

Published in final edited form as:

Q J Exp Psychol (Hove). 2009 June ; 62(6): 1198–1215. doi:10.1080/17470210802391599.

The contribution of familiarity to recognition memory is a function of test format when using similar foils

Ellen Migo¹, Daniela Montaldi¹, Kenneth A. Norman², Joel Quamme², and Andrew Mayes¹

¹School of Psychological Sciences, University of Manchester, UK

²Department of Psychology, Princeton University, USA

Abstract

Patient Y.R., who suffered hippocampal damage that disrupted recollection but not familiarity, was impaired on a yes/no (YN) object recognition memory test with similar foils. However, she was not impaired on a forced-choice corresponding (FCC) version of the test that paired targets with corresponding similar foils (Holdstock et al. 2002). This dissociation is explained by the Complementary Learning Systems (CLS) neural-network model (Norman & O'Reilly 2003) if recollection is impaired but familiarity is preserved. The CLS model also predicts that participants relying exclusively on familiarity should be impaired on forced-choice non-corresponding (FCNC) tests, where targets are presented with foils similar to other targets. The present study tests these predictions for all three test formats (YN, FCC, FCNC) in normal participants using two variants of the remember/know procedure. As predicted, performance using familiarity alone was significantly worse than standard recognition on the YN and FCNC tests, but not on the FCC test. Recollection in the form of recall-to-reject was the major process driving YN recognition. This adds support to the interpretation of patient data according to which, hippocampal damage causes a recollection deficit that leads to poor performance on the YN test relative to FCC.

Keywords

familiarity; recollection; recognition memory; hippocampus; remember/know

There is increasing consensus that recognition memory is supported by the underlying processes of recollection and familiarity (Yonelinas 2002; Mayes, Montaldi & Migo 2007; Wixted 2007). Recollection involves the recall of information associated with an item from a previous encounter, whereas familiarity is a feeling of memory in the absence of retrieval of this additional information. A variety of evidence converging from animal lesion work (e.g. Fortin, Wright & Eichenbaum 2004; Eacott & Easton 2007), fMRI (see Skinner & Fernandes 2007 for a review) and amnesic patients (e.g. Aggleton & Shaw 1996) suggests that the hippocampus is critical for recall, whereas familiarity depends on other medial temporal lobe cortices, notably the perirhinal cortex (see Brown & Aggleton 2001 for a review).

Particularly compelling evidence for the recollection/familiarity distinction comes from patients who have selective hippocampal damage and preserved familiarity (Vargha-Khadem, Gadian, Watkins, Connelly, Van Paesschen & Mishkin 1997; Holdstock, Mayes, Roberts, Cezayirli, Isaac, O'Reilly et al. 2002; Yonelinas, Kroll, Quamme, Lazzara, Sauve, Widaman et al. 2002; Bastin, Van der Linden, Charnallet, Denby, Montaldi, Roberts et al.

2004; Aggleton, Vann, Denby, Dix, Mayes, Roberts et al. 2005; Barbeau, Felician, Joubert, Sontheimer, Ceccaldi & Poncet 2005; Turriziani, Serra, Fadda, Caltagirone & Carlesimo 2008; but see also Manns, Hopkins, Reed, Kitchener & Squire 2003). Measuring familiarity using the remember/know (RK) procedure (Tulving 1985; Gardiner 1988) with patients can be problematic as they can struggle to understand the instructions (Baddeley, Vargha-Khadem & Mishkin 2001). Instead, selective recollection deficits have been investigated using event-related potential indices (Düzel, Vargha-Khadem, Heinze & Mishkin 2001), receiver operating characteristics (Aggleton et al. 2005; Cipolotti, Bird, Good, Macmanus, Rudge & Shallice 2006), the process dissociation procedure (PDP) (Bastin et al. 2004; see Jacoby 1991 for PDP method) and structural equation modelling (Quamme, Yonelinas, Widaman, Kroll & Sauve 2004).

Recent findings from Holdstock et al. (2002) suggest that *test format* can have a large effect on recognition performance in patients with selective recollection deficits. Patient Y.R. suffered damage limited to the hippocampus after a purported ischemic incident (see Mayes, Holdstock, Isaac, Montaldi, Grigor, Gummer et al. 2004 for brain volume data) and was left with impaired recollection (Mayes, Holdstock, Isaac, Hunkin & Roberts 2002), but intact familiarity (Holdstock et al. 2002). Holdstock et al. (2002) designed a recognition memory test where each target item was different, but had three very similar foils (see Figure 2). Y.R. was substantially impaired in a yes/no (YN) version of the test but performed normally on a forced-choice corresponding (FCC) version, where each target was presented with its similar foils (Holdstock et al. 2002; see Figure 3). The same dissociation has been seen in amnesic Mild Cognitive Impairment (MCI) patients (Westerberg, Paller, Weintraub, Mesulam, Holdstock, Mayes et al. 2006). Volume measures of amnesic MCI patients indicate a similar pathology to Y.R., in that they show hippocampal volume reductions (Bell-McGinty, Lopez, Meltzer, Scanlon, Whyte, DeKosky et al. 2005).

It has been suggested that forced-choice tests in general are easier to solve than YN tests using familiarity (i.e. even with dissimilar foils), as they allow *relative* familiarity judgements to be made between targets and foils (Parkin, Yeomans & Bindschaedler 1994). However, Y.R. could perform relatively normally on a variety of item recognition tests in both forced-choice *and* YN formats (Mayes et al. 2002). The Complementary Learning Systems (CLS) model of recognition memory (Norman & O'Reilly 2003) provides a more refined account of patient performance patterns in the object recognition memory test described above. The CLS model is a biologically constrained dual-process computational model of memory, composed of a neocortical component that supports familiarity processing and a hippocampal component that supports recollective processing (see Norman & O'Reilly 2003 for full details). In tests where targets and foils are very similar, the CLS model predicts that participants' ability to discriminate between them on a standard YN recognition test will depend on recollection of specific studied details. However, if participants are given a FCC test, the model predicts that both familiarity and recollection can support good discrimination performance.

Figure 1 illustrates why, according to the model, familiarity is more useful on FCC tests versus YN tests. When foils are similar to targets, the mean difference in familiarity between targets and their similar foils will be small, relative to the overall level of variability in the familiarity for different targets. Because of this, the familiarity distributions associated with targets and foils will overlap strongly. However, because there is strong *covariance* in the familiarity scores triggered by targets and their corresponding similar foils, targets should be reliably (slightly) more familiar than their corresponding foils (this point was originally made by Hintzman 1988; see also Norman & O'Reilly 2003)¹. In a YN test, each item appears alone as a test probe, and subjects must set a familiarity criterion in order to decide whether the item is old or new. Since the distributions overlap so closely, it is not possible

for a single criterion to separate the distributions, so YN discrimination should be very poor. In a FCC test, each target (e.g. C) is assessed relative to its own corresponding foils (e.g. C'), rather than the other foils (A', B', D', E'). Therefore participants can distinguish targets from foils in FCC by consistently accepting the most familiar item on each trial (provided that the resolution of familiarity is sufficiently fine). See Norman and O'Reilly (2003) for details on how familiarity and recollection success is predicted to change as target-foil similarity is manipulated and for other manipulations that can alter the relative success of familiarity and recollection in recognition memory tasks.

According to the CLS model, the patients' relatively spared performance on FCC tests is attributable to use of corresponding foils in the test, not the forced-choice test format *per se*. Norman & O'Reilly (2003) discuss how FCC performance can be compared to performance on a *forced-choice non-corresponding* (FCNC) test, where each target is presented with foils that correspond to other targets (Hall 1979; Tulving 1981). In this test format, participants cannot utilise the reliable familiarity difference between a target and corresponding foils because, as with YN recognition, these items are not presented simultaneously. Any given foil will be nearly as likely, on average, to be more familiar as it is to be less familiar than the target. Thus, as in YN, there is no principled decision rule that will allow targets to be reliably distinguished from foils based on familiarity alone. As such, the model predicts that familiarity-based discrimination on FCNC tests will be poor, and (consequently) participants will be forced to rely on recollection. The FCNC test may also be a better test to compare against the FCC format, since both tests have the same number of items and therefore the same list length. Another advantage of the FCNC test (relative to the YN test) is that forced-choice estimates of performance are relatively unaffected by response bias, whereas YN estimates of performance need to be corrected for response bias using assumptions that may not always be upheld (Macmillan & Creelman 2005; see discussion of this issue).

The prediction that familiarity-based discrimination should be better on FCC versus FCNC tests is not specific to the CLS model; it is also true for single-process 'global matching' models of recognition (e.g. MINERVA 2; Hintzman 1988). The key difference between the CLS model and global matching models is that, in models like MINERVA 2, familiarity is the *only* process driving recognition. As such, FCC performance should always be better than FCNC performance, whereas the CLS model's predictions depend on the extent to which participants are relying on recollection versus familiarity. When participants are relying exclusively on familiarity, FCC performance should be better than FCNC performance. However, if participants make use of recollection when performing the task, performance on FCC and FCNC tests can be similar. Whether performance on FCC and FCNC tests is *exactly* matched will depend on the levels of recollection used in each format. Since familiarity is sufficient to solve the FCC tests, participants may use less elaborate recall in that format, leading to higher overall recollection levels in the FCNC test. The CLS model allows for performance to be matched between the tests because it includes this contribution of recollection. Single process global matching models can never predict that FCC and FCNC performance would be the same.

Recently, it has been suggested that the YN/FCC dissociation could be attributable to the effects of study-test delay and interference rather than the ability of familiarity to solve different test formats (Bayley, Wixted, Hopkins & Squire 2008). In a group of patients with selective hippocampal damage, performance in both formats was impaired, but YN performance deteriorated during the test. Bayley et al. (2008) suggest that the poorer

¹Formally, the variance of the familiarity difference between a studied item and its corresponding foil can be expressed as follows: Let S be the familiarity of the studied item and let F be the familiarity of its corresponding foil. $Var(S-F) = Var(S) + Var(F) - 2 \times Cov(S, F)$. If $Cov(S, F)$ is large, $Var(S - F)$ will be small.

performance in the YN test in other patient work could be due to the YN test having more items than FCC and therefore, a longer average study-test delay. Additionally, in the YN test, foils related to a target appear on multiple trials and some targets appear several times (to encourage participants to treat each trial independently) which may cause a build-up of interference that selectively impairs patients' performance on later items.

The difference in performance in these patients relative to Y.R., is not unexpected, as this group has been shown to have impaired recollection *and* familiarity (Wais, Wixted, Hopkins & Squire 2006), unlike other patients where selective hippocampal lesions impaired recollection but left familiarity preserved (e.g. Holdstock et al. 2002; Bastin et al. 2004; Aggleton et al. 2005; Barbeau et al. 2005; Holdstock, Mayes, Gong, Roberts & Kapur 2005; Turriziani et al. 2008). No early-late deterioration in performance across the YN test has been found for Y.R. (Holdstock, personal communication) or using a larger group of 12 amnesic MCI patients (Westerberg, Florczak, Parrish, Weintraub, Mesulam, Mayes et al. 2006). Y.R. also performed normally on the FCC test after a 30 minute delay (Holdstock et al. 2002). Importantly, all the patient results (Holdstock et al. 2002; Westerberg, Florczak et al. 2006; Westerberg, Paller et al. 2006; Bayley et al. 2008) which compare sections of the test against each other are limited, since all participants received the same fixed order of items; this is usual in standardised tests of memory such as the Doors and People (Baddeley, Emslie & Nimmo-Smith 1994) and Warrington Recognition Memory tests (Warrington 1984). However, the procedure leaves the potential for item effects to confound split-half analyses, especially with small groups or case studies.

This paper aims to investigate normal healthy performance on these tests with similar foils that have previously been exclusively used with patients. The original patient data was interpreted within a dual-process framework, where Y.R.'s selective hippocampal damage and resultant recollection deficit led to her poor YN performance. In the FCC format, with concurrent target and foil presentation, her preserved familiarity enabled her to achieve relatively normal performance levels by making relative familiarity judgements. These experiments can clarify whether this explanation, an a priori prediction of the CLS model, can be supported or whether alternate theories based on interference and study-test delay are appropriate.

In Experiment 1, we used an instructional manipulation to vary normal participants' use of recollection versus familiarity. Participants received either standard recognition instructions or instructions to use familiarity only (Montaldi, Spencer, Roberts & Mayes 2006; Mayes et al. 2007). From the patient studies and the CLS model, we predicted that use of familiarity-only instructions should impair YN and FCNC, but not FCC performance relative to standard recognition. In this experiment, we used random item orders at study and test, thereby allowing us to address the issues raised by Bayley et al. (2008). If study-test delay and/or interference are having a greater impact on those who perform poorly in the YN format, regardless of memory process used, there should be a correlation between deterioration during the test and overall performance. The FCNC test also acts as a control for study-test delay as it has the same number of items as FCC, and so can be used to identify any interference effects. In Experiment 2, the YN test was investigated further using a justified-recollection/familiarity (justified-RF) procedure, in which participants were asked to describe which process they were using to make their recognition judgment.

Experiment 1

The goal of this experiment was to determine whether the CLS model's predictions about test format effects hold true in normal participants. It used a computerised version of the original object recognition memory task, rather than the cards used in patient work

(Holdstock et al. 2002; Westerberg, Florczak et al. 2006; Westerberg, Paller et al. 2006), with recognition and familiarity-only conditions. Although a longer delay was used and the FCNC test format was introduced, no other changes were made to the procedure to keep the method as consistent as possible with the patient work.

For the familiarity-only condition, participants were trained to understand the difference between recollection and familiarity, and then instructed to try to use only familiarity in the test phase. This procedure differs from the standard RK instructions (e.g. Rajaram 1993) by asking subjects to indicate only whether the item seemed familiar or not, as opposed to being asked whether it was 'remembered' or 'known' (Montaldi et al. 2006; see also Mayes et al. 2007). This modified-RK procedure uses the terms 'recollection' and 'familiarity' in the instructions (instead of 'remember' and 'know') because the terms 'remember' and 'know' have strong pre-existing associations from outside the experimental context. A recently published associative recognition study using this modified-RK procedure (Quamme, Yonelinas & Norman 2007) found that control participants given 'familiarity-only' instructions performed similarly to amnesic patients with selective recollection deficits. In an fMRI paradigm using this familiarity-only procedure, increases in subjective familiarity strength were associated with decreases in perirhinal but not hippocampal activity, whereas inadvertent recollections (which participants were asked to report) were associated with increased hippocampal activity (Montaldi et al. 2006). These findings show that participants in modified-RK experiments can selectively 'filter out' recollection and it also has been used to link familiarity and recollection reports to their neural substrates. We predict that performance using familiarity-only should be impaired relative to standard recognition for the FCNC and YN tests, but it should be relatively preserved for the FCC test.

When similar foils in the form of switched plurality item recognition tests have been investigated, ROC curves have been found to be relatively linear (Rotello, Macmillan & Van Tassel 2000). This is in keeping with the idea that recollection is important for good performance, and it also suggests that YN performance would be better modelled by a threshold based estimate (e.g. P_r or hit rate minus false alarm rate, Snodgrass & Corwin 1988) than an equal variance signal detection estimate such as d' (Macmillan & Creelman 2005). We do not expect that the threshold assumption will be entirely accurate; however, some assumption of this sort is needed to compute sensitivity on the YN test. As stated above, a key advantage of the FCNC test (relative to the YN test) is that it allows us to compute sensitivity without making (possibly incorrect) assumptions about the underlying nature of memory signal. When the patient data for Y.R. and the amnesic-MCI patients were reanalysed using P_r as the performance indicator, the pattern of results remained the same: YR was still impaired relative to controls on the YN test (z score=-4.48; Holdstock, Personal Communication), as were the amnesic-MCI patients relative to their controls ($p=0.011$; Westerberg, Personal Communication). Throughout this paper, YN results from both performance estimates (d' and P_r) are reported.

Method

Participants

Ninety-six students (average age 19.8 years, 12 males) were recruited from the Universities of Manchester and Liverpool. Some took part to obtain course credits while others were paid £3 for their time. Ethical approval was obtained from the School of Psychological Sciences Research Ethics Committee, University of Manchester. Sixteen participants took part in each test format-condition combination, where each participant completed a single test under standard recognition conditions or the familiarity-only condition.

Stimuli

The stimuli used were taken from Holdstock et al (2002). They are black and white silhouettes of everyday items and animals. Each picture has four different versions, one to be used as a target with three highly similar foils (see Figure 2 for examples). The high similarity between targets and foils was verified with a perceptual discrimination task performed by healthy controls. When deciding whether or not a target and foil were identical, participants made more errors and had longer reaction times for the stimuli in the object recognition memory task than for any other set of visual recognition memory stimuli administered to Y.R. (Holdstock et al. 2002). These stimuli therefore have more perceptually similar targets and foils than any other tests given to Y.R., which included the Doors and People Doors Test (Baddeley et al. 1994) and the Warrington Recognition Memory Test (Warrington 1984). The choice of target for each quartet of pictures was random. Two equivalent picture sets were used, each consisting of twelve items. These sets were identical to those used previously and half the participants used each picture set.

Procedure

Both conditions (recognition and familiarity-only) involved a study phase, a ten-minute delay and a recognition test phase (see Figure 3 for procedure summary). Participants were first shown four examples of picture quartets in order to encourage them to concentrate and to make them aware of the high similarity of foils. These example pictures were not tested in the recognition phase. One picture from each quartet was selected as a target with the three remaining pictures being used as similar foils in the test phase. The target pictures in a set of twelve were presented twice each, in a random order each time, for three seconds on each presentation. On the first presentation, participants were asked to make a natural or man-made judgement and on the second presentation, they were asked to study the details of the picture.

In the recognition condition, the ten-minute delay was filled with a mental arithmetic test; participants were not given any special instructions about what strategy to use on the recognition test. In the familiarity-only condition, participants were given clear written instructions at the start of the delay explaining the difference between familiarity and recollection. Participants then had to give an example of each to satisfy the experimenter that they understood. After this distinction was clear, the remainder of the ten-minute delay was filled with the mental arithmetic questions. At test, participants in the familiarity-only condition were asked to use familiarity and to refrain from trying to recall details. Participants were also instructed to report if they (inadvertently) recalled details on a particular trial (see Montaldi et al. 2006 for details).

For the FCC task the quartet for each picture was presented and for the FCNC task the original target was presented with randomly chosen foils from different targets. For these tasks there were therefore twelve trials in the test phase and the correct response was counterbalanced across the four choices. In the YN task, some targets were repeated to encourage participants not to be guided by their previous responses. Four targets were presented once, four targets were presented twice and four were presented three times, making a total of 60 trials. Only the response to the first presentation of a target was scored, ensuring that no single item had an undue impact on the overall performance measure. Instructions and stimuli were presented and responses recorded using the E-Prime software (Psychology Software Tools, Pittsburgh, PA)

Results

Data for each test format and condition are presented in Table 1. In the familiarity-only condition, trials on which recollection was reported were excluded from the analysis of familiarity performance (FCC: $M=2.38$ trials, $SD=1.63$; FCNC: $M=2.13$ trials, $SD=1.71$; YN: $M=4$ trials, $SD=4.02$). In all reported ANOVAs, the effect of test format compared YN, FCC and FCNC (or separate pairings of these formats as specified) and the effect of condition compared recognition with familiarity-only. YN performance has been estimated using P_r (see introduction) and also with d' . For the d' analysis, hit and false alarm rates were systematically corrected for any floor and ceiling effects as recommended by Snodgrass and Corwin (1988). FCC and FCNC results were converted to d' values using the table provided in Hacker and Ratcliff (1988). For the P_r results, to compare YN performance to the forced-choice formats, we converted the performance scores for each test format into a common metric of standard deviations below recognition mean (see Table 2), using percentage correct for FCC and FCNC.

An overall ANOVA based on the P_r data, including all three formats in both conditions showed a strong trend to significance for format, $F(2,90)=2.746$, $MSE=2.285$, $p=0.070$ (result with d' ; $F(2,90)=1.766$, $MSE=0.515$, $p=0.177$), a significant effect of condition, $F(1,90)=34.493$, $MSE=28.696$, $p<0.001$ (result with d' ; $F(1,90)=30.987$, $MSE=9.035$, $p<0.001$), and a trend to significance for the interaction, $F(2,90)=2.746$, $MSE=2.285$, $p=0.070$ (result with d' ; $F(2,90)=2.828$, $MSE=0.825$, $p=0.064$). Planned comparisons using P_r for YN and percentage correct for FCC and FCNC show that performance in the familiarity-only condition was impaired relative to recognition for YN testing, $t(30)=3.999$, $p<0.001$ (result with d' ; $t(30)=3.098$, $p=0.004$), and FCNC testing, $t(26.600)=4.762$, $p<0.001$ (result with d' ; $t(24.664)=4.979$, $p<0.001$), but not for FCC testing, $t(30)=1.478$, $p=0.150$ (result with d' ; $t(30)=1.650$, $p=0.109$).

Separate ANOVAs comparing the test formats in pairs were then carried out. The results of this analysis were consistent with the CLS model's predictions: Looking at just the FCC and FCNC tests, using proportion correct, performance on these tests was closely matched in the recognition condition, $t(30)=0.364$, $p=0.72$ (result with d' : $t(30)=0.267$, $p=0.792$). A two-way ANOVA of test format and condition showed a significant interaction, $F(1,60)=5.597$, $MSE=1795.04$, $p=0.021$ (result with d' ; $F(1,60)=5.018$, $MSE=1.584$, $p=0.029$), illustrating that FCNC performance was more severely impaired than FCC in the familiarity-only condition relative to recognition. There was also a significant main effect of condition, $F(1,60)=19.672$, $MSE=6309.59$, $p<0.001$ (result with d' ; $F(1,60)=21.429$, $MSE=6.767$, $p<0.001$), and a trend to significance for format, $F(1,60)=3.184$, $MSE=1021.31$, $p=0.079$ (result with d' ; $F(1,60)=3.223$, $MSE=1.018$, $p=0.078$). For the YN versus FCC comparison, there was a trend to an interaction between test format and condition, $F(1,60)=3.510$, $MSE=3.087$, $p=0.066$ (result with d' ; $F(1,60)=0.582$, $MSE=0.167$, $p=0.448$), but there was a non-significant interaction when comparing the YN and FCNC tests, $F(1,60)=0.037$, $MSE=0.031$, $p=0.848$ (result with d' ; $F(1,60)=2.657$, $MSE=0.722$, $p=0.108$)².

²We also ran a version of these comparisons where we did not exclude trials on which recollection was reported in the familiarity only condition. All of the results that were previously significant (at $p<0.05$) remained significant and all of the results that were previously non-significant (at $p<0.05$) remained non-significant, except for some of the d' results involving the YN format. The changed results were the overall interaction between test format and recognition vs. familiarity only (now significant, $p=0.039$), the planned comparison of YN performance in the recognition condition vs. the familiarity only condition (now non-significant, $p=0.084$), and the interaction between YN and FCC (now significant, $p=0.033$). The finding that familiarity-only instructions had less of an effect on YN performance with recollection trials included (vs. excluded) is expected, given our hypothesis that recollection should be more useful than familiarity on YN tasks.

In order to assess whether interference or study-test delay were contributing to our results, we first ran split-half analyses for all test conditions and formats (comparing performance in the first half of the test versus the second half of the test). First and second half scores were compared with t tests for all conditions and formats using with P_r and d' for YN and percentage correct and d' for FCC and FCNC. There were no significant effects, largest $t(15)=-1.041$, $p=0.314$. We looked to see whether weaker memories are associated with a greater decline in performance from the first to the second half. Overall performance for each participant was correlated against a split-half index (first half minus second half scores; where a larger positive score indicates a greater performance drop). With the YN test, no correlation was present using the data from both conditions together or separately with either P_r or d' , largest $r=-0.114$, $p=.675$. A parallel analysis with the FCNC results also found no significant correlations (using percentage accuracy as a performance indicator), largest $r=0.099$, $p=0.589$. This illustrates that there is no relationship between poorer memory, regardless of whether this is underlain by familiarity and/or recollection, and a greater deterioration in performance over the test.

Discussion

The data from this experiment match those found using the patient YR (Holdstock et al. 2002) and the amnesic-MCI patients (Westerberg, Florczak et al. 2006; Westerberg, Paller et al. 2006), where familiarity processing alone can lead to relatively normal performance on an FCC test, but not on a YN test. This is in line with predictions from the CLS model of recognition memory (Norman & O'Reilly 2003). The additional CLS model predictions about the FCNC test format were also supported, where familiarity performance was impaired relative to recognition. As predicted, there was an interaction between FCC and FCNC, where familiarity performance was much worse in the FCNC condition relative to recognition. It is also important that recognition performance in FCNC and FCC was closely matched. Although the CLS model predicts that this can happen when recollection is available, other global matching memory models predict a universal impairment in FCNC relative to FCC (e.g. Hintzman 1988). The dual-process model therefore provides the best account of these results. No split-half differences, interactions or correlations between split-half differences and performance were found in the data. There is, therefore, no evidence that study-test delay or interference effects can provide an alternative explanation for the results.

The fact that different performance estimates (i.e. d' and P_r) can slightly change the significance of the analyses is important. Neither estimate provides an entirely accurate measure of recognition performance; d' assumes an equal-variance signal detection model of recognition and P_r assumes a threshold model. We believe that the threshold model may be more appropriate for these data as linear recognition ROCs have been reported when using similar foils (Rotello et al. 2000), but as ROC data from an experiment designed for patients with such few trials is impossible to collect, this remains speculative. For this reason, results using both performance estimates have been included. To clarify further the role of recollection in the YN test, Experiment 2 was carried out to understand how healthy controls carry out the task.

Experiment 2

In Experiment 2, participants were given the YN test using the 'justified-RF' procedure. Unlike standard RK procedures, participants were asked to verbalise their decision-making process for each item using the terms familiarity and recollection and to explain, as appropriate, what they recollected at the time. One of the key advantages of the justified-RF procedure is that it allows not only quantification of the use of recollection to *accept* items

as old ('recall-to-accept'), but also separate quantification of the use of recollection to *reject* items as being new ('recall-to-reject'). Based on the CLS model's predictions, we expected that performance in this paradigm would be primarily driven by recollection in the form of recall-to-reject processing.

Participants were able to report their use of recollection and familiarity in their decision making process in real time because this task is difficult and a recognition decision takes a number of seconds. The real time aspect is an advantage of the method as it allows participants to be prompted for extra information when required, to ensure that responses are correctly categorised. One key concern about the method is that asking for responses to be verbalised imposes demand characteristics on participants that might encourage them to make increased use of recollection, leading to changes in overall performance (see Diana, Reder, Arndt & Park 2006 for a discussion of demand characteristics of the RK procedure in general). To address this possibility, we compared overall performance levels and familiarity performance in Experiment 1 and Experiment 2; as discussed below, we found that performance levels were not significantly different across the two experiments. This suggests that the demand characteristics of the justified-RF procedure did not substantially alter subjects' strategies, and (as such) the results of Experiment 2 can be used to shed light on the processes used during Experiment 1.

Participants

Sixteen students (average age 18.1 years, 4 males) were recruited from the Universities of Manchester and Liverpool. Some took part to obtain course credits while others were paid £3 for their time. Ethical approval was obtained from the School of Psychological Sciences Research Ethics Committee, University of Manchester. Four participants who had ceiling levels of recollection (with 8 or less trials on which familiarity was reported) were replaced.

Stimuli and Procedure

The stimuli and overall procedure were identical to the YN condition from Experiment 1, except for the procedure used at test. All participants received training on the difference between recollection and familiarity at the start of the study-test delay period. On each test trial, participants had to verbally describe how their responding on that trial was guided by recollection and familiarity. No participants reported using a mixture of both processes to solve the task. Since the decision making process was verbalised in real time there were occasions when participants reported finding an item familiar and then reported a subsequent recollection. The high-target foil similarity means that almost all items will feel familiar. Therefore a feeling of familiarity is often not sufficient for participants to make a decision, unlike in more standard YN tests with more different foils. If participants had not reported a final recognition decision based on familiarity before the recollection occurred, the trial was scored as a recollection trial. This is because it is not possible to judge what decision would have been made using only familiarity. If recollection occurred, participants were asked to say what it was they recollected. Before starting the test phase, participants were given examples of how responses could be justified using recollection (using a specific detail from encoding), familiarity (using a general non-specific feeling of memory), or guessing (no feeling of why a decision was being made). Participants were discouraged from guessing but instructed to state when they were making a guess without any influence from feelings of familiarity and recollection (however, only two guess responses were reported which were excluded from the totals for those participants). The entire test phase was recorded on audio cassette tape or digital voice recorder.

Results

Table 3 shows the distribution of the responses by process for all participants as hits, misses, false alarms and correct rejections. Familiarity rates have been calculated assuming stochastic independence, since on trials where recollection is reported, you cannot also assess whether familiarity was present at a level that would have led to the same decision (Yonelinas & Jacoby 1995; Yonelinas et al. 2002; see Appendix for equations for calculating familiarity performance). To identify whether the justified-RF procedure changed performance levels relative to Experiment 1, independent samples *t* tests were used to compare performance (recognition and familiarity) from Experiments 1 and 2 (a summary of performance estimates from both experiments is shown in Table 4). Using both P_r and d' , no significant differences were found, largest $t(30)=-1.152$, $p=0.258$ for recognition levels measured with d' . There is no evidence that the use of the justified-RF procedure changed overall recognition or the success of familiarity in isolation, so concerns that the justified-RF procedure would alter participants' performance do not appear to be warranted.

The information on correct rejections and misses that was collected allowed us to calculate the contribution of recall-to-reject processing. Recall-to-reject was indexed by correct rejections minus misses (i.e., trials on which recollection correctly or incorrectly identified items as new) and recall-to-accept was indexed by hits minus false alarms (i.e. trials on which recollection correctly or incorrectly identified trials as old). Recall-to-reject performance was significantly better than recall-to-accept, $t(15)=3.451$, $p=0.004$.

To investigate the nature of the reported recollection false alarms (an average of 3.75 trials per participant), performance on these trials was categorised by cause from the recorded responses from participants: 65.0% were 'correct' recollections that were not diagnostic, 26.7% were incorrect recollections and 8.3% could not be classified (e.g. noise loss from tape). Of the 16 participants, nine did not have any 'incorrect' recollections.

Discussion

The use of the justified-RF procedure carried a risk that by asking participants to verbalise their decision process, it would have encouraged them to recollect more information than the participants in Experiment 1. This could be a particular problem in this test where recollection is critical for task success. The lack of a difference in overall recognition and in familiarity performance between Experiments 1 and 2 shows that the use of this method has not changed the way participants complete the task as far as overall success is concerned. The method allowed information to be collected on how participants reject items as old, which enabled us to index recall-to-reject performance. This is particularly relevant to tests like the similar foil YN task used here because recall-to-reject is likely to be a major contributor to performance. The results indicate that overall performance is largely driven by recall-to-reject, as performance using recall-to-reject was significantly higher than recall-to-accept.

Recollection, although more useful than familiarity, is not a guarantee of success in this task, as indicated by the presence of false alarms and misses. It is important to address the level of false alarms due to recollection (see Table 3), which does not fit comfortably with the idea that recollection is a high threshold process that is only triggered by studied items (e.g. Yonelinas 1994). Based on participants' justifications, it appears that for the majority of the time (65%), participants are recollecting actual studied details when they make recollection false alarms and this result is peculiar to this task with such target-foil similarity. False alarms occur because participants think that a particular (correctly) recollected detail is diagnostic, when in fact it is not. For example, in the house picture quartet in Figure 2, if

participants study the first house and then are tested with the fourth house, they might (correctly) recollect having studied a house with the door in the centre, and then (incorrectly) conclude based on this information that the item was studied. Given the very high level of target-foil similarity in this study, calling an item 'old' based on recollection of particular details is a very risky strategy, because those details were often also be present in foils. Using a recall-to-reject strategy is much safer than using a recall-to-accept strategy: Although matching recollection is rarely diagnostic of an item having been studied, mismatching recollection (e.g., recollecting that the chimney was on the left in the studied house, when the chimney is on the right in the presented test item) is a reliable indicator that the test stimulus is a foil (for a discussion of the importance of diagnostic recollection see Dobbins, Kroll, Yonelinas & Liu 1998).

The results of Experiment 2 also help to explain why familiarity-only instructions in Experiment 1 boosted false alarms and hits. In Experiment 2, we found that that recollection was primarily used to reject items, rather than accept them. We also found that recollection was used both to correctly reject lures and (to a lesser extent) to incorrectly reject studied items. These results imply that removing recollection should lead to a large increase in false alarms and a smaller increase in hits; this was exactly the pattern that we observed in Experiment 1.

General Discussion

Taken together, these experiments clearly demonstrate that recollection is required for success in YN and FCNC tests with high target-foil similarity. In contrast, the FCC task can be largely solved using familiarity alone. These results converge with prior results showing impaired YN performance but relatively spared FCC performance in a patient with selective hippocampal damage and recollection impairments, but with intact familiarity (Holdstock et al. 2002) and with groups of amnesic MCI patients (Westerberg, Florczak et al. 2006; Westerberg, Paller et al. 2006). This direct testing of the use of familiarity in healthy controls supports the interpretation of the patient data related to the use of preserved familiarity and the predictions of the CLS model.

In Experiment 1, separate groups of participants completed recognition memory tests (using very similar foils) in YN, FCC and FCNC formats in one of two conditions: standard recognition versus familiarity-only. As predicted by the CLS model, performance using familiarity alone was poorer than standard recognition in the YN and FCNC tests but not the FCC test. In Experiment 2, the YN test was investigated further using a justified-RF procedure that quantified how often recollection was used to accept items as old or reject items as new. The change in procedure of making participants report the basis of their recognition decision did not change the recognition and familiarity performance estimates for the test compared with those from Experiment 1, showing that further results are not simply an artefact of the method. The results showed that overall recognition sensitivity was mainly driven by recollection, in the form of recall-to-reject processing. Recall-to-accept was less reliable at solving the task. This result could not have been obtained using a standard RK procedure since this does not measure reasons behind any rejections of items as new (i.e. misses and correct rejections) and thus cannot capture the contribution of recall-to-reject.

The FCNC/FCC dissociation shown in Experiment 1 was more reliable than the YN/FCC dissociation. As a forced-choice test, the FCNC test does not have the same issues as the YN test over the choice of performance estimate used (i.e. P_r versus d'). The FCNC test is also better matched to the FCC test in that they are both forced-choice tests (here four-choice forced-choice tests) and both use the same number of items. Given these two factors, we

believe that future patient work exploring test format effects using targets and similar foils should focus on FCNC/FCC differences, rather than YN/FCC differences.

Our split-half analyses in Experiment 1 allow us to address whether other factors are contributing to the observed test format dissociations in patients. The study-test delay and interference hypotheses (Bayley et al. 2008) both predict that performance for those with weaker memories should worsen during the YN and FCNC tests. These hypotheses apply irrespective of whether these weaker memories depend on recollection and/or familiarity. Study-test delay effects would result in a split-half effect for all test formats, with the greatest effect in the YN test as it is longer. Interference effects would result in a split-half difference for YN and FCNC tests. The lack of *any* differences between first- and second-half performance in Experiment 1, even in the FCNC and YN familiarity-only conditions where performance was very poor, suggests that study-test delay and interference effects are not major factors. The correlations between overall memory performance (including both conditions) and split-half changes were also all non-significant, suggesting that those who perform badly on the test are not more susceptible to delay or interference effects.

A recent fMRI study using highly perceptually similar targets and foils in a YN test showed a dissociation in hippocampal and perirhinal activity (Danckert, Gati, Menon & Köhler 2007). Hippocampal activity discriminated between correct and incorrect trials whereas perirhinal activity reflected subjective opinion (i.e. hit and false alarm activity was matched, as was the activity for misses and correct rejections). The interpretation of these data must be limited, since recollection and familiarity reports were not collected and the targets were all similar. However, taking hippocampal activity as a reflection of recollection and perirhinal activity as a reflection of familiarity, the data are consistent with our interpretation of the patient and healthy participant results; familiarity/perirhinal processing cannot support good performance in a YN test with high target-foil similarity and recollection/hippocampal processing is important for good recognition levels. Other studies have used the RK procedure to explore the contributions of recollection and familiarity to forced-choice and YN tests (e.g. Khoe, Kroll, Yonelinas, Dobbins & Knight 2000; Kroll, Yonelinas, Dobbins & Frederick 2002; Bastin & Van der Linden 2003; Cook, Marsh & Hicks 2005). However, none of these experiments used test materials constructed in the same manner as the object recognition memory test here. When foils are not highly similar to studied items, the CLS model predicts that recollection and familiarity can both support good performance, regardless of test format.

It is worth noting that the CLS model only predicts test format dissociations (i.e., impaired FCNC and YN, spared FCC) in patients with selective recollection deficits. If familiarity processing is also damaged, the model predicts that patients should show impaired performance on both test formats (given that they both use stimuli with high target-foil similarity). Thus far, patient results appear to be consistent with these predictions. Patients who (in other tests) show evidence for impaired recollection and spared familiarity, show impaired YN performance and relatively spared FCC performance (e.g. Holdstock et al. 2002). Likewise, patients who (in other tests) show evidence of impaired recollection *and* familiarity show impairment on both the YN and FCC version of this test (Bayley et al. 2008). If this pattern of results continues to hold, it raises the possibility that one could diagnose selective recollection deficits by looking for test format dissociations (spared FCC, impaired FCNC and/or YN).

Finally, our FCC and FCNC results pose a challenge for single-process global-matching models of memory (e.g. Hintzman 1988). As discussed in the Introduction, global-matching models base recognition judgments in their entirety on familiarity and therefore they predict that FCC performance should always be higher than FCNC performance. In Experiment 1,

we found an FCC advantage over FCNC in the familiarity-only condition. However, in the standard recognition condition, where recollection could also contribute to performance, FCC and FCNC performance was matched. This pattern fits well with the CLS model's predictions; however, the lack of an FCC advantage in the standard recognition condition clearly contradicts the predictions of models like MINERVA 2 (Hintzman 1988).

Summary and conclusions

These experiments have shown that participants' ability to discriminate studied items from similar foils is a function of test format and whether participants are relying on recollection or familiarity. When participants have access to recollection, performance is comparable across test formats, but participants relying exclusively on familiarity perform much better on FCC tests (where they have a chance to make relative familiarity judgments between studied items and corresponding foils) than on YN and FCNC tests (where they do not). This pattern of results matches prior results from patients with selective hippocampal damage, with resultant selective recollection deficits, and confirms the predictions of the CLS dual-process computational model.

References

- Aggleton JP, Shaw C. Amnesia and recognition memory: A re-analysis of psychometric data. *Neuropsychologia*. 1996; 34(1):51–62. [PubMed: 8852693]
- Aggleton JP, Vann SD, Denby C, Dix S, Mayes AR, Roberts N, Yonelinas AP. Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia*. 2005; 43(12):1810–1823. [PubMed: 16154457]
- Baddeley, AD.; Emslie, H.; Nimmo-Smith, I. *Doors and people: A test of visual and verbal recall and recognition*. Bury St. Edmunds, England: Thames Valley Rest Co; 1994.
- Baddeley AD, Vargha-Khadem F, Mishkin M. Preserved recognition in a case of developmental amnesia: Implications for the acquisition of semantic memory? *Journal of Cognitive Neuroscience*. 2001; 13(3):357–369. [PubMed: 11371313]
- Barbeau EJ, Felician O, Joubert S, Sontheimer A, Ceccaldi M, Poncet M. Preserved visual recognition memory in an amnesic patient with hippocampal lesions. *Hippocampus*. 2005; 15(5):587–596. [PubMed: 15884094]
- Bastin C, Van der Linden M. The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology*. 2003; 17(1):14–24. [PubMed: 12597069]
- Bastin C, Van der Linden M, Charnallet A, Denby C, Montaldi D, Roberts N, Mayes AR. Dissociation between recall and recognition memory performance in an amnesic patient with hippocampal damage following carbon monoxide poisoning. *Neurocase*. 2004; 10(4):330–344. [PubMed: 15788271]
- Bayley PJ, Wixted JT, Hopkins RO, Squire LR. Yes/no recognition, forced-choice recognition, and the human hippocampus. *Journal of Cognitive Neuroscience*. 2008; 20(3):505–512. [PubMed: 18004951]
- Bell-McGinty S, Lopez OL, Meltzer CC, Scanlon JM, Whyte EM, DeKosky ST, Becker JT. Differential cortical atrophy in subgroups of mild cognitive impairment. *Archives of Neurology*. 2005; 62(9):1393–1397. [PubMed: 16157746]
- Brown MW, Aggleton JP. Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*. 2001; 2(1):51–61.
- Cipolotti L, Bird C, Good T, Macmanus D, Rudge P, Shallice T. Recollection and familiarity in dense hippocampal amnesia: A case study. *Neuropsychologia*. 2006; 44(3):489–506. [PubMed: 16023686]
- Cook GI, Marsh RL, Hicks JL. Revisiting the role of recollection in item versus forced-choice recognition memory. *Psychonomic Bulletin & Review*. 2005; 12(4):720–725. [PubMed: 16447387]

- Danckert SL, Gati JS, Menon RS, Köhler S. Perirhinal and hippocampal contributions to visual recognition memory can be distinguished from those of occipito-temporal structures based on conscious awareness of prior occurrence. *Hippocampus*. 2007; 17(11):1081–1092. [PubMed: 17696171]
- Diana RA, Reder LM, Arndt J, Park H. Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*. 2006; 13(1):1–21. [PubMed: 16724763]
- Dobbins IG, Kroll NEA, Yonelinas AP, Liu Q. Distinctiveness in recognition and free recall: The role of recollection in the rejection of the familiar. *Journal of Memory and Language*. 1998; 38(4):381–400.
- Düzel E, Vargha-Khadem F, Heinze HJ, Mishkin M. Brain activity evidence for recognition without recollection after early hippocampal damage. *Proceedings of the National Academy of Sciences*. 2001; 98(14):8101–8106.
- Eacott MJ, Easton A. On familiarity and recall of events by rats. *Hippocampus*. 2007; 17(9):890–897. [PubMed: 17604354]
- Fortin NJ, Wright SP, Eichenbaum H. Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature*. 2004; 431(7005):188–191. [PubMed: 15356631]
- Gardiner JM. Functional-aspects of recollective experience. *Memory and Cognition*. 1988; 16(4):309–313.
- Hall JF. Recognition as a function of word frequency. *The American Journal of Psychology*. 1979; 92(3):497–505.
- Hintzman DL. Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*. 1988; 95(4):528–551.
- Holdstock JS, Mayes AR, Gong Q, Roberts N, Kapur N. Item recognition is less impaired than recall and associative recognition in a patient with selective hippocampal damage. *Hippocampus*. 2005; 15(2):203–215. [PubMed: 15390152]
- Holdstock JS, Mayes AR, Roberts N, Cezayirli E, Isaac CL, O'Reilly RC, Norman KA. Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? *Hippocampus*. 2002; 12(3):341–351. [PubMed: 12099485]
- Jacoby LL. A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*. 1991; 30(5):513–541.
- Khoe W, Kroll NE, Yonelinas AP, Dobbins IG, Knight RT. The contribution of recollection and familiarity to yes-no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia*. 2000; 38(10):1333–1341. [PubMed: 10869576]
- Kroll NE, Yonelinas AP, Dobbins IG, Frederick CM. Separating sensitivity from response bias: Implications of comparisons of yes-no and forced-choice tests for models and measures of recognition memory. *Journal of Experimental Psychology: General*. 2002; 131(2):241–254. [PubMed: 12049242]
- Macmillan, NA.; Creelman, CD. *Detection theory: A user's guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates; 2005.
- Manns JR, Hopkins RO, Reed JM, Kitchener EG, Squire LR. Recognition memory and the human hippocampus. *Neuron*. 2003; 37(1):171–180. [PubMed: 12526782]
- Mayes A, Montaldi D, Migo E. Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*. 2007; 11(3):126–135. [PubMed: 17270487]
- Mayes AR, Holdstock JS, Isaac CL, Hunkin NM, Roberts N. Relative sparing of item recognition memory in a patient with adult-onset damage limited to the hippocampus. *Hippocampus*. 2002; 12(3):325–340. [PubMed: 12099484]
- Mayes AR, Holdstock JS, Isaac CL, Montaldi D, Grigor J, Gummer A, Cariga P, Downes JJ, Tsivilis D, Gaffan D, Gong Q, Norman KA. Associative recognition in a patient with selective hippocampal lesions and relatively normal item recognition. *Hippocampus*. 2004; 14(6):763–784. [PubMed: 15318334]
- Montaldi D, Spencer TJ, Roberts N, Mayes AR. The neural system that mediates familiarity memory. *Hippocampus*. 2006; 16(5):504–520. [PubMed: 16634088]

- Norman KA, O'Reilly RC. Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*. 2003; 110(4):611–646. [PubMed: 14599236]
- Parkin AJ, Yeomans J, Bindschaedler C. Further characterization of the executive memory impairment following frontal lobe lesions. *Brain and Cognition*. 1994; 26(1):23–42. [PubMed: 7986491]
- Quamme JR, Yonelinas AP, Norman KA. Effect of unitization on associative recognition in amnesia. *Hippocampus*. 2007; 17(3):192–200. [PubMed: 17203466]
- Quamme JR, Yonelinas AP, Widaman KF, Kroll NE, Sauve MJ. Recall and recognition in mild hypoxia: Using covariance structural modeling to test competing theories of explicit memory. *Neuropsychologia*. 2004; 42(5):672–691. [PubMed: 14725804]
- Rajaram S. Remembering and knowing: Two means of access to the personal past. *Memory and Cognition*. 1993; 21(1):89–102.
- Rotello CM, Macmillan NA, Van Tassel G. Recall-to-reject in recognition: Evidence from roc curves. *Journal of Memory and Language*. 2000; 43(1):67.
- Skinner EI, Fernandes MA. Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologia*. 2007; 45(10):2163–2179. [PubMed: 17445844]
- Snodgrass JG, Corwin J. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*. 1988; 117(1):34–50. [PubMed: 2966230]
- Tulving E. Similarity relations in recognition. *Journal of Verbal Learning and Verbal behavior*. 1981; 20(5):479–496.
- Tulving E. Memory and consciousness. *Canadian Psychology*. 1985; 26(1):1–12.
- Turriziani P, Serra L, Fadda L, Caltagirone C, Carlesimo GA. Recollection and familiarity in hippocampal amnesia. *Hippocampus* 9999. (9999): NA.
- Vargha-Khadem F, Gadian DG, Watkins KE, Connelly A, Van Paesschen W, Mishkin M. Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*. 1997; 277(5324):376–380. [PubMed: 9219696]
- Wais PE, Wixted JT, Hopkins RO, Squire LR. The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron*. 2006; 49(3):459–466. [PubMed: 16446148]
- Warrington, EK. Recognition memory test. Windsor: NFER-Nelson Publishing Company; 1984.
- Westerberg, CE.; Florczak, SM.; Parrish, TB.; Weintraub, S.; Mesulam, MM.; Mayes, AR.; Reber, PJ.; Paller, KA. When memory does not fail: Examining the neural basis of familiarity-based recognition in mild cognitive impairment. Marburg, Germany: Neuroimaging and Psychological Theories of Human Memory; 2006.
- Westerberg CE, Paller KA, Weintraub S, Mesulam MM, Holdstock JS, Mayes AR, Reber PJ. When memory does not fail: Familiarity-based recognition in mild cognitive impairment and alzheimer's disease. *Neuropsychology*. 2006; 20(2):193–205. [PubMed: 16594780]
- Wixted JT. Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*. 2007; 114(1):152–176. [PubMed: 17227185]
- Yonelinas AP. Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 1994; 20(6):1341–1354.
- Yonelinas AP. The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*. 2002; 46(3):441–517.
- Yonelinas AP, Jacoby LL. The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language*. 1995; 34(5):622–643.
- Yonelinas AP, Kroll NE, Dobbins I, Lazzara M, Knight RT. Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*. 1998; 12(3):323–339. [PubMed: 9673991]
- Yonelinas AP, Kroll NEA, Quamme JR, Lazzara MM, Sauve M-J, Widaman KF, Knight RT. Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience*. 2002; 5(11):1236–1241.

Appendix

The equations for calculating familiarity based hit and false alarm in a standard remember/know experiment are (taken from Yonelinas, Kroll, Dobbins, Lazzara & Knight 1998):

$$F_{Old} = \frac{P(C'F_{old})/old}{1 - P(C'R_{old})/old}$$

$$F_{New} = \frac{P(C'F_{old})/new}{1 - P(C'R_{old})/new}$$

A key difference between the justified-RF task and the standard remember/know procedure is that, in the justified-RF task, items can be recollected as either old *or* recollected as new. To allow for the latter possibility, we have amended the above equations so the denominator includes both a “recollection as old” term and a “recollection as new” term.

$$F_{Old} = \frac{P(C'F_{old})/old}{1 - P(C'R_{old})/old - P(C'R_{new})/old}$$

$$F_{New} = \frac{P(C'F_{old})/new}{1 - P(C'R_{old})/New - P(C'R_{new})/new}$$

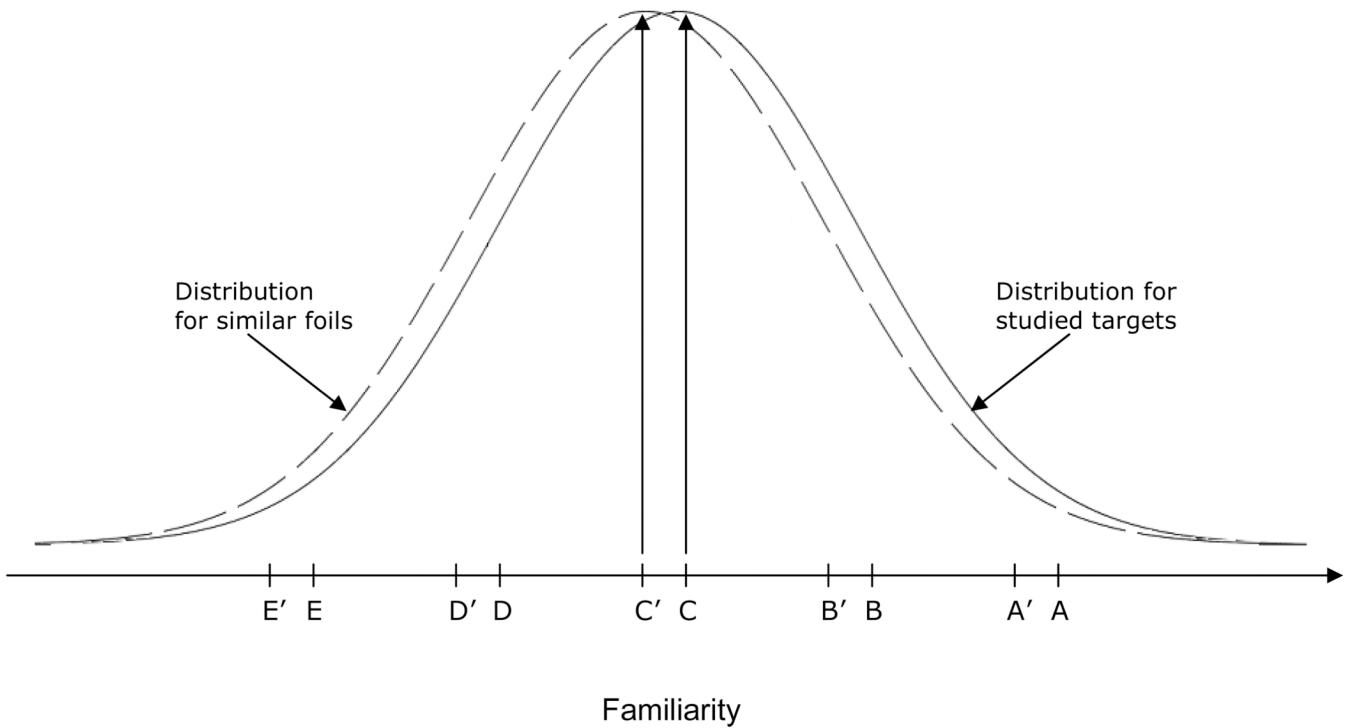


Figure 1.

Illustration of why familiarity discriminates well between studied items and similar foils on forced-choice-corresponding (FCC) tests, but not on yes-no (YN) or forced-choice non-corresponding (FCNC) tests. A, B, C, D, E stand for studied items and A', B', C', D', E' stand for foils that are similar to A, B, C, D, and E (respectively). An item's position on the X axis indicates its familiarity. The solid line represents the familiarity distribution for the target items while the dashed line represents the distribution for the foils. FCC discrimination is good because each item is reliably more familiar than its corresponding foil ($A > A'$, $B > B'$, and so on). YN and FCNC discrimination is poor because the distribution of studied-item familiarity scores overlaps strongly with the distribution of foil familiarity scores; this is indicated by the two overlapping bell curves (centred on C and C', respectively). This high degree of overlap means that there is no way to set a familiarity criterion that reliably separates studied items from foils (e.g. A' is more familiar than D).

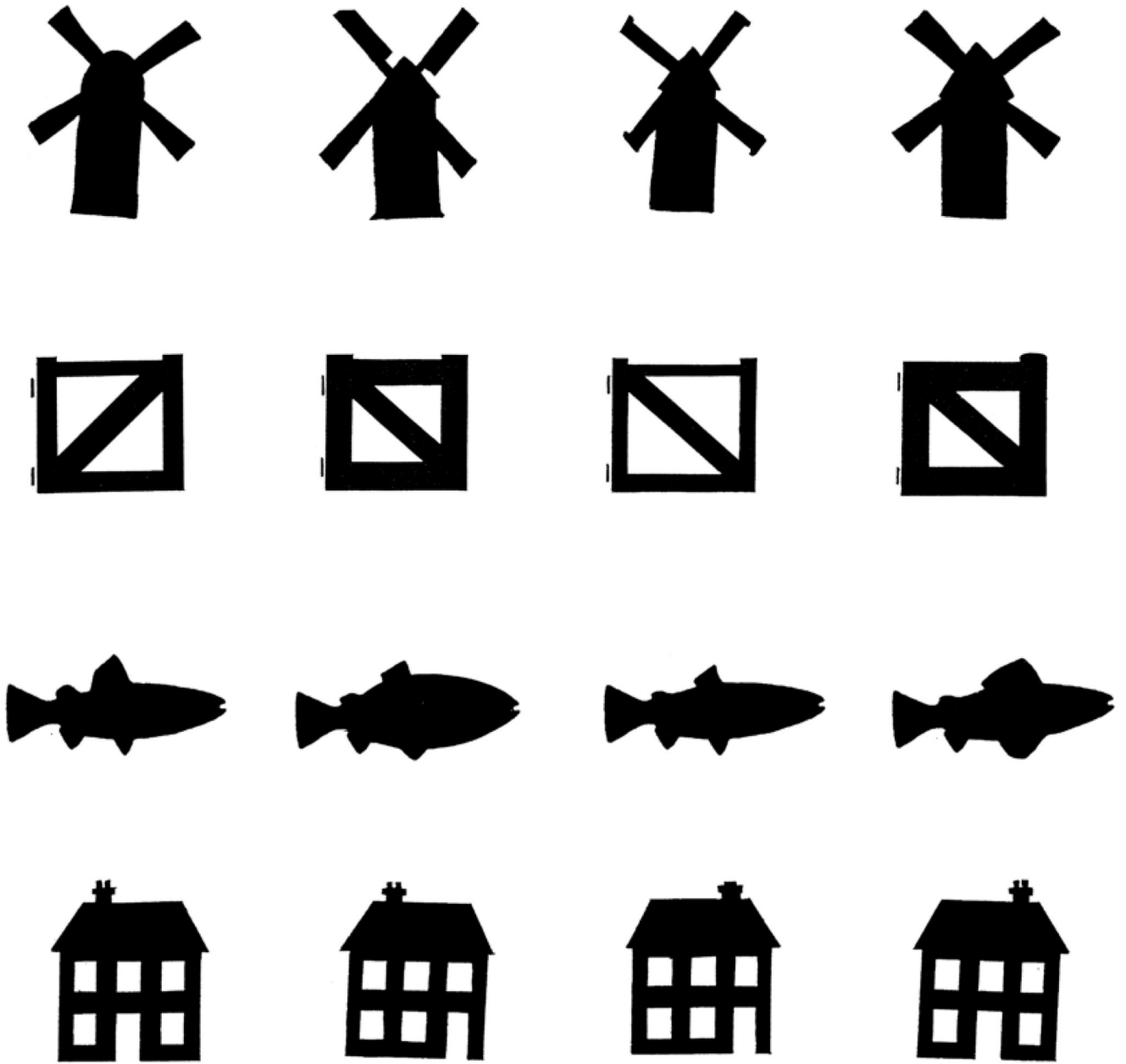


Figure 2. Examples of picture quartets used at stimuli. The target was randomly selected from the four pictures, with the three remaining pictures being used as similar foils.

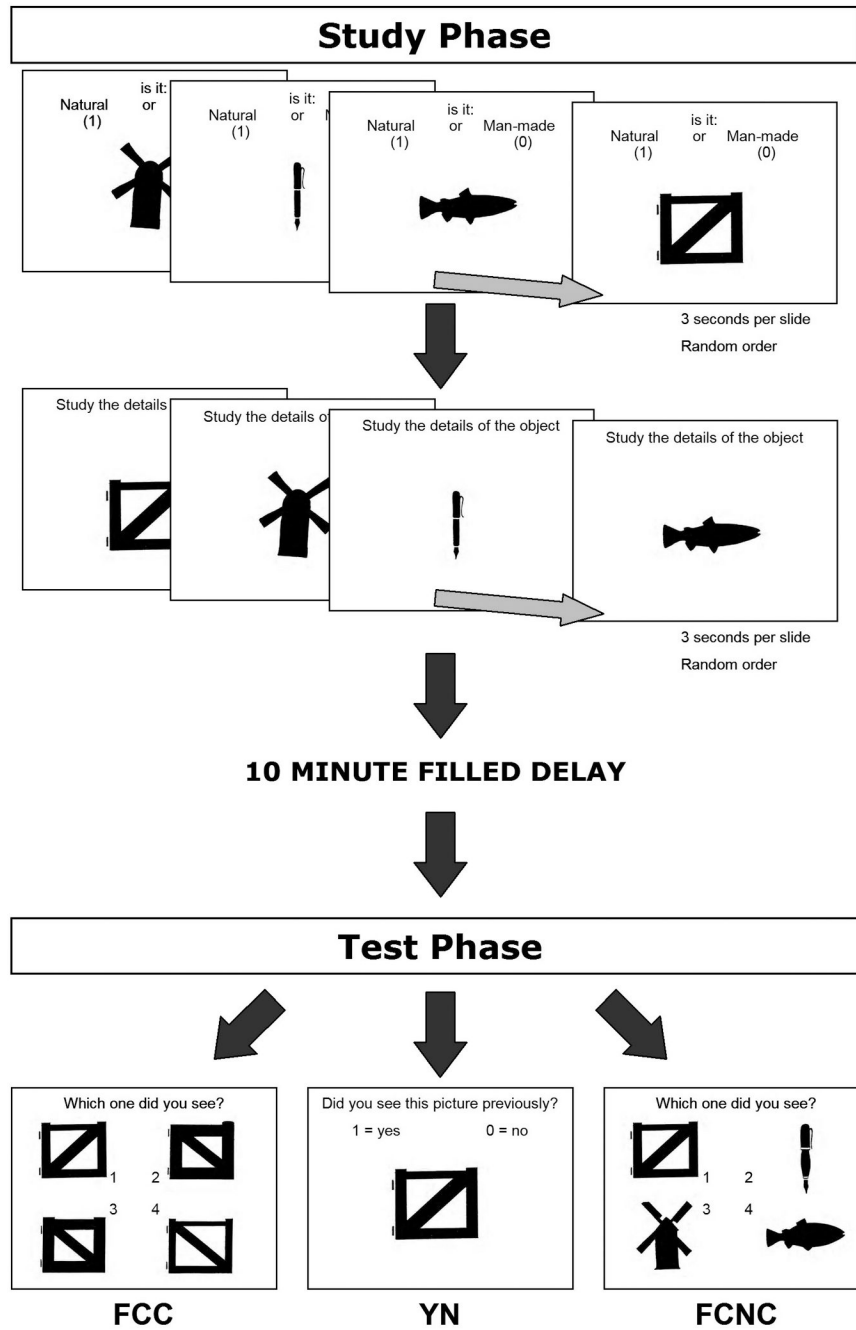


Figure 3.
Experimental procedure.

Table 1

Raw average scores for each test format and condition

For FCC and FCNC this is given as proportion correct and for YN hit and false alarm rates are given. All scores are to 3 significant figures.

Condition	FCC		FCNC		YN	
	Accuracy	SE	Accuracy	SE	Hit Rate	SE
Recognition	.604	.048	.630	.053	.719	.038
Familiarity-only	.512	.040	.326	.036	.858	.031
					.292	.027
					.645	.037

Table 2
Performance in all three test formats in Experiment 1 as measured by standard deviations below recognition mean

For FCC and FCNC this was calculated using percentage correct and for YN this was calculated using P_r .

Condition	FCC		FCNC		YN	
	Accuracy	SE	Accuracy	SE	Accuracy	SE
Familiarity-only	-.478	.051	-1.445	.043	-1.357	.057

Table 3
Responses collated across all participants, separated for familiarity and recollection
 Familiarity performance has assumed statistical independence, see Appendix.

	Old Items				New Items			
	Hit Rate		Miss Rate		Correct Rejection Rate		False Alarm Rate	
	Average	SE	Average	SE	Average	SE	Average	SE
Overall	0.750	0.039	0.250	0.039	0.733	0.035	0.267	0.035
Recollection	0.448	0.038	0.120	0.032	0.591	0.038	0.107	0.013
Familiarity (raw rates)	0.302	0.031	0.130	0.033	0.152	0.112	0.151	0.028
Familiarity (independence)	0.743	0.059	0.257	0.059	0.517	0.070	0.483	0.070

Table 4
Performance estimates for recognition and familiarity performance from Experiment 1 and Experiment 2

Results are reported using both P_r (hits minus false alarms) and d' .

	P_r		d'					
	Experiment 1	Experiment 2	Experiment 1	Experiment 1	Experiment 2	Experiment 2		
	Accuracy	SE	Accuracy	SE	Accuracy	SE		
Recognition	.437	.040	.492	.044	1.153	.127	1.359	.127
Familiarity	.210	.037	.261	.067	.613	.120	.572	.174