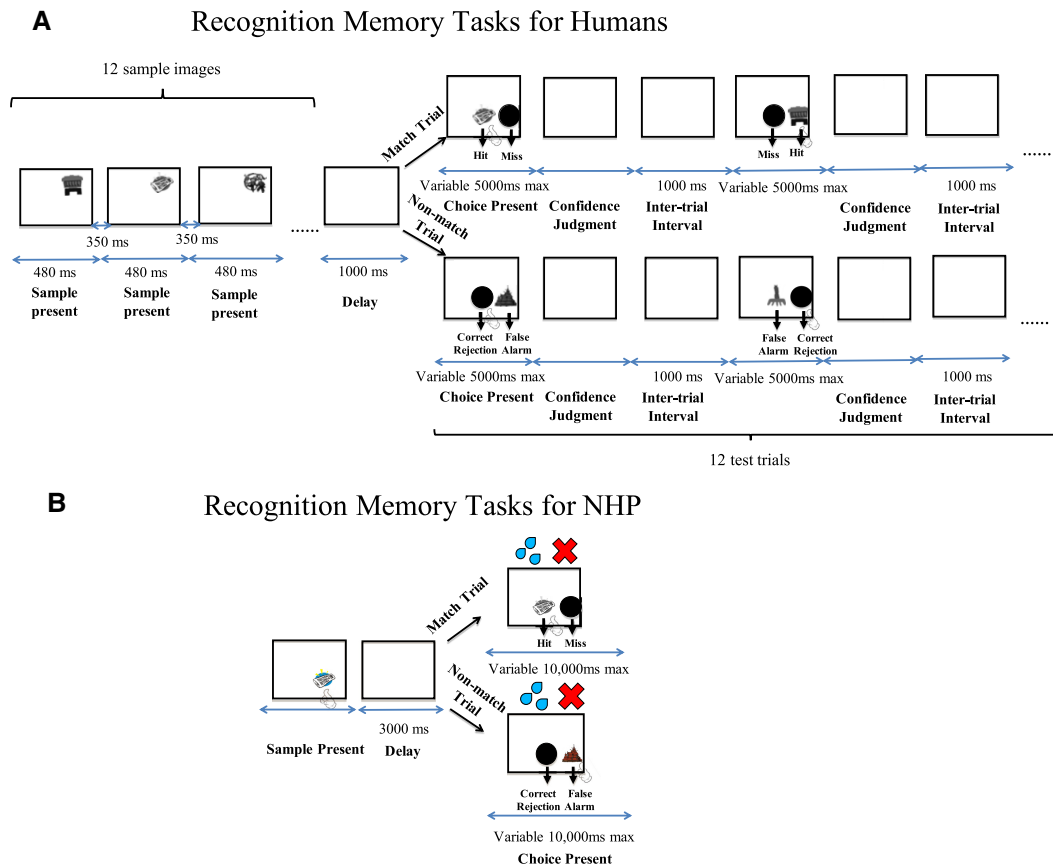


FIGURE 1. Diagram of the match/nonmatch recognition memory paradigms for human participants (*top panel, A*) and a NHP (*bottom panel, B*). (*A*) In the recognition memory tasks for humans, participants were instructed to remember 12 gray-scale sample images, then after a delay of 1000 msec, they were instructed to touch the test image if it matched the sample image; or selected the standard “nonmatch button” if not match (the maximal responding time for humans was 5000 msec); and then report their confidence judgments using hand gestures in each test trial; they completed blocks of 12 sample trials and then 12 test-trials in this way (15 blocks in total per session). (*B*) In the NHP recognition memory tasks the NHP viewed one sample image, and experienced a delay of 3000 msec, and then was trained to touch the test image if it matched or to select a standard nonmatch button if not match (the maximal responding time for animal was 10,000 msec); this trial structure repeated throughout the session with some samples and test-items being novel (trial-unique) and others familiar within (but not between) daily sessions.

each of 15 blocks). For analysis of recognition errors, for each participant, we first grouped their 180 trials into 10 response time (RT) bins (therefore, each bin contained 10% of each participant’s trials), and then calculated false alarm error rate, miss error rate, and d' (an index of discriminability/accuracy) independently for each RT bin.

To compare our results to the results from the study of Basile and Hampton, we also chose the EVSD model to interpret the discriminability/accuracy. D' (d') was calculated according to the output of an EVSD model wherein new-items and old-items are represented by two Gaussian distributions along the dimension of memory strength with equal variance. The d' measure depends both on the separation (difference in standardized means of new-item and old-item distributions) and the spread (standard deviation) of new-item and old-item distributions. As the EVSD model assumes the variance of new-item and old-item distributions are equal, d' is effectively the distance between new-item and old-item distributions (expressed as a z score with respect to its magnitude relative to the standard deviations of the distributions), reflecting sensitivity or discriminative ability to judging old-items from new-items. In our recognition error analysis, d' was calculated in each RT bin, as an indication of discriminability/accuracy (see Table 1 EVSD model for mean RTs represented in each bin).

Similar to plots of recognition errors and d' found in Basile and Hampton’s macaque study (Basile and Hampton 2013), we found that our human participants (Fig. 2) exhibited a slight U-shaped function: their false alarm error rate was significantly higher in our shortest latency RT bin than in the next two RT bins in which RT was lower (first vs. second bin: paired t -test, $t_{(26)} = 2.431$, $P = 0.022$; first vs. third bin: paired t -test, $t_{(26)} = 3.338$, $P = 0.003$), after the third RT bin it rose again (third vs. fourth bin: paired t -test, $t_{(26)} = -2.656$, $P = 0.013$; third vs. tenth bin: paired t -test, $t_{(26)} = -5.896$, $P < 0.001$). On the other hand their miss error rate increased (left to right on Fig. 2) from the fastest RT bins to the slowest RT bins (first vs. tenth bin: paired t -test, $t_{(26)} = -6.214$, $P < 0.001$). Discriminability/accuracy, measured as d' , demonstrated an increasing then decreasing trend across the 10 latency bins, whose peak-to-valley value was significantly larger than zero (second vs. tenth bin: paired t -test, $t_{(26)} = 6.958$, $P < 0.001$). It should be noted that the peak value of d' ($d' = 2.57$) appeared in the second RT bin (average time: 1225.0 msec). The peak time of false alarm error occurrence has been considered to correspond to the onset of familiarity (Basile and Hampton 2013), and the peak time of discriminability/accuracy (in parallel with a suppression of false alarm rate due to recollection/recollective-like memory processes) was considered to evidence the onset of recollection (Basile and Hampton 2013); accordingly, the patterns and time

TABLE 1. Mean response latencies for binned human data in Figure 2

Bin Number	1	2	3	4	5	6	7	8	9	10
EVSD model	1074.7±24.2	1225.0±30.7	1337.2±35.9	1443.4±41.8	1559.3±48.3	1685.0±54.0	1834.8±62.0	2050.8±72.4	2358.2±86.8	3089.1±118.5
DPSD model	1048.5±29.1	1209.4±38.1	1327.4±46.1	1442.9±56.6	1554.5±64.7	1677.0±73.1	1839.6±84.1	2067.0±100.3	2358.0±118.9	3058.7±170.4

Each bin contained 10% of the responses of each participant. For each response bin, we calculated the mean response latency (\pm SEM) in milliseconds for each human participant, and then averaged those mean latencies to produce a group mean latency for each bin.

profiles of recognition errors against response time in humans were similar to the ones found in the previous macaque study of Basile and Hampton (2013) which also showed evidence for fast-familiarity and slow-recollection (FF/SR) in a similar behavioral task. The confidence ratings were also plotted across the 10 RT bins, which showed a decreasing trend in confidence with longer RT; this is further consistent with participants guessing to a greater extent in longer RT bins.

To further investigate the latency of recollection and familiarity in human participants, we looked for correspondence between the aforementioned analyses and the ROC-derived approach for deriving familiarity and recollection indices. So we next applied the DPSD model to ROC data derived from each of the 10 RT bins separately. The DPSD model is more complex form of model containing a single threshold process of recollection and a pure signal-detection process of familiarity; below is the algorithm of

this model (Koen et al. 2017) we used:

$$p(\text{Lure}|\text{CL} \geq \text{CL}_i) = (1 - R_n) * \Phi(0, 1),$$

$$p(\text{Target}|\text{CL} \geq \text{CL}_i) = R_0 + (1 - R_0) * \Phi\left(\frac{d'_F - C_i}{\sigma_F}\right).$$

For all trials with confidence levels equal to or larger than the i th confidence rating bin (i.e., $\text{CL} \geq \text{CL}_i$), the DPSD model predicts the cumulative false alarm rate in lure-item trials and hit rate in target-item trials. Briefly, in the target and lure distributions (Φ is the cumulative normal distribution), R_0 is a target threshold parameter labeled recollection of old stimuli, and R_n is a lure threshold parameter labeled “recollection” of lures. As recollection is assumed to be a threshold process in the DPSD model, stimuli with a strength value above the R_0 threshold will be classified as a target. In the DPSD

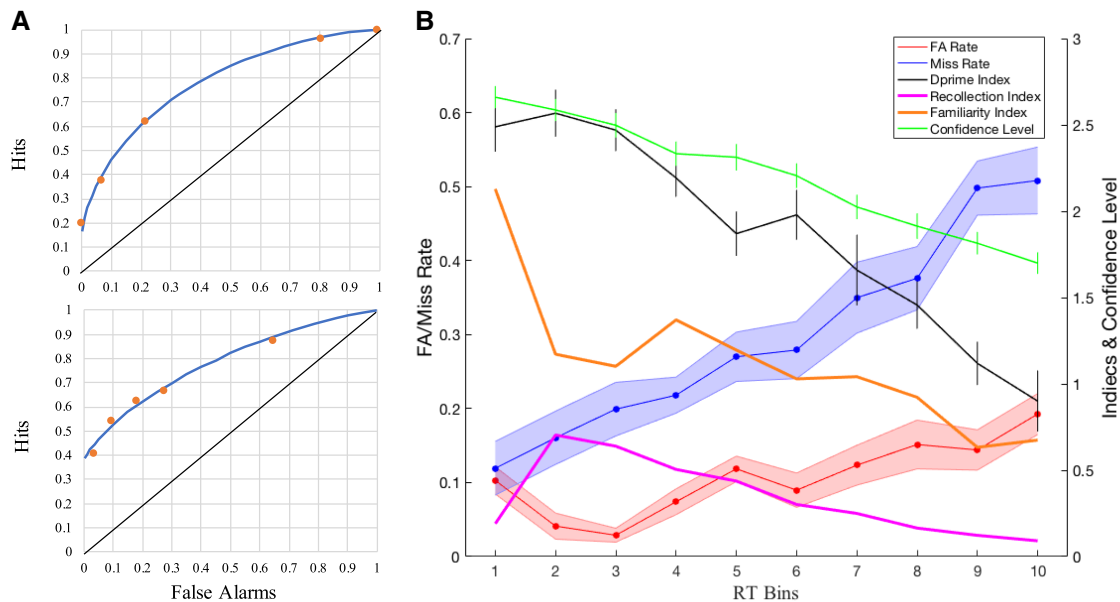


FIGURE 2. (A) Examples of ROCs fitted from the data of two human participants in Experiment 1. (*Top panel*) After curve fitting, recollection index of this participant is 0.153 and familiarity index is 0.925. (*Bottom panel*) After curve fitting, recollection index of that participant is 0.382 and familiarity index is 0.560; note that the higher recollection index in the *bottom* panel reflects the ROC curve shape being heightened on the *left* (corresponds to focus on higher confidence choices) and hence asymmetrical about the diagonal, whereas the lower recollection and higher familiarity indices in the *top* panel reflects the ROC curve shape being more symmetrical and curvilinear with respect to the diagonal; (B) Recognition errors, discriminability/accuracy d' , confidence level, and familiarity/recollection indices as a function of response time in humans. Each bin contains 10% of each participant's trials (ranging from bin 1 which contains the fastest 10% of trials to bin 10 which contains the slowest 10% of trials). Error bars are ± 1 SEM. False alarm errors (red solid line) and miss errors (blue solid line) were almost the same in the first RT bin; then miss errors increased, while false alarm errors initially decreased and then later increased across RT bins. Based on EVSD model, discriminability/accuracy (d' , black solid line) demonstrated an increasing then decreasing trend across RT bins and reached the summit in the second RT bin. Confidence level (green solid line) decreased along RT bins, consistent with participants guessing in the longest RT bin. Based on DPSD model, recollection index (magenta solid line) demonstrated an inverted U-shape against the 10 latency bins; while familiarity index (orange solid line) showed a decreasing trend, whose peak value appeared in the first RT bin.

model, as we don't care about the lure distribution, we make it a standard normal Gaussian function with a mean of 0 and a unit standard deviation ($R_0 = 0$). The other two parameters, d'_F and σ_F , are estimates of familiarity. In the DPSD model we used, familiarity is defined as a signal-detection process in which memory strength of target and lure items falls into Gaussian functions. In our DPSD model, we assume variances of target and lure items are equal ($\sigma_F = 1$), thus familiarity is defined as an EVSD model with its parameter d'_F . Based on the above assumptions, the algorithm of DPSD model becomes:

$$p(\text{Lure} | \text{CL} \geq \text{CL}_i) = \Phi(-C_i),$$

$$p(\text{Target} | \text{CL} \geq \text{CL}_i) = R_0 + (1 - R_0) * \Phi(d'_F - C_i).$$

When fitting ROC data to the DPSD model, sufficient trials are required to make ROC curve fitting sufficiently reliable (Yonelinas 2002; Yonelinas and Parks 2007). Therefore, for each human participant, we took the 180 trials divided into 10 RT bins from fastest to slowest and to ensure sufficient trials in each RT bin for curve fitting, trials in the corresponding RT bins of different participants were all combined. So, for example, the 10% fastest trials from each participant came to bin 1, and 10% slowest trials from each participant came to bin 10. Therefore, for analysis, there were 486 trials (i.e., 10% of total number of trials from all 27 participants) in each RT bin. Then for each RT bin independently, a ROC was plotted cumulatively for each confidence level according to the proportion of correct "old" judgments against proportion of incorrect "old" judgments. The recollection index (R_0) and familiarity index (d'_F) were extracted by fitting the model to data minimizing the squared difference between the observed and predicted data in each cumulative confidence rating bin. Yonelinas' research group provides an Excel worksheet which facilitates ROC curve fitting and index extraction (website link: <https://yonelinas.faculty.ucdavis.edu/roc-analysis/>). In this way, we calculated recollection and familiarity index across 10 RT bins and plotted the values on the same figure as the error-rates (and EVSD derived d') given both analysis approaches considered 10 RT bins (Fig. 2; see Table 1 DPSD model for average latencies represented by each RT bin).

The recollection index demonstrated an inverted U-shape across the 10 latency bins, whose peak value (recollection index = 0.70) was in the second RT bin (average time: 1209.4 msec), while the familiarity index showed a decreasing trend from a peak value (familiarity index = 2.13) in the first RT bin (average time: 1048.5 msec). The mean RTs for recollection/familiarity indices in each bin is shown in Table 1. It should be noted that the time-course of a significant switch from clearly high familiarity and clearly low recollection to somewhat lower familiarity values accompanied by somewhat higher recollection values appeared from the first to the second RT bin, thereafter both indices decreased as RT prolonged consistent with the longest latency RT bins being associated with increased rates of "guessing" (also consistent with confidence ratings reached their lowest point), and also consistent with our observed steady decline in discriminability/accuracy (see plot of d' on Fig. 2) as RT increased. This is also consistent with the aforementioned evidenced that miss rate steadily increases to 50% across higher RT bins. Notably, the average time of peak value of discriminability/accuracy or d' in EVSD model was in the second RT bin (average time: 1225.0 msec) and this corresponded to the RT bin (average time: 1209.4 msec) in which the peak recollection index occurred. Taken together our observations provide strong evidence that FF/SR processes exist in humans just as observed by Basile and Hampton in macaques, using broadly similar paradigms. Moreover, the interpretation/conclusions from the FF/SR approach appear consistent with the time profile of indices of

recollection and familiarity derived from the ROC approach, further validating the approach.

Experiment 2: NHP behavioral study

In this experiment, the animal had completed five sessions with more than 300 trials per session, and all of the data from these five sessions were put into the analyses. Similar to our above analyses on recognition errors in humans, we next analyzed macaque behavioral data across five sessions in a similar task; we likewise grouped trials into 10 RT bins, with each bin containing 10% of macaque's trials. In each RT bin, we calculated false alarm rate, miss rate and d' (discriminability/accuracy index based on EVSD model). This analysis was done separately for sample images that were novel (Fig. 3B) and those that were familiar (Fig. 3C) to the macaque (also see Table 2 for mean RTs represented in each bin for images as being novel and familiar to the monkey separately).

Consistent with previous results from the Basile and Hampton (2013) macaque study, which considered only highly familiar stimuli consisting of a repeating small set of just two stimuli (Fig. 3A), our plots of recognition errors for viewing larger numbers of familiar images (Fig. 3B) and also novel images (Fig. 3C). Both demonstrated a U-shape function in the false alarm error rate profile across the 10 RT bins whereas the miss error rate profile increased across the 10 RT bins. Discriminability/accuracy, measured as d' in the EVSD model, produced an inverted U-shape from the fastest RT bin to slowest RT bin, whose peak-to-valley value was larger than zero. Consistent with the main conclusion in Basile and Hampton's study and with the pattern of recognition errors in our human study (Experiment 1), we found a similar pattern of recognition errors and d' (discriminability/accuracy) in the macaque: at the short response latencies, errors were mostly false alarms consistent with performance being driven primarily by familiarity; at moderate response latencies, the false alarm rate dropped and d' was higher consistent with longer-latency recollective-like processes reducing the false alarm rate; and at the longest response latencies, both false alarm rate and miss rate increased and d' reached its lowest point consistent with monkeys guessing.

Although the familiarity level of sample pictures didn't affect the U-shape function of false alarm errors and the inverted U-shape function of d' as a function of RT, the RT bins in which the minimal values of false alarm errors and the RT bins in which the maximum discriminability/accuracy (d') were different for the different familiarity levels. To be specific, the first valley value of false alarm errors (FA rate = 0.1) and the peak value of d' ($d' = 2.66$) were in the second RT bin (average time 900.7 msec) for familiar pictures (Fig. 3B); while the first valley value of false alarm errors (FA rate = 0.1) and the peak value of d' ($d' = 2.56$) were in the fourth RT bin (average time 1070.9 msec) for novel pictures (Fig. 3C). The mean RTs for novel and familiar stimuli in each RT bin is shown in Table 2. As the peak value of d' (also corresponding to the first decrease in false alarm error rate) may indicate the start of recollection, it appears that novel images induce a switch from familiarity driven to recollection/recollective-like driven contributions to recognition memory later with respect to time-course, compared with familiar pictures.

Discussion

This study investigated the time course of familiarity and recollection/recollective-like processes contributing to recognition memory in both humans and NHP. A first key aim was to examine relationships between ROC derived indices of recollection and familiarity with recognition errors derived measures of "fast-familiarity" and "slow-recollection" (FF/SR) in humans. A second

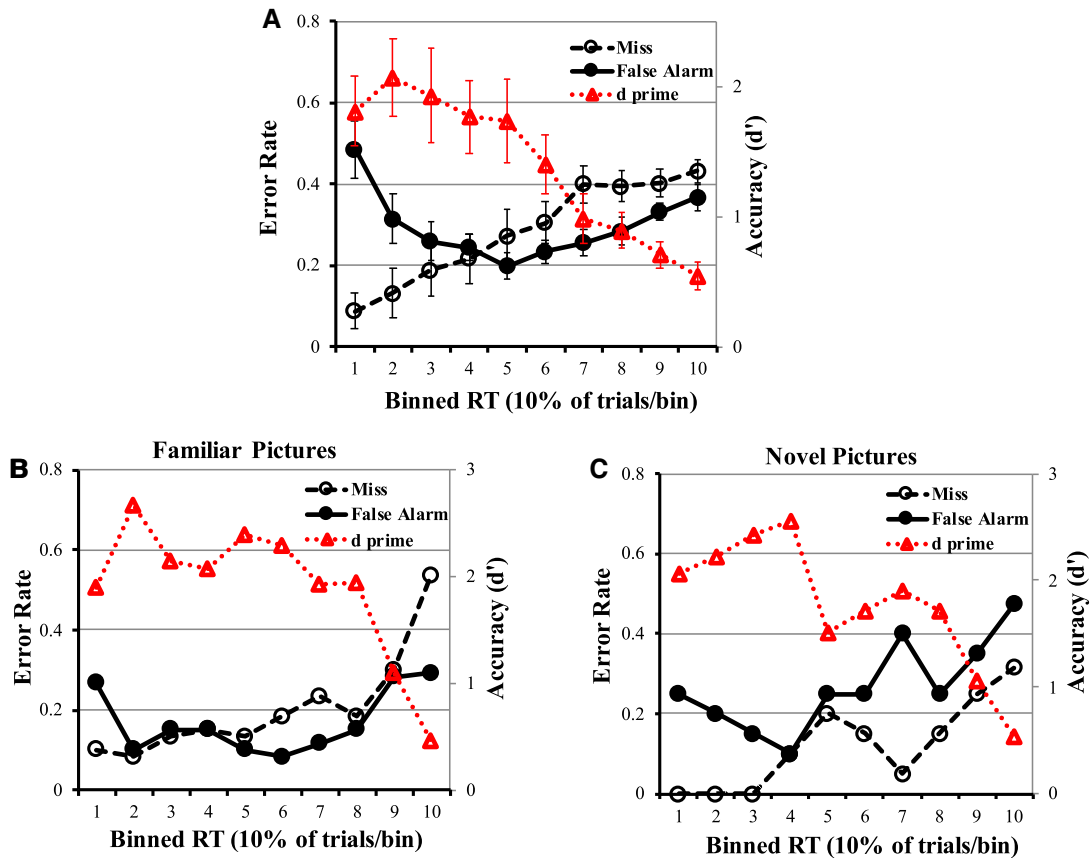


FIGURE 3. False alarm error rate (solid line with filled circles), miss error rate (dashed line with hollow circles), and discriminability/accuracy d' (dotted line with hollow triangles) as a function of response time bin in the macaque. Each bin contains 10% of each macaque's trials (ranging from bin 1 which contains the fastest 10% of trials to bin 10 which contains the slowest 10% of trials). Panels A–C depict recognition errors and discriminability/accuracy from: (A) data, replotted, from Basile and Hampton (2013); (B) from our macaque when viewing familiar pictures; and (C) from our macaque when viewing novel pictures. All the panels show that false alarm errors revealed a U-shape function, miss errors increased, and d' produced an inverted U-shape in relative incidence across the 10 RT latency bins in the NHP.

key aim was to compare FF/SR analyses in humans and NHP. In the current study, using a similar task structure across species, human participants and a macaque performed an object recognition memory task (with distinct match and nonmatch trials, see methods, to distinguish two types of errors, false alarms and misses). Our expectations were that: at the shortest response latencies, false alarm errors would be high, attributed to familiarity; intermediate-latency responses would be more accurate and associated with fewer false alarm errors, attributed to an onset of recollection or recollection-like processes reducing the false alarm errors; and at the longest response latencies, both false alarm errors and miss errors would increase in frequency and discriminability/accuracy would reach its lowest point suggestive of guessing to a greater degree. Indeed, in our macaque study, most of the short-latency

errors were false alarm errors, and at the intermediate response latencies, false alarm errors decreased along with an increase of d' , consistent with a slow recollective-like process to suppress false alarm errors and increase discriminability/accuracy. In humans, the short-latency errors were occupied by miss and false alarm errors equally; at intermediate response latencies, false alarm errors showed a decreasing trend along with an increasing of d' , which indicates a similar start point of recollection-driven memory. Thus, it is concluded that both humans and macaque can use “recollect to reject” to override false alarm errors driven by familiarity and their response profiles are broadly similar on the FF/SR paradigm.

Robust evidence in support of “fast familiarity and slow recollection” has been provided by many human recognition memory studies (Doshier 1984; Gronlund and Ratcliff 1989; Doshier and

TABLE 2. Mean response latencies for binned macaque data in Figure 3

Bin Number	1	2	3	4	5	6	7	8	9	10
Novel Image	781.8 ± 174.1	936.0 ± 84.4	1002.2 ± 103.2	1070.9 ± 111.2	1180.4 ± 129.3	1306.2 ± 162.8	1463.7 ± 173.4	1657.1 ± 211.4	1927.8 ± 255.5	3215.4 ± 1888.8
Familiar Image	714.8 ± 207.6	900.7 ± 51.0	973.4 ± 54.9	1048.3 ± 60.3	1133.5 ± 69.6	1245.9 ± 67.1	1384.8 ± 58.2	1563.8 ± 93.1	1941.5 ± 286.0	3689.9 ± 1688.4

Each bin contained 10% of the responses of the monkey. For each response bin, we calculated the mean response latency (\pm SEM) in milliseconds for images as being novel and familiar to the monkey.

Rosedale 1991; Hintzman and Curran 1994; Rotello and Heit 2000; Rotello et al. 2000; Brainerd and Reyna 2005; Matzen et al. 2011). For example, Gronlund and Ratcliff (1989) found little or no evidence for familiarity-based retrieval of studied items at extremely brief retrieval time (i.e., <200 ms), but as retrieval time lengthened (i.e., 300–500 msec), familiarity-based evidence accumulated accompanied by increasing false alarm errors, and as retrieval time became longer (i.e., more than 500 msec), recollection-based evidence was taken into account and accumulated to suppress false alarm errors (Gronlund and Ratcliff 1989; Brainerd and Reyna 2005). Consistent with these studies, our human and macaque results showed a U-shaped curve of false alarm errors against response time as detailed earlier (and see Figs. 2, 3); however we did not find the corresponding very low false alarm probability at extremely low response times as Gronlund and Ratcliff (1989) and other showed. One possible reason for this might be that our studies did not include such extreme short response time periods. Indeed, in the study of Gronlund and Ratcliff (1989) their very short (<600 ms) latencies included both sample-to-test item delay plus response time (also see the review by Brainerd and Reyna 2005) but in our macaque study, the delay between sample and test image was already 3000 msec and in our human study was 1000 msec. Instead, the U-shaped functions in our study, seen also in the macaque study of Basile and Hampton (2013), are consistent with the higher response times (wherein an inverted U-curve is seen) in the study of Gronlund and Ratcliff, in which false alarm errors reached highest point at short response latency and then decreased.

In this study we compared the response time profiles for fast-familiarity (FF) and slow-recollection (SR) in a macaque and humans; for each species we plotted recognition errors and discriminability/accuracy (indicated by d' in EVSD model) against response time (see Figs. 2, 3). The summit of discriminability/accuracy for humans appeared in the second RT bin, which indicates that a switch from relying primarily upon familiarity to relying upon recollection has already occurred by around 1225.0 msec (see Table 1 EVSD model for mean RTs represented in the second bin). In Figure 2, to compare ROC-derived indices of recollection and familiarity to the FF/SR interpretations we also plotted curves through the values obtained from deriving familiarity and recollection indices on the data in each bin from fitting ROC data to the DPSD model. The peak values of the ROC-derived familiarity index occurred much earlier in RT Bin 1 (corresponding to average peak time of familiarity index of 1048.5 msec) than the peak values of the ROC-derived recollection index in RT Bin 2 (corresponding to average peak time of recollection index of 1209.4 msec). The timing profiles of the ROC-derived indices of recollection and familiarity correspond with and corroborate the timing profiles of the FF/SR based interpretations further supporting and validating the FF/SR approach in humans. Moreover, the peak ROC-derived recollection index based on DPSD model fell in the second RT bin in which the discriminability/accuracy (d' , calculated based on recognition errors in EVSD model) was also in peak, suggesting recognition error distributions can be used as a proxy for estimating the onsets of recollection processes in humans. In our macaque study, the trends of recognition errors and discriminability/accuracy were similar to those in humans (compare Fig. 3B,C with Fig. 2). Accordingly, if the peak time of discriminability/accuracy can be used as a proxy to estimate the onset of recollection processes within the response time profile, then we estimate that an analogous switch from relying primarily upon familiarity to relying upon recollection in the macaque was at an average time of around 900.7 msec for familiar sample images (Fig. 3B), and with average time of 1070.9 msec for novel sample images.

The extent to which image repetition affects familiarity or recollection is an issue of debate. In some studies with humans repre-

sentation of studied items improves both familiarity and recollection; image repetition was found to increase false alarm errors if participants were under pressure to respond quickly but had no effects on false alarm errors if participants had more time to respond (Jacoby et al. 1998; Jacoby 1999; Jones and Jacoby 2001). Other studies using the remember-know (R/K) procedure (Tulving 1985) found that repetition of items enhanced recollection/remembering rather than familiarity/knowing (Mánytlá and Cornoldi 2002; Pitarque et al. 2015). In our macaque study, as mentioned above, the estimate switch time from familiarity to recollective-like processes occurred earlier for familiar than novel sample images. It appears that image repetition expedites the onset of recollection-like process so as to enhance the performance of recognition memory in the NHP. In a recent macaque electrophysiological and human ERP study, image familiarization sharpened the response dynamics of inferotemporal neurons of macaques and of human extrastriate visual cortex (Meyer et al. 2014). This finding illustrates that repeated viewing of an image modulated neural responses in both macaque and human visual areas, which may be conducive for efficient coding of stimuli. However, as only one animal was involved in this current study, and as we didn't carry out a similar investigation of image repetition effects on familiarity/recollection effect in humans, our conclusions as regards the difference between familiar and novel images should be appreciated as being limited and so we encourage future studies to follow up on investigations on whether this intriguing behavioral pattern can be replicated in both species.

Some authors have argued that the introduction of stimuli with preexisting familiarity into recognition-memory experiments might necessitate the use of "recollection for recognition" decisions (e.g., Bird 2017), such logic could affect interpretation of trials that contribute to the shortest RT bin in our NHP study (see Fig. 3B) if they are therefore deemed to draw greatly upon recollection or recollective-like processes contrary to our interpretation that they do not. We considered this issue from the outset by designing our study such that stimuli used in the preexperimental task-training stages were intentionally not used again in the formal experiment, so we can exclude consideration of long-term familiarity of the stimuli affecting macaque's recognition performance in our Experiment 2. However, short-term (i.e., within-session) familiarity remains an issue as some different or additional processes may support recognition of recently familiar stimuli compared to recognition of trial-unique stimuli (this is of course even more of an issue for Basile and Hampton's (2013) recognition memory task as they applied a very small set-size in their task wherein the same 2 images were repeatedly presented throughout the task). In another study, Basile and Hampton (2012) argued that when macaques perform recognition memory with familiar stimuli their memory is under "active cognitive control" based on their observations that recognition of familiar but not novel stimuli could be impaired by concurrent task demands imposed in the delay period. Human fMRI indicates that prefrontal cortex is more active during remembrance of familiar images (Stern et al. 2001) and it has also been argued (e.g., Bird 2017) that while extra-hippocampal areas in the medial temporal lobe can support recognition of unfamiliar items, the hippocampus itself is required to support recognition memory for previously experienced (i.e., familiar) items. However, even if it is presumed that there is greater involvement of hippocampus and/or prefrontal cortex in task performance, that does not necessitate involvement of "recollection" per se. An alternative view is that the additional involvement of context (temporal context in the case of recently familiar stimuli) might lead to the task becoming dependent upon recollection but the involvement of context does not imply recollection either; this is attested to by the substantial literature on episodic-like memory in animals that highlights that even contextually structured,

what-where-when, episodic-like memories that are also capable of being flexibly deployed still do not amount to memories that necessitate recollection per se (Griffiths et al. 1999; Clayton et al. 2001, 2003a,b). Another possibility to consider is that recognition of recently familiar stimuli, and the interference that is elicited, recruits recency memory processes distinct to item recognition per se (e.g., our own earlier study showed that NHPs with fornix transection were selectively impaired at judging relative recency of individual items in a short list but were not impaired at recognition of items in the list [Charles et al. 2004]), such that recognition memory for recently familiar stimuli may be based upon a recency signal as opposed to recollection. Moreover, there is good evidence from neuronal recordings in NHPs that signals for long-term stimulus familiarity exist in several extrahippocampal areas including TE, perirhinal, and entorhinal cortex (Xiang and Brown 1998; Hölscher et al. 2003) so hippocampal involvement may not be essential in support. While this literature as a whole leads us to expect different brain regions/networks to be recruited and/or to interact (e.g., frontal-temporal interactions) in recognition memory for recently familiar versus trial unique stimuli, it also persuades us to against conclusions that recollection would be required for performance in trials that fall in our shortest time-bins in our NHP study in Experiment 2. Empirical evidence for this study also appears to support this view as we note that the false alarm rate and d' values are very similar irrespective of whether stimuli were recently familiar or novel (Fig. 3: panels B vs. C). In our human experiment (Experiment 1), all the sample images are “trial-unique” so human participants’ performance there cannot be influenced by relative recency.

A small number of influential animal models (with rodents and more recently macaques) used the ROC approach to assess contributions of familiarity and recollection-like processes to recognition memory. These studies manipulated animals’ biases to respond “old” or “new” to stimuli by varying the relative amount of reward corresponding to correct old and correct new responses across various levels (Fortin et al. 2004; Sauvage et al. 2008; Guderian et al. 2011). As stated earlier, these types of manipulation are complex and time-consuming to train but most importantly the bias levels cannot be treated equivalently to humans’ confidence judgments, with the result that interpretation of ROC curves in animals has been questioned and hotly debated (Eichenbaum et al. 2008; Wixted and Squire 2008). We consider those animal models important and influential but at the same time we recognize both the practical and theoretical limitations. Indeed, this has been a major motivation for the current study which we consider successful in strengthening the interpretation of a proxy for differentiating between familiarity and recollection-like processes in animals, namely by investigating the time-course patterns of recognition errors. Our data further validates this FF/SR approach as providing valuable insight into these processes and hence future investigations of familiarity and recollection-like processes in animals, which are in need of advancing given the scarcity of animal models of recollection/recollective-like memory processes (see also Hampton 2001; Hampton et al. 2004; Hampton and Hampstead 2006; Basile et al. 2009; Paxton and Hampton 2009; Basile and Hampton 2011 for some other approaches to look at other elements of episodic memory in animals).

Finally, we turn to the general conceptual question of the use of recognition memory tasks, as opposed to study-test paradigms, to probe familiarity and recollection. While it is logical and appealing to think of the DMS recognition memory task (with individual stimuli, or short lists) as a key probe of working memory and to contrast that procedure to study-test paradigms (with longer lists) as key probes of long-term declarative memory, that does not reflect how these tests were differentially used and reported in the literature during the development of theories about the neural substrates of

declarative memory. A key question when considering the appropriateness of different paradigms for probing elements of declarative memory is to what extent they draw upon familiarity-based recognition and recollective-like memory; whether the task itself probes long-term, intermediate-term, short-term, or near-instantaneous memory may be a complementary and to some extent dissociable issue. Certainly, the object DMS task (and its nonmatching-to-sample variant, DNMS) has a long history in animal models (especially in NHP research); in the (predominantly) frontal lobe oriented literature the task is often referred to as a working memory task or delayed-response task. However, in the (predominantly) MTL oriented declarative memory literature it is more often just referred to as recognition memory or DMS/DNMS and one early influential theory proposed that hippocampal, parahippocampal, perirhinal, and entorhinal cortices together comprised a MTL memory system specialized for establishing long-term declarative memory (Squire and Zola-Morgan 1991). This theory was supported by highlighted observations as to the relative magnitude of impairments upon short-term DNMS (i.e., delays from seconds up to several minutes) following lesions of different extents within the MTL in macaques (e.g., see Fig. 4; Squire and Zola-Morgan 1991); subsequent factor analysis reported 2/3 of the test paradigms that provided “good measure of declarative memory” in monkeys were trial-unique DNMS tasks (Zola Morgan et al. 1994). In contrast, unambiguously long-term associative memory tasks such as concurrent discrimination learning (CDL) were instead regarded as nondeclarative memory habit-learning probes, especially since large lesions within the so-called “MTL declarative memory system” initially failed to significantly impair CDL (e.g., Malamut et al. 1984), a distinction subsequently eroded upon discovery that both DNMS and CDL were in fact found to be either impaired or unimpaired depending upon an entirely different factor, namely stimulus set-size, highlighting instead the greater importance of demands placed by task on stimulus identification and discrimination (Eacott et al. 1994; Buckley and Gaffan 1997, 2006). The status of DMS/DNMS as purported tests of declarative memory had nonetheless become pervasive in the MTL literature. Accordingly, it is perhaps unsurprising that when new paradigms such as confidence-report-derived ROCs and remember/know procedures that aimed to formally dissociate familiarity-based recognition and recollection also adopted some short-term memory paradigms (e.g., Yonelinas et al. 2002, 2005). As reviewed earlier, these new paradigms have been highly influential and have generally succeeded in dissociating familiarity-based recognition and recollection across humans, rodents, and macaques, supporting the notion that when investigating dissociations of those particular processes, the long-term, medium-term, or short-term nature of the task may not be of primary importance. A useful parallel may be drawn here with another prominent theory, namely relational memory theory of hippocampal function which was also originally conceived as a theory of long-term relational memory (Eichenbaum 2004), subsequently argued to also apply to short-term memory (Hannula et al. 2006), and now is arguably extendable to encompass hippocampal dependent simultaneous perceptual discrimination of spatial feature configurations/relations (Lee et al. 2005, 2006). In short, we argue that the short-term nature of the DMS tasks we have used here in both humans and NHP are suitable both for investigating similar mnemonic processes across primate species and for probing familiarity and recollection as these processes are not exclusive to long-term memory.

In summary, the conclusions from this study are that (i) “fast familiarity” and “slow recollection” provide comparable measures in both humans and macaque performing similar tasks; (ii) both species applied a similar strategy of “recollect to reject” to countermand false alarm errors based on recollection/recollective-like memory processes; and (iii) image familiarization expedites

switches from familiarity to recollection-like processes in macaque; and (iv) the time course of familiarity and recollection can be estimated relative to the time-to-peak value of discriminability/accuracy based on recognition errors, and this estimation method is validated by DPSD models used in humans.

Materials and Methods

Experiment 1: human behavioral study

Participants

Twenty seven participants (18 male, 9 female, age range 18–30 yr) took part in this study. Participants were fluent English speakers, right-handed, and had normal or corrected-to-normal vision. Prior to the study, all the participants provided written consent and went through safety screening check to make sure they had no history of previous or current neurological or psychiatric conditions and were not taking any psychoactive medication. All the participants received monetary compensation for their participation at a standard rate for volunteers in Oxford. This study was carried out with the approval of Medical Sciences Interdivisional Research Ethics Committee, University of Oxford.

Task stimuli and apparatus

The task was an object recognition memory task similar to the one used in the NHP, similarly programmed using Turbo Pascal (Borland), run under DOS on a desktop PC and presented on a 20.1 inch color touchscreen (TFT LCD TS200H GNR). The object images used in the task were clip-art images as in NHP study, but in order to increase difficulty (in light of a pilot study wherein performance was close to ceiling) the images were all converted to gray-scale and the contrast toned down in an attempt to make them harder to discriminate from each other. Additionally, the samples were presented in lists followed by lists of choice trials, again to reduce ceiling effects in the human version of the NHP task. Each image subtended 10° of visual angle in width and 10° in height to the participant sitting facing the screen. The sample image was always presented on the right top of the screen, positioned +12° horizontal and -12° vertical from the center of the screen. The test images were presented on the right bottom of the screen: one was positioned 0° horizontal and +5° vertical from the center of the screen; and the other one was positioned +23° horizontal and +5° vertical from the center of the screen. The background color to the screen was white.

Participants sat with their eyes a distance of 25 cm from the screen, wearing earplugs, resting their chins on a chin-rest and their foreheads on a head holder to stabilize their head position throughout the experiment. They were instructed to respond to items by touching them on the screen and gestured their confidence ratings using their right hands. For example, they indicate by raising fingers (1, 2, or 3) whether their opinion corresponding to their being somewhat confident (1), moderately confident (2), or absolutely confident (3) in their judgment as to whether they considered the test-item to be old (i.e., presented before in the preceding list as a sample) or new (not seen before in the preceding list as sample). None of the samples in this task were used in more than one list so all stimuli were trial-unique (and hence compared to the novel/trial-unique stimuli in the NHP task).

Behavioral task

Prior to the experiment, participants were given instructions on how to perform the task, and how to indicate their confidence judgments (see above), and then they were introduced to the behavioral task in a short training/practice session (comprised of one list of 12 samples and then 12 test-item trials) of the task to provide an opportunity to become familiarized with task procedure. We used lists in the human version to avoid ceiling effects as pilot investigations revealed the NHP task with single samples and single test-trials to be too easy for participants. The experimental session itself contained three subsessions, with each subsession

containing 15 blocks of trials. Participants took a 10–15 min break between each sub-session. Each block contained an encoding phase (i.e., a list of 12 sample images), and then a short delay (i.e., 1000 msec), and then a test phase (i.e., a list of 12 test-item trials). The task structure is depicted, for one block, in Figure 1 (top panel, A); during the encoding phase, 12 sample images were presented sequentially on the screen (individually for 480 msec), with inter-stimulus-interval as 350 msec. Participants were instructed to view and try to remember each sample image. After the sample phase (all 12 sample items) was completed, a blank screen was presented for 1 sec (delay), and then the test phase commenced (all 12 test trials). Test trials were either “match trials” or “nonmatch trials.” In each test trial, either an identical stimulus to one of the preceding 12 samples (i.e., “match trials”) was shown as a test image, or a novel and previously unseen image (i.e. “nonmatch trials”) was shown as a test image, and that test image appeared on the screen together with a black circle (the left/right position of the black circle and the test-trial image were randomized between trials). The maximal time allowed for a response for human participants was 5000 msec. The match and nonmatch trials were also counter balanced (six of each per test phase of 12 test-trials) and were put in a random order. Just as in the NHP version of the task (Experiment 2) participants were instructed to touch the test image if they thought it matched one of the 12 sample images, or touch the standard “nonmatch button” (i.e., the black circle) if they thought the test image was not a match. After responding to the test image, participants were instructed to rate their confidence as to whether the test item was new or old using a scale of 1–3 by making three different movements with their fingers, corresponding to somewhat, moderate, and absolute confidence. Participants were further instructed to try to use the entire range of confidence responses as best they could and not simply select the extremes of confidence as defaults.

Experiment 2: macaque behavioral study

Subject

One adult female rhesus macaque (*Macaca mulatta*, age 8 yr, weigh 10–13 kg) participated in this experiment. All animals in our laboratory are socially housed (or socially housed for as long as possible if later precluded, for example, by repeated fighting with cage-mates despite multiple regrouping attempts) and all are housed in enriched environments (e.g., swings and ropes and objects, all within large pens with multiple wooden ledges at many levels) with a 12 h light–dark cycle. The NHP always had ad libitum water access 7 d/wk. Most of its daily food ration of wet mash and fruit and nuts and other treats was delivered in the automated testing/lunch-box at the end of each behavioral session (this provided “jack-pot” motivation for quickly completing successful session performance; supplemented by trial-by-trial rewards for correct choices in the form of drops of smoothie delivered via a sipping tube) and this was supplemented with fruit and foraging mix in the home enclosure. All animal training and experimental procedures were performed in accordance with the guidelines of the UK Animals (Scientific Procedures) Act of 1986, licensed by the UK Home Office, and approved by Oxford’s Committee on Animal Care and Ethical Review.

Task stimuli and apparatus

The object recognition memory task was programmed using Turbo Pascal (Borland), run under DOS on a desktop PC. Visual stimuli used in the task were clip-art images in color, which were presented on a 20.1 inch color touch-sensitive screen (TFT LCD TS200H GNR). Those clip-art images used in the task were from a large pool of several thousand unique images. Each image subtended 5° of visual angle in width and 5° in height to the subject when presented on the screen. Both the sample image and test images were presented on the right bottom of the screen. The sample image was positioned +10° horizontal and +5° vertical from the center of the screen, with equal distance to the two test images. The test images were presented at the same horizontal level as the sample image:

one was positioned 0° horizontal and +5° vertical from the center of the screen; and the other one was positioned +20° horizontal and +5° vertical from the center of the screen. The background color to the screen was white. In each session, images were randomly chosen from the pool without replacement and were not reused on the other testing days.

The animal was seated in a primate chair (Rogue Research, Inc.) in front of the touch screen with its head-fixed and while it performed the recognition memory task in a magnetic-shielded, and partially sound-attended, testing-box. A window in the front of the chair provided its access to the touch-screen itself. The distance between the monkey and touch screen was fixed at 50 cm enabling the animal to touch the screen easily. An infrared camera was used to monitor the general status of the monkey in the box. A peristaltic pump device located on top of the box fed smoothie reward through a tube and to a spout positioned in the vicinity of the animal's mouth. Below the screen was also an automated lunch-box which contained the majority of the animal's daily meal (wet mash and fruits and nuts, etc.) and which opened immediately at the end of the task.

Behavioral task

In the NHP study, a similar short-term recognition memory task design has been exploited with the main difference being samples and test phases were not blocked. In each trial, animal viewed a sample image (shown in the right bottom of the screen) to remember and touched it to progress. Then after the 3000 msec delay period, the NHP was tested with two choice stimuli (one an object image and the other a black circle, left-right randomized between trials). The NHP was required to make a choice to the touchscreen to either the object test-image stimulus or to the black circle stimulus. The object stimulus was either the identical stimulus to the sample seen earlier in the trial or it was not identical to the sample. The animal was rewarded by delivery of 10 mL of smoothie for touching the test-item image if it matched the sample image (these we refer to as “match trials”), or it was rewarded for selecting the standard “nonmatch button” (i.e., the black circle) if the test-item was a nonmatch (these we refer to as “nonmatch trials”). The maximal time allowed for a response for animal was 10,000 msec. The next trial started after a 3000 msec intertrial interval after a correct response, or after a 10 sec “time-out” after any kind of error response. A schematic of the recognition memory task is illustrated in Figure 1 (bottom panel, B). Accordingly, on *match trials* the animal could either make a correct response (“hit”) or an incorrect response (“miss”) whereas on *nonmatch trials* the animal could either make a correct response (“correct rejection”) or an incorrect response (“false alarm”). The use of the black circle allowed hit/miss rate to be calculated independently of correct rejection/false-alarm rate; in this way the paradigm is similar to one previously used by Basile and Hampton (2013). A key element of task design in this paradigm for the macaque, enabled by the nonmatch “button” (i.e., black circle), is that miss errors and false-alarm errors may be distinguished independently in which their latencies can be used for estimating the time of familiarity and recollection in the macaques. We also varied the degree to which stimuli were either familiar or novel in the session (typically there were up to 80 pairs of novel stimuli that were trial-unique in each session, and six other familiar stimuli, grouped into three pairs, that repeated many times through the session). This allows us to investigate how the image familiarization (repetition of novel sample images) changes the relationship between recognition errors and response time.

Acknowledgments

This work was supported by Wellcome Trust Strategic Award Grant (Ref: WT101092MA) and MRC Project Grant (Ref: MR/K005480/1). We thank the team of expert animal technicians and veterinary staff and anaesthetists for their very high standards of animal care and husbandry throughout. We thank BM Basile for sharing the data for replotting Figure 3A.

References

- Basile BM, Hampton RR. 2011. Monkeys recall and reproduce simple shapes from memory. *Curr Biol* **21**: 774–778. doi:10.1016/j.cub.2011.03.044
- Basile BM, Hampton RR. 2012. Dissociation of active working memory and passive recognition in rhesus monkeys. *Cognition* **126**: 391–396. doi:10.1016/j.cognition.2012.10.012
- Basile BM, Hampton RR. 2013. Recognition errors suggest fast familiarity and slow recollection in rhesus monkeys. *Learn Mem* **20**: 431–437. doi:10.1101/lm.029223.112
- Basile BM, Hampton RR, Suomi SJ, Murray EA. 2009. An assessment of memory awareness in tufted capuchin monkeys (*Cebus apella*). *Anim Cogn* **12**: 169–180. doi:10.1007/s10071-008-0180-1
- Bird CM. 2017. The role of the hippocampus in recognition memory. *Cortex* **93**: 155–165. doi:10.1016/j.cortex.2017.05.016
- Brainerd CJ, Reyna VF. 2005. *The science of false memory*. Oxford University Press, New York.
- Brainerd CJ, Gomes CFA, Moran R. 2014. The two recollections. *Psychol Rev* **121**: 563–599. doi:10.1037/a0037668
- Brainerd CJ, Nakamura K, Lee W-FA. 2019. Recollection is fast and slow. *J Exp Psychol Learn Mem Cogn* **45**: 302–319. doi:10.1037/xlm0000588
- Buckley MJ, Gaffan D. 1997. Impairment of visual object-discrimination learning after perirhinal cortex ablation. *Behav Neurosci* **111**: 467–475. doi:10.1037/0735-7044.111.3.467
- Buckley MJ, Gaffan D. 2006. Perirhinal cortical contributions to object perception. *Trends Cogn Sci* **10**: 100–107. doi:10.1016/j.tics.2006.01.008
- Charles DP, Gaffan D, Buckley MJ. 2004. Impaired recency judgments and intact novelty judgments after fornix transection in monkeys. *J Neurosci* **24**: 2037–2044. doi:10.1523/JNEUROSCI.3796-03.2004
- Clayton NS, Griffiths DP, Emery NJ, Dickinson A. 2001. Elements of episodic-like memory in animals. *Philos Trans R Soc B Biol Sci* **356**: 1483–1491. doi:10.1098/rstb.2001.0947
- Clayton NS, Bussey TJ, Dickinson A. 2003a. Can animals recall the past and plan for the future? *Nat Rev Neurosci* **4**: 685–691. doi:10.1038/nrn1180
- Clayton NS, Bussey TJ, Emery NJ, Dickinson A, Suddendorf T, Busby J. 2003b. Prometheus to Proust: the case for behavioural criteria for “mental time travel” [1] (multiple letters). *Trends Cogn Sci* **7**: 436–437. doi:10.1016/j.tics.2003.08.003
- Curran T. 2000. Brain potentials of recollection and familiarity. *Mem Cognit* **28**: 923–938. doi:10.3758/BF03209340
- Dede AJO, Squire LR, Wixted JT. 2014. A novel approach to an old problem: analysis of systematic errors in two models of recognition memory. *Neuropsychologia* **52**: 51–56. doi:10.1016/j.neuropsychologia.2013.10.012
- Dosher BA. 1984. Discriminating preexperimental (semantic) from learned (episodic) associations: a speed-accuracy study. *Cogn Psychol* **16**: 519–555. doi:10.1016/0010-0285(84)90019-7
- Dosher BA, Rosedale G. 1991. Judgments of semantic and episodic relatedness: common time-course and failure of segregation. *J Mem Lang* **30**: 125–160. doi:10.1016/0749-596X(91)90001-Z
- Eacott MJ, Gaffan D, Murray EA. 1994. Preserved recognition memory for small sets, and impaired stimulus identification for large sets, following rhinal cortex ablations in monkeys. *Eur J Neurosci* **6**: 1466–1478. doi:10.1111/j.1460-9568.1994.tb01008.x
- Eichenbaum H. 2004. Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron* **44**: 109–120. doi:10.1016/j.neuron.2004.08.028
- Eichenbaum H, Fortin NJ, Ergorul C, Wright SP, Agster KL. 2005. Episodic recollection in animals: “If it walks like a duck and quacks like a duck...”. *Learn Motiv* **36**: 190–207. doi:10.1016/j.lmot.2005.02.006
- Eichenbaum H, Yonelinas AP, Ranganath C. 2007. The medial temporal lobe and recognition memory. *Annu Rev Neurosci* **30**: 123–152. doi:10.1146/annurev.neuro.30.051606.094328
- Eichenbaum H, Sauvage MM, Fortin NJ, Yonelinas AP. 2008. ROCs in rats? Response to wixted and squire. *Learn Mem* **15**: 691–693. doi:10.1101/lm.1133808
- Fortin NJ, Wright SP, Eichenbaum H. 2004. Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature* **431**: 188–191. doi:10.1038/nature02853
- Griffiths D, Dickinson A, Clayton N. 1999. Episodic memory: what can animals remember about their past? *Trends Cogn Sci* **3**: 74–80. doi:10.1016/S1364-6613(98)01272-8
- Gronlund SD, Ratcliff R. 1989. Time course of item and associative information: implications for global memory models. *J Exp Psychol Learn Mem Cogn* **15**: 846–858. doi:10.1037/0278-7393.15.5.846
- Guderian S, Brigham D, Mishkin M. 2011. Two processes support visual recognition memory in rhesus monkeys. *Proc Natl Acad Sci* **108**: 19425–19430. doi:10.1073/pnas.1117078108
- Hampton RR. 2001. Rhesus monkeys know when they remember. *Proc Natl Acad Sci* **98**: 5359–5362. doi:10.1073/pnas.071600998
- Hampton RR, Hampstead BM. 2006. Spontaneous behavior of a rhesus monkey (*Macaca mulatta*) during memory tests suggests memory

- awareness. *Behav Processes* **72**: 184–189. doi:10.1016/j.beproc.2006.01.007
- Hampton RR, Zivin A, Murray EA. 2004. Rhesus monkeys (*Macaca mulatta*) discriminate between knowing and not knowing and collect information as needed before acting. *Anim Cogn* **7**: 239–246. doi:10.1007/s10071-004-0215-1
- Hannula DE, Tranel D, Cohen NJ. 2006. The long and the short of it: relational memory impairments in amnesia, even at short lags. *J Neurosci* **26**: 8352–8359. doi:10.1523/JNEUROSCI.5222-05.2006
- Hintzman DL, Curran T. 1994. Retrieval dynamics of recognition and frequency judgments: evidence for separate processes of familiarity and recall. *J Mem Lang* **33**: 1–18. doi:10.1006/jmla.1994.1001
- Hölscher C, Rolls ET, Xiang J. 2003. Perirhinal cortex neuronal activity related to long-term familiarity memory in the macaque. *Eur J Neurosci* **18**: 2037–2046. doi:10.1046/j.1460-9568.2003.02903.x
- Jacoby LL. 1991. A process dissociation framework: separating automatic from intentional uses of memory. *J Mem Lang* **30**: 513–541. doi:10.1016/0749-596X(91)90025-F
- Jacoby LL. 1999. Ironic effects of repetition: measuring age-related differences in memory. *J Exp Psychol Learn Mem Cogn* **25**: 3–22. doi:10.1037/0278-7393.25.1.3
- Jacoby LL, Jones TC, Dolan PO. 1998. Two effects of repetition: support for a dual-process model of know judgments and exclusion errors. *Psychon Bull Rev* **5**: 705–709. doi:10.3758/BF03208849
- Jones TC, Jacoby LL. 2001. Feature and conjunction errors in recognition memory: evidence for dual-process theory. *J Mem Lang* **45**: 82–102. doi:10.1006/jmla.2000.2761
- Koen JD, Barrett FS, Harlow IM, Yonelinas AP. 2017. The ROC Toolbox: a toolbox for analyzing receiver-operating characteristics derived from confidence ratings. *Behav Res Methods* **49**: 1399–1406. doi:10.3758/s13428-016-0796-z
- Lee ACH, Buckley MJ, Pegman SJ, Spiers H, Scahill VL, Gaffan D, Bussey TJ, Davies RR, Kapur N, Hodges JR, et al. 2005. Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus* **15**: 782–797. doi:10.1002/hipo.20101
- Lee ACH, Buckley MJ, Gaffan D, Emery T, Hodges JR, Graham KS. 2006. Differentiating the roles of the hippocampus and perirhinal cortex in processes beyond long-term declarative memory: a double dissociation in dementia. *J Neurosci* **26**: 5198–5203. doi:10.1523/JNEUROSCI.3157-05.2006
- Malamut BL, Saunders RC, Mishkin M. 1984. Monkeys with combined amygdalo-hippocampal lesions succeed in object discrimination learning despite 24-hour intertrial intervals. *Behav Neurosci* **98**: 759–769. doi:10.1037/0735-7044.98.5.759
- Malmberg KJ. 2008. Recognition memory: a review of the critical findings and an integrated theory for relating them. *Cogn Psychol* **57**: 335–384. doi:10.1016/j.cogpsych.2008.02.004
- Mánytlá T, Cornoldi C. 2002. Remembering changes: repetition effects in face recollection. *Acta Psychol (Amst)* **109**: 95–105. doi:10.1016/S0001-6918(01)00054-3
- Matzen LE, Taylor EG, Benjamin AS. 2011. Contributions of familiarity and recollection rejection to recognition: evidence from the time course of false recognition for semantic and conjunction lures. *Memory* **19**: 1–16. doi:10.1080/09658211.2010.530271
- Meyer T, Walker C, Cho RY, Olson CR. 2014. Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. *Nat Neurosci* **17**: 1388–1394. doi:10.1038/nn.3794
- Parks CM, Yonelinas AP. 2007. Moving beyond pure signal-detection models: comment on wixted (2007). *Psychol Rev* **114**: 188–201. doi:10.1037/0033-295X.114.1.188
- Paxton R, Hampton RR. 2009. Tests of planning and the Bischof-Köhler hypothesis in rhesus monkeys (*Macaca mulatta*). *Behav Processes* **80**: 238–246. doi:10.1016/j.beproc.2008.12.016
- Pitarque A, Sales A, Meléndez JC, Algarabel S. 2015. Repetition increases false recollection in older people. *Scand J Psychol* **56**: 38–44. doi:10.1111/sjop.12168
- Rotello CM, Heit E. 2000. Associative recognition: a case of recall-to-reject processing. *Mem Cogn* **28**: 907–922. doi:10.3758/BF03209339
- Rotello CM, MacMillan NA, Van Tassel G. 2000. Recall-to-reject in recognition: evidence from ROC curves. *J Mem Lang* **43**: 67–88. doi:10.1006/jmla.1999.2701
- Sauvage MM, Fortin NJ, Owens CB, Yonelinas AP, Eichenbaum H. 2008. Recognition memory: opposite effects of hippocampal damage on recollection and familiarity. *Nat Neurosci* **11**: 16–18. doi:10.1038/nn2016
- Scalici F, Caltagirone C, Carlesimo GA. 2017. The contribution of different prefrontal cortex regions to recollection and familiarity: A review of fMRI data. *Neurosci Biobehav Rev* **83**: 240–251. doi:10.1016/j.neubiorev.2017.10.017
- Squire L, Zola-Morgan S. 1991. The medial temporal lobe memory system. *Science* **253**: 1380–1386. doi:10.1126/science.1896849
- Stern CE, Sherman SJ, Kirchoff BA, Hasselmo ME. 2001. Medial temporal and prefrontal contributions to working memory tasks with novel and familiar stimuli. *Hippocampus* **11**: 337–346. doi:10.1002/hipo.1048
- Tulving E. 1985. Memory and consciousness. *Can Psychol Can* **26**: 1–12. doi:10.1037/h0080017
- Wixted JT. 2007. Dual-process theory and signal-detection theory of recognition memory. *Psychol Rev* **114**: 152–176. doi:10.1037/0033-295X.114.1.152
- Wixted JT, Squire LR. 2008. Constructing receiver operating characteristics (ROCs) with experimental animals: cautionary notes. *Learn Mem* **15**: 687–690. doi:10.1101/lm.1077708
- Wixted JT, Stretch V. 2004. In defense of the signal detection interpretation of remember/know judgments. *Psychon Bull Rev* **11**: 616–641. doi:10.3758/BF03196616
- Xiang J-Z, Brown MW. 1998. Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology* **37**: 657–676. doi:10.1016/S0028-3908(98)00030-6
- Yonelinas AP. 2001. Components of episodic memory: the contribution of recollection and familiarity. *Philos Trans R Soc B Biol Sci* **356**: 1363–1374. doi:10.1098/rstb.2001.0939
- Yonelinas AP. 2002. The nature of recollection and familiarity: a review of 30 years of research. *J Mem Lang* **46**: 441–517. doi:10.1006/jmla.2002.2864
- Yonelinas AP, Jacoby LL. 1994. Dissociations of processes in recognition memory: effects of interference and of response speed. *Can J Exp Psychol* **48**: 516–535. doi:10.1037/1196-1961.48.4.516
- Yonelinas AP, Parks CM. 2007. Receiver operating characteristics (ROCs) in recognition memory: a review. *Psychol Bull* **133**: 800–832. doi:10.1037/0033-2909.133.5.800
- Yonelinas AP, Kroll NEA, Quamme JR, Lazzara MM, Sauvé MJ, Widaman KF, Knight RT. 2002. Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nat Neurosci* **5**: 1236–1241. doi:10.1038/nn961
- Yonelinas AP, Otten LJ, Shaw RN, Rugg MD. 2005. Separating the brain regions involved in recollection and familiarity in recognition memory. *J Neurosci* **25**: 3002–3008. doi:10.1523/JNEUROSCI.5295-04.2005
- Yonelinas AP, Aly M, Wang W, Koen JD. 2010. Recollection and familiarity: examining controversial assumptions and new directions. *Hippocampus* **20**: 1178–1194. doi:10.1002/hipo.20864
- Zola Morgan S, Squire LR, Ramus SJ. 1994. Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system. *Hippocampus* **4**: 483–495. doi:10.1002/hipo.450040410

Received January 7, 2020; accepted in revised form April 21, 2020.