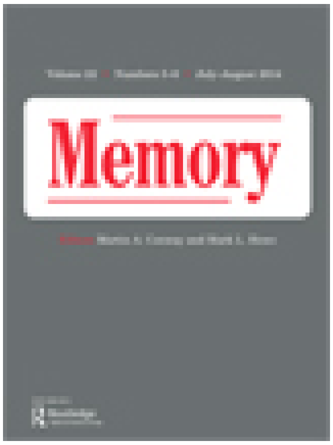

Memory

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pmem20

Context, remember-know recognition judgements, and ROC parameters

Salvador Algarabel [a] & Alfonso Pitarque [a]

[a] University of Valencia , Spain

Published online: 02 Jul 2007.

To cite this article: Salvador Algarabel & Alfonso Pitarque (2007) Context, remember-know recognition judgements, and ROC parameters, Memory, 15:5, 477-494, DOI: 10.1080/09658210701312226

To link to this article: http://dx.doi.org/10.1080/09658210701312226

MEMORY, 2007, 15 (5), 477–494

Psychology Press
Taylor & Francis Group

Context, remember–know recognition judgements, and ROC parameters

Salvador Algarabel and Alfonso Pitarque
University of Valencia, Spain

Recent work (e.g., Dunn, 2004; Heathcote, 2003) has questioned the necessity of postulating two processes to explain recognition memory. As part of this trend, strength theories of the remember–know methodology have gained in support. We present three experiments with pictorial material in which we force participants to use differential contextual information at test. Participants were required to give remember–know judgements and confidence ratings for each test stimulus. Hits, false alarms, remember–know data, and discrimination indices indicated systematic variations as a function of the availability and use of contextual information. Moreover, when we normalised the receiver operating characteristic data in terms of z-scores, the slopes were lower than 1 and slightly concave. Additionally, we computed the same set of statistical indices suggested by Wixted and Stretch (2004), with mixed results. Overall, we think that the data support a two-factor theory of remember–know and recognition, although many results fit well signal detection views of the task. Finally, the idea that remember and know responses are pure manifestations of recollection and familiarity seems difficult to sustain. We think that a productive use of the remember–know methodology involves the minimisation of the bias factors that may contaminate the responses, in addition to the introduction of the experimental manipulations needed to promote recollective and/or familiarity processes.

We have all experienced the feeling of knowing somebody because they look familiar to us although we cannot identify the source of this feeling. At other times a similar encounter elicits a rich series of images of the previous occurrence. In the recognition research literature both situations translate into a dispute between those who defend the view that the data can be explained by a single familiarity mechanism, and those who propose a duality of familiarity and recollection processes.

Traditional single-process views in their diverse variants (Donaldson, 1996; Dunn, 2004; Gillund & Shiffrin, 1984; Glanzer, Kim, Hilford, & Adams, 1999; Hilford, Glanzer, Kim, & DeCarlo, 2002; Hintzman, 1984, 1986; Hirshman & Master, 1997; Inoue & Bellezza, 1998; Murdock, 1974, 1982; Pike, 1984; Ratcliff, Sheu, & Gronlund, 1992; Shiffrin & Steyvers, 1997; Xu & Bellezza, 2001) analyse recognition performance from the point of view of signal detection theory. All items presented at test vary in the level of familiarity they produce. The decision as to whether a presented stimulus is categorised as new or old depends on the result of the comparison between the familiarity produced by the stimulus and a reference criterion set by the participant.

By contrast, all dual models (Yonelinas, 2002) view recognition as the product of a mixture of familiarity and retrieval or recollection. Historically, the models of Atkinson and Juola (1974), Mandler (1980), and Jacoby & Dallas (1981) are

Address correspondence to: Salvador Algarabel, Facultad de Psicologia, Universidad de Valencia, Blasco Ibanyez, 21 46010 Valencia, Spain. E-mail: Salvador.Algarabel@uv.es

This research was supported by Grant SEJ2004-02541 from The "Dirección General de Investigación Científica y Técnica (Spanish Ministry of Education and Technology").

We thank John T. Wixted for his thoughtful comments on an earlier version of this article.

http://www.psypress.com/memory DOI:10.1080/09658210701312226

well-known precursors, although the theory put forward by Yonelinas (1994) and the dual views born around the remember–know (Tulving, 1985) methodology have been the most influential. Yonelinas's model postulates that familiarity and recollection are two qualitatively different cognitive mechanisms involved in the generation of specific evidence needed for a recognition decision. In this model, familiarity is described by a signal detection model with equal variances, which behaves in a way similar to the classical analysis of recognition by traditional one-process models. In contrast, recollection reflects an all-or-none or threshold process whereby contextual information about the study episode is retrieved, and contributes to a decision. Both processes are initiated in parallel, and their weight in the recognition decision is dependent on a number of factors, among which the testing conditions are particularly relevant. Additional and relevant characteristics include the fact that familiarity is faster (e.g., Boldini, Russo, & Avons, 2004), and it is assumed to support a wide range of confidence responses, as opposed to recollection that supports mostly high-confidence decisions.

The Yonelinas model is a specific version of the broad idea that there are two independent sources of evidence contributing to recognition. In this paper we link this broad, general two-process idea to many researchers who use the remember–know methodology (Gardiner & Richardson-Klavehn, 2000; Tulving, 1985) despite not being formulated as a formal and more specific model of recognition. One could argue that this loose view of the two-process model is identical to the signal detection model (see Wixted & Stretch, 2004), in which a final decision is taken based on the combined evidence. However, we think that if a different neural basis and different behavioural properties may be linked to the two postulated mechanisms, beyond simple differences in criteria, then both points of view should be treated separately.

In a remember–know experiment, people are asked whether they are able to recall peripheral details associated with the target (remember response) or simply if they judge that the target stimulus was studied although they are not able to place it in any specific context. Originally, both types of judgements were intended to reflect characteristics associated with episodic and semantic memory, respectively. However, the division between episodic and semantic memory is no longer so clearly maintained, and the dichotomy between remember and know recognition responses has been taken as indicating dual processes (Gardiner & Java, 1983), particularly through the demonstration of functional dissociations between both types of judgements in relation with the influence of a number of variables (see Gardiner & Richardson-Klavehn, 2000). According to this more contemporary view of the remember–know task, the responses reflect the influence of recollection on the one hand and familiarity in the absence of recollection on the other, respectively. An alternative signal detection interpretation of the remember–know data has emerged (Donaldson, 1996; Dunn, 2004; Hirshman & Master, 1997; Inoue & Bellezza, 1998; Wixted & Stretch, 2004; Xu & Bellezza, 2001), which sees remember responses as simply indexing greater memory strength than know responses, and categorised as such because their memory strength surpasses a decision criterion set by the participant.

The testing of these theories relies heavily on the tools provided by signal detection theory, and in the case of the remember–know methodology in the finding of experimental dissociations between both response categories. In particular, when hits (proportion of times that one correctly identifies a studied item) are plotted against false alarms (proportion of times that one incorrectly identifies a novel stimulus as a studied one), a receiver operating curve is obtained. Three parameters calculated from this basic function are of special relevance to recognition memory theories: its form when intact, and when transformed by a z-normalisation (z-ROC), and the slope of the z-transformed function. When the basic ROC function is analysed its form is almost always nonlinear, leading to the rejection of a whole family of linear threshold models (see Murdock, 1974). Usually, the function is convex and asymmetrical along the negative diagonal, as indicated by a significant component of the quadratic constant of the equation fitted to the data. All models discussed in this paper either single or dual, agree in this prediction. However, if the ROC function is replotted in z-ROCs coordinates, most single-process models would predict a linear function, meaning a non-significant quadratic component of the corresponding quadratic equation. However, Yonelinas's model predicts a curvilinear z-ROC function when recollection is present (Yonelinas, 1994). Therefore if the condition that familiarity is governed by a signal detection model with equal variances is to be held in the model, the z-ROC function must be curvilinear. The fact that the slope of the z-ROC functions is around 0.80 (Ratcliff et al., 1992), means that the study phase leads to an increase in dispersion of the familiarity values of the corresponding items. This is so because the slope of the z-ROC linear function is the ratio of the new to the old standard deviations of the respective distributions.

As indicated previously, those who defend dual positions using the remember–know methodology do so arguing the independence of both types of judgements given the large set of variables showing dissociations between them (Gardiner & Richardson-Klavehn, 2000). An important characteristic of the methodology, and a weakness for the critics, is that it is introspective in nature (see Dunn, 2004; Inoue & Bellezza, 1998; Xu & Bellezza, 2001). However, we think that the final decision on the goodness of the remember–know (R/K) procedure depends on the evidence available in favour of the existence of different determinants for both types of responses, and the demonstration that the dissociations found in the literature cannot be explained by a single process.

With regard to the first aspect, the determinants of remember and know judgements remain largely unexplored (Hicks, Marsh & Ritschel, 2002; Java, Gregg & Gardiner, 1997; Perfect, Mayes, Downes & van Eijk, 1996). Perfect et al. (1996) asked their participants about memories of spatio-temporal contextual details (word order, quadrant in which it was originally written, changes in typeface and size, and so on) systematically presented and manipulated in the study phase. They found that people were able to provide more details when they made a remember judgement than when they said they knew. In the latter case, participants were also able to provide a certain level of contextual information, although in lesser amounts. These researchers also realised that the accuracy of the contextual information people had was quite low (see also Hicks et al., 2002).

However, a recent paper by Dudukovic and Knowlton (2006) has shown that the number of episodic details is more clearly associated with remember responses, and a long retention interval produces a change from remember to know, accompanied by decrease in contextual details. In one of our previous papers (Algarabel, Pitarque & Gotor, 2006), we also showed a change from remember to know judgements as a function of retention interval. Reder, Donavos, and Erickson (2002) have also shown that the matching of font type between study and test produced better recognition and remember responses than mismatches or the use of less distinctive fonts. Finally, Macken (2002) has observed that changing context between study and test affected recollection, but not familiarity, although there was no certainty that people were not using overall familiarity in responding. These demonstrations do not prove either of the two theoretical points of view that we are discussing, but could be taken into account as supporting evidence in favour of acceptance of the existence of two mechanisms for recognition.

More recently, however, Rotello, Macmillan, Reeder, and Wong (2005) have questioned the viability of the procedure used regularly in remember–know experiments. They identified two different set of instructions used in the literature, and categorised them as conservative (as in Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996) and liberal (as in Rajaram, 1996). When they measured simultaneously confidence judgements and remember–know responses under both sets of instructions, they observed that hits and false alarms varied accordingly, and more important, that signal detection theories could provide a better account of the data. The fact that introspective judgements are subject to bias is understandable, but the question remains whether the procedure is able to capture the basic mechanisms of recognition under circumstances in which people have sufficient opportunities to recollect. This is the case in Dudukovic and Knowlton's paper (2006) in which participants were exposed to rich pictorial elements as part of the procedure. Nevertheless this is an open question, of particular importance given the criticisms of Rotello et al. (2005).

The second previously mentioned problematic aspect of the R/K procedure has to do with recent claims by single-process proponents (e.g., Dunn, 2004) that the dissociations found in the R/K procedure do not provide unequivocal support for two-process theories. In particular, Dunn (2004) has shown convincingly that the results obtained in a large compilation of experiments with regard to those arguments could also be predicted from a signal detection model with equal variances. It is also important to point out that the analysis brought about by Dunn does not try to explain simultaneously critical data not obtained with the R/K procedure. To this category belongs the need to provide an account of the values of the slopes of the z-ROC functions, usually around 0.80, incompatible with the signal detection model with equal variances. In conclusion, current research has shed some doubts on the validity of the remember–know methodology and the corresponding dual views, based on the apparent contamination of the data by criterion issues, and the possible alternative explanation by signal detection models of the dissociation experiments.

The main motivation for the current series of experiments is to assess the one- and two-process interpretations of the remember–know procedure, and to evaluate the methodology itself as a valuable tool to analyse recognition. Although the neurophysiological experimentation shows an overwhelming support for the two-process view of recognition (see a recent review in Aggleton & Brown, 2006), there has been a progressive trend in the behavioural literature to show increased support for the signal detection view (e.g., Dunn, 2004; Heathcote, 2003; Wixted & Stretch, 2004), and its interpretation of the results obtained with the remember–know methodology. We believe that this is due to the susceptibility of the task to a number of extraneous factors, besides those cognitive processes directly involved in recognition (e.g., Rotello & Macmillan, 2006; Rotello et al., 2005). Therefore, in the experiments that follow we try to require the use of contextual evidence to have a chance of getting a correct response. In addition, we expect that this influence in objective discrimination translates into the subjective judgements, giving clues to the validity of the R/K procedure. Finally, we use ROC parameters to determine whether this influence can be subsumed under a single variable or is the result of two different processes. We do not subscribe to the view that recollection is always used in an item recognition experiment (see Diana, Reder, Arndt, & Park, 2006) as a necessary part of a general two-process view, but assume that encoding and testing conditions establish its use to varying degrees.

In the three experiments we describe below, we use real-world scenes as stimuli and we look for independent evidence that remember and know responses reflect two independent processes. The reason is that pictures are complex and rich stimuli in which all parts are coherently interrelated. When one is confronted with a partial or degraded view of a previously seen scene, and this is the rule when we have a mental image from a past event in mind, it is very likely that one has to use a recall to confirm or similar strategy to aid in the decision. This is in opposition to the situation in which one sees a stimulus originally presented with an impoverished non-interactive context (see Baddeley, 1982; Murnane, Phelps, & Malmberg, 1999), which may render it useless as a cue for retrieval. The use of pictorial stimuli in testing one- versus two-factor accounts of recognition is not new (e.g., Chandler, 1994; Dobbins, Kroll & Liu, 1998), although our approach is different, as we are trying to address some of the concerns elicited by recent signal detection approaches (e.g., Dunn, 2004).

We present three experiments with very similar logic; namely to manipulate the test stimulus in such a way that a participant has to explicitly use different degrees of contextual information to confirm or reject the test stimulus. We define ''the context'' of an item as the encoding aspects or operations carried out at study and stored in memory, not physically present at test, but that can help in the recognition decision. In the first experiment, we define a sizeable portion of a picture as ''context'' and the remaining part as the study item. At test, only the latter is presented. It is obvious that there are many other dimensions of the study situation that are not directly under our control but that may also play a role (mental images, etc.). However, we expect that our ''contextual manipulation'' overshadows those other possible functional factors not directly manipulated. The design of the second experiment is identical to the first one, except for the increment in the time interval between ''context'' and ''study item'' at presentation time. In the third experiment, presentation is identical for all participants; that is, they see complete photographs, but at test time some people recognise them on the basis of the whole stimulus and others on a small portion of the original photograph. We predict that all conditions in which the test stimulus is a partial representation of the original picture, or there exists the possibility of using the ''context'' presented at study in the recognition phase, will lead to heavier use of retrieval mechanisms than those in which the stimulus is identical in both cases, and also to a higher level of recognition. The additional question of interest is whether those retrieval mechanisms brought into play contribute as a unique entity (one-process theories) or not (two-process theories) to recognition. With these goals in mind we also required a subjective judgement (remember or know) and a confidence in the yes–no response.

In this way we will analyse the objective data (proportion correct or discriminability indices), subjective judgements, and ROC parameters obtained from the confidence responses.

Data analysis will be carried out looking at the different proportions of responses partitioned by type of judgement, and their relations to confidence (see Wixted & Stretch, 2004). Very recently, Wixted and Stretch have put together a compelling set of arguments in favour of the detection model based on new analyses of the remember–know data. We will present a similar set of analyses for the present data, although we were not aware of Wixted and Stretch's approach at the time we planned and ran our experiments. In particular, Wixted and Stretch (2004) centred most of their analyses on the relation of false alarms with other parameters calculated from the data, claiming that the signal detection view is uniquely supported by data analyses carried out on them. We will present the results obtained from our experiments based on the idea that false alarms can also be subject to illusory recollective processes.

We will also analyse key ROCs parameters, relating them to the rest of the data. Specifically, we will look at three parameters: the curvature of the untransformed ROCs and the linearity and slope of the z-ROCs. According to past analysis (Hilford et al., 2002; see also Heathcote, 2003), ROCs must be curvilinear if the underlying mechanism is not threshold based. On the other hand, the z-ROCs would be nonlinear if Yonelinas's dual model is correct, and linear according to signal detection based alternatives. Although the different models can be accommodated to explain apparently contradictory results, the presence of a nonlinear quadratic component in the standardised data is associated with the use of recollection, according to Yonelinas's model. An additional specific point of his theory is that it cannot predict the simultaneous presence of a significant quadratic constant and a slope lower than 1 in the z-transformed data. A final comparison of interest is between the z-ROC slopes estimated from the confidence ratings, and the same parameter from the remember–know judgements. It has recently been indicated (Rotello, Macmillan, & Reeder, 2004) that if the signal detection model is correct, both slopes should be equal. This prediction is similar to that between discrimination calculated from all data and from only the remember responses (see Donaldson, 1996). Rotello et al. (2004) have found that both slopes are different, with the one estimated from the remember data not different from 1, as in the source recognition literature (e.g., Hilford et al., 2002). In this paper we will have the opportunity to compare the z-ROC and the two-point z-ROC slopes calculated from the confidence and remember–know data respectively, obtained from the same data set. Given that we will have the confidence and remember–know responses for the same set of stimuli, we will carry out the same analyses that Wixted and Stretch (2004) have carried out recently, trying to generate new predictions on the relation between both variables from the point of view of the signal detection model.

As in previous papers (Glanzer et al., 1999; Hilford et al., 2002), we estimated the intercept and slopes of z-ROC by maximum likelihood and least square procedures for the group data. Given that the correlations between both estimations were 0.997 (intercepts) and 0.9983 (slopes) for the seven conditions of our three experiments (Hilford et al., 2002), we used least square estimations throughout the paper because of their convenience. All statistical fits were carried out on the individual data, except the estimations for the contrast between z-ROCr and global z-ROCs.

EXPERIMENT 1

Method

Participants. The participants were 96 psychology student volunteers at the University of Valencia (Spain) who participated for extra course credit. They were allocated randomly into three groups of 32 participants.

Materials and apparatus. Two sets of 68 photographs of human and natural landscapes from different parts of the world, taken by one of the authors and closely matched in content and appearance, were formed. The photographs were about 800 pixels high $\times$ 526 wide, and they were projected onto computer screens set at a resolution of 1024×768 pixels. We created two versions of each photograph. The study item consisted of the small part of an image that could be seen between two central white concentric circles drawn inside two squares of 166×166 and 107×107 pixels, respectively (see Figure 1A). The remaining area of the photograph was

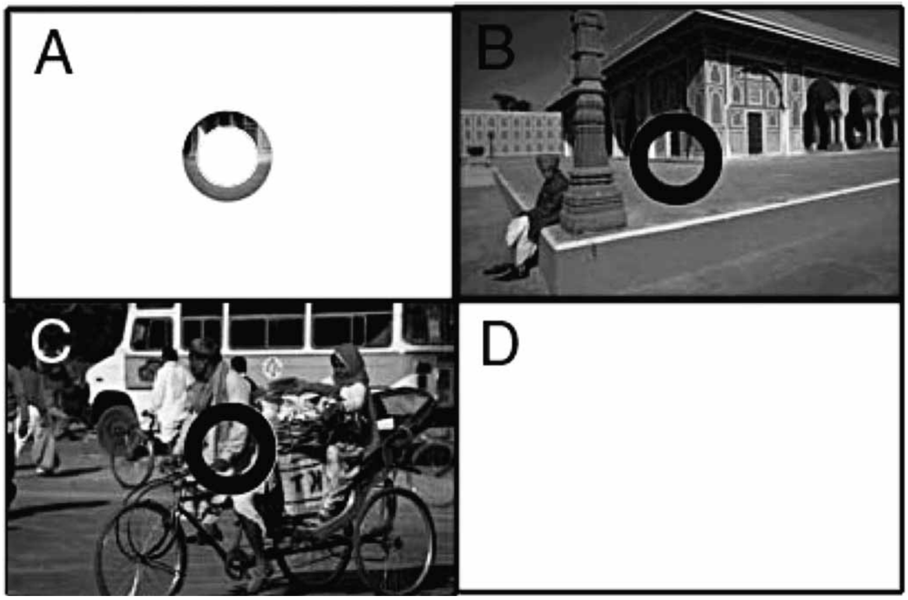

Figure 1. Examples of stimuli presented in Experiments 1 and 2: The study item common to all conditions (A), and the congruent (B), incongruent (C), and context absent (D) manipulation. Original in colour.

masked in white. Given the small depicted area of each picture, the viewable area was hardly recognisable. A second version of the picture, which would be used as context in some conditions, was created in which the visible area was reversed and the item was masked (see Figure 1B and 1C). From each list 34 pictures were assigned to the study condition, with the first and last two not being tested. The remaining pictures within each set were assigned to the recognition test. List order and type were counterbalanced across participants.

Study and test lists were randomised individually for each participant and presented using e-prime software for experimental control (Schneider, Eschman, & Zuccolotto, 2002).

Design and procedure. The experiment consisted of two study–test cycles in which a list of items was presented for study and then mixed with a new set for a recognition test. The same process was repeated with a second list. There were three different between-subjects' conditions differentiated by the type of context presented at study. In the congruent condition, the masked part of the study image was presented for 3 seconds (Figure 1B), followed 100 ms later by the study item for 2 seconds (Figure 1A). The two complementary images never coexisted. In the incongruent condition, the context (Figure 1C) was unrelated to the study item (Figure 1A) and came from a different picture. In the third condition no context was presented, being replaced by solid white (Figure 1D), followed by the study item (Figure 1A). The incongruent contexts were extracted from the "new" items to be presented at test. The intertrial interval was 1 second, during which the screen went blank. The recognition test consisted of the presentation of the 30 studied items mixed with 30 new ones. Participants were asked to recognise the test item on a 6-point confidence scale, from very confident that it was presented, to very confident that it was not. The keys corresponding to the letters D, E, G, H, J, and K were labelled for that purpose. Participants were further instructed to report their state of awareness in case of positive recognition (instructions modelled according to Rajaram, 1996). A remember response was defined as when recognition was accompanied by specific details experienced at the item's occurrence, including the specific context presented. A know response was defined as when one has the feeling that the item was presented, but was unable to bring to mind any specific details about its occurrence in the list. The know and remember responses were assigned to the V and B keys on the computer keyboard.

TABLE 1
Mean proportions (and standard errors) for recognition as a function of context condition for Experiment 1

	Remember		*Know*		
Context	*Hits*	*False Alarms*	*Hits*	*False Alarms*	*d′*
Absent	.46 (.03)	.07 (.01)	.22 (.02)	.10 (.01)	1.45 (.08)
Congruent	.64 (.02)	.06 (.01)	.17 (.02)	.15 (.02)	1.77 (.10)
Incongruent	.41 (.03)	.19 (.03)	.21 (.02)	.18 (.02)	0.69 (.12)

Results and discussion

In the Appendix we reproduce the raw data by confidence and condition for the three experiments reported here. The recognition scores for Experiment 1 are summarised in Table 1, which shows the mean proportions of responses separately by conscious state, as well as the discrimination indices. The significance level for all the statistical tests was $p < .05$, unless otherwise noted. As the interaction between the list and the remaining variables was not significant, we dropped it from further analyses and simply combined the data from both lists.

Discrimination and remember–know data. There were different levels of hits depending on the context condition (see Table 1), $F(2, 93) = 27.21$, $MSE = 0.01$, $p < .01$. Newman-Keuls tests showed that the congruent context produced the highest level of performance, followed by the no context condition, and leaving the incongruent context condition in last position. This difference in performance was reflected in the remember judgements, $F(2, 93) = 18.71$, $MSE = 0.02$, $p < .01$, which, in the congruent context condition, were superior to the rest on the Newman-Keuls tests. In contrast, the differences on know hits were marginally significant, $F(2, 93) = 2.48$, $MSE = 0.01$, $p = .09$. Context manipulation produced an effect on d', $F(2, 93) = 29.82$, $MSE = 0.33$, $p < .01$. All contrasts were significantly different as shown by the Newman-Keuls tests. The false alarm rates showed an opposite pattern, $F(2, 93) = 15.41$, $MSE = 0.02$, $p < .01$, with the incongruent context producing the highest false alarm rates on the Newman-Keuls test. The very same pattern of results was observed in the analysis of remember false alarms, $F(2, 93) = 10.76$, $MSE = 0.02$, $p = .01$. In "know" false alarms, $F(2, 93) = 5.68$, $MSE = 0.01$, $p = .01$, only the incongruent differed from the no context condition.

Remember responses were made with higher confidence than know responses in the three conditions: $t(31) = 4.68$, $t(31) = 17.80$, and $t(31) = 9.31$, $p < .01$, for the absent, congruent, and incongruent conditions, respectively. Similarly, the remember false alarms were also made with higher confidence than know false alarms, $t(31) = 2.79$, $t(31) = 6.35$, and $t(31) = 8.32$, $p < .01$, respectively, as well as the remember false alarms in relation to know hits, but only for the congruent and incongruent conditions, $t(31) = 6.91$, $t(31) = 5.76$, $p < .01$, respectively. There was a positive correlation between the know hit and know false alarm rates across participants ($r = .44$, $p < .01$).

ROC and z-ROC analyses. Table 2 presents the parameters obtained in the transformation of the group's ratings into ROCs and z-ROCs by least square fitting. Figure 2 presents the plot of the ROCs and z-ROCs data for the three experiments. The three ROC quadratic constants differed from zero, $t(31) = 5.78$, $p < .01$ (no context); $t(31) = 4.68$, $p < .01$ (congruent); $t(31) = 5.11$, $p < .01$ (incongruent), indicating that the three curves were convex. The z-ROC quadratic constants (see Table 2) of the no context and incongruent conditions were significantly different from zero, $t(31) = 3.07$, $p < .01$, and $t(31)$

TABLE 2
Statistics (mean and standard error) for the ROCs, *z*-ROCs of Experiment 1

	ROC		*z*-ROC			
Context	*quadratic constant*	*R2*	*linear slope*	*R2*	*quadratic constant*	*R2*
Absent	−1.53 (0.27)	.975	0.72 (0.03)	.987	0.08 (0.02)	.997
Congruent	−1.61 (0.34)	.963	0.78 (0.08)	.996	0.05 (0.06)	.998
Incongruent	−0.53 (0.10)	.999	0.91 (0.07)	.995	0.10 (0.04)	.999

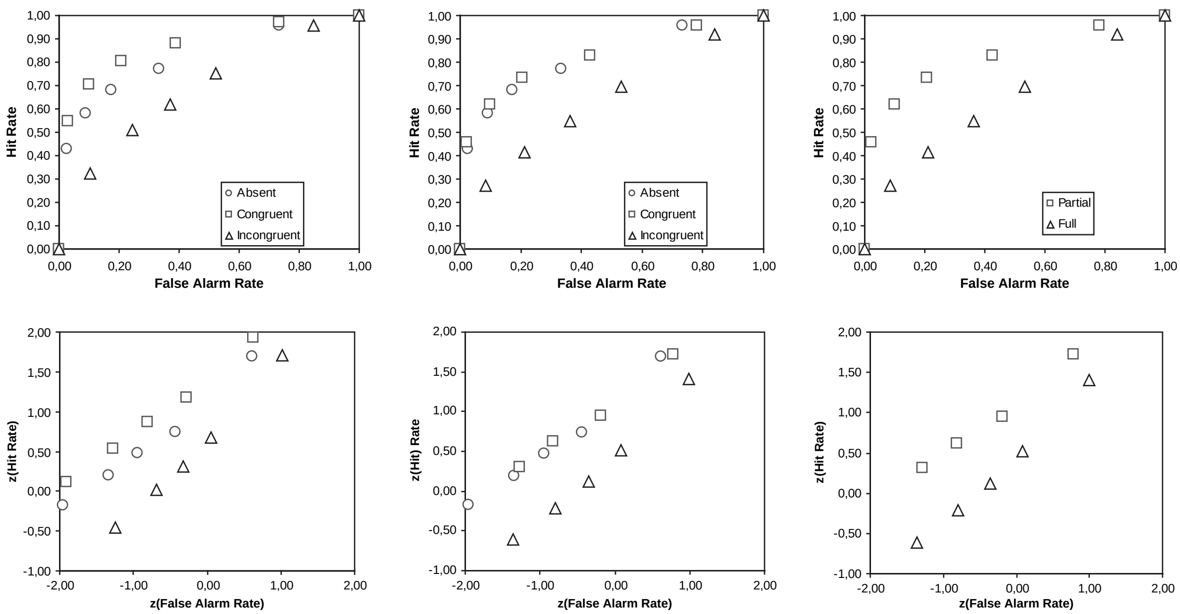

Figure 2. Receiver operating characteristics (top figures) and receiver operating characteristics plotted in standard scores (bottom figures) for the different conditions of each experiment (left: Experiment 1; centre: Experiment 2; right: Experiment 3).

$= 2.36$, $p = .025$, respectively. Power for detecting a "medium" effect size (0.5 standard deviation; Cohen, 1988) in the congruent condition was 0.63. These results, when significant, indicated that the functions were concave. In contrast, the z-ROC quadratic constant of the related condition was not different from zero, $t < 1$. With regard to the z-ROC linear slopes, the absent as well as the congruent conditions differed from 1, $t(31) = 10.24$, $p < .01$; $t(31) = 2.95$, $p < .01$, whereas the incongruent one did not, $t(31) = 1.27$.

As indicated in the introduction, we also computed the two-points z-ROC slopes estimated from the remember–know judgements, to be able to compare them with the z-ROC slopes estimated from the whole set of data. This computation is carried out by subtracting normalised remember hits from overall hits, and dividing by the corresponding difference in normalised false alarms (see Rotello et al., 2004, p. 590). This computation was carried out on the global data, given the large number of divisions by zero when the individual data were taken for the analysis. The z-ROCr values were 1.09, 0.70, and 0.97 versus z-ROC values of 0.72, 0.70, and 0.96[1] for the control, congruent, and incongruent conditions respectively.

This first experiment indicates that subjective judgements can, under normal conditions, be very susceptible to systematic contextual manipulations and not only to bias (as in Rotello et al., 2005). As expected, remember responses were highest in the congruent condition followed by the no context and incongruent conditions. This is a very systematic proof that people are using mental evidence to identify the test item as studied or not. A similar argument can be made with regard to the know responses. Note that the evidence categorised as "know" by people in the "congruent" condition is almost nil (.17 hits versus .15 false alarms) in contrast to the remaining two conditions, as expected given the decreasing saliency of the context in the "absent" condition and the misleading context in the "incongruent" one. The important question now is whether those differences in discrimination can be attributed to one or two sources of information. We expected that the differences in the z-ROC quadratic constants would be different from zero first and foremost in the congruent, and possibly in the "absent", manipulation because we did not attempt to preclude completely the possibility that the participants used their own contextual events in retrieval (other contextual details not manipulated by the experimenter). A similar reasoning follows with regard to the slopes computed from the z-ROCr and global z-ROC confidence data. The data fit this general

[1] Note the similarity between these z-ROC values estimated globally and the values estimated individually (see Table 2).

prediction except for the main condition; that is, the ''congruent'' one. A possible explanation of this discrepancy lies in the use of a very short time interval between context and item presentation at encoding, and the partial stimulus presentation at test may have not encouraged the use of recollection enough because the whole stimulus could pop out. A second aspect of the data is the high standard error of estimation of that quadratic constant (0.06), sometimes present in these types of experiment given the difficulty of including more stimuli for study. Following this reasoning, we replicated the first experiment using a longer-context item at study.

EXPERIMENT 2

Method

Participants. The participants were 64 psychology student volunteers at the University of Valencia (Spain) who participated for extra course credit. They were allocated randomly into two groups of 32 participants.

Materials and procedure. The materials and procedure were identical to those used in Experiment 1, except for the delay interval between the context and the target stimuli. Instead of 100 ms, this delay was set at 1000 ms. This value was a compromise between the need to differentiate both stimuli, and an acceptable level of recognition performance to be achieved on the basis of the test stimuli.

Results and discussion. The recognition scores for Experiment 2 are summarised in Table 3. In the following analysis, the control group data of the previous experiment were also introduced as

TABLE 3
Mean proportions (and standard errors) for recognition as a function of context condition for Experiment 2

	Remember		*Know*		
Context	*Hits*	*False Alarms*	*Hits*	*False Alarms*	d'
Absent	.46 (.03)	.07 (.01)	.22 (.02)	.10 (.01)	1.45 (.08)
Congruent	.52 (.03)	.05 (.01)	.21 (.02)	.15 (.02)	1.53 (.08)
Incongruent	.32 (.02)	.17 (.03)	.23 (.02)	.19 (.02)	0.52 (.10)

The control condition (Absent) data come from Experiment 1.

a baseline, given that both experiments are identical except for the delay interval, which is not defined in the control group.

Discrimination and remember–know data. There were different levels of hits depending on the context condition (see Table 3), $F(2, 93) = 21.15$, $MSE = 0.01$, $p < .01$. Newman-Keuls tests showed that the incongruent context produced the lowest level of performance, whereas no differences were found between the remaining ones. The same pattern of results was found analysing d', $F(2, 93) = 5.48$, $MSE = 0.23$, $p < .01$, and the proportions of remember responses, $F(2, 93) = 15.65$, $MSE = 0.02$, $p < .01$. No difference was found in the know category, $F(2, 93) < 1$, $MSE = 0.01$. With regard to false alarms, differences were found between conditions, $F(2, 93) = 21.52$, $MSE = 0.02$, $p < .01$, with the unrelated context producing the highest rates by Newman-Keuls tests. The very same pattern of results was obtained with remember false alarms, $F(2, 93) = 15.75$, $MSE = 0.01$, $p < .01$. In case of know false alarms, the control condition led to the lowest rate, $F(2, 93) = 7.08$, $MSE = 0.01$, $p = .01$.

Remember responses were made with higher confidence than know responses both in the congruent and incongruent conditions, $t(31) = 11.17$, $t(31) = 5.52$, $p < .01$, respectively. Similarly, the remember false alarms were also made with higher confidence than know false alarms, $t(31) = 8.73$, $t(31) = 3.81$, $p < .01$, respectively, as well as the remember false alarms in relation to know hits, $t(31) = 4.58$, $t(31) = 3.09$, $p < .01$, respectively.

ROC and z-ROC analyses. The absent and congruent ROC quadratic constants differed from zero (see Table 4), $t(31) = 5.78$, $p < .01$, $t(31) = 9.08$, $p < .01$, respectively, but not the incongruent one, $t(31) = 1.40$, $p > .05$. The z-ROC quadratic constants of both congruent and incongruent conditions did not differ from zero, $t(31) = 1.48$, $p > .05$, $t(31) = 1.82$, $p = .078$, respectively, although the absent condition did, $t(31) = 3.07$, $p < .01$. Finally, all z-ROC linear slopes differed from 1, $t(31) = 10.24$, $p < .01$; $t(31) = 11.54$, $p < .01$, $t(31) = 2.05$, $p = .05$.

The z-ROCr slopes were 0.72 and 0.98 and 0.64 and 0.86 when estimated from the remember or confidence data, for the congruent and incongruent conditions respectively. This discrepancy may indicate that R responses cannot be taken simply as higher-confidence responses, as the signal detection model postulates.

TABLE 4
Statistics (mean and standard error) for the ROCs of Experiment 2

	ROC		*z*-ROC			
Context	*quadratic constant*	*R2*	*linear slope*	*R2*	*quadratic constant*	*R2*
Absent	−1.53 (0.27)	.975	0.72 (0.03)	.987	0.08 (0.02)	.997
Congruent	−1.21 (0.13)	.967	0.70 (0.03)	.992	0.04 (0.02)	.998
Incongruent	−0.26 (0.18)	.999	0.90 (0.05)	.993	0.16 (0.09)	.999

The control condition (Absent) data come from Experiment 1.

Despite a drop in accuracy from the first to the second experiment, the parameters resulting from the fits are very similar to those obtained in the first one. In this experiment key parameters of the incongruent condition became non-significant. Variability and lower accuracy could explain this discrepancy. Given the variability of the previous results, we attempted to simplify the design of the next experiment, trying to achieve more stable statistical estimates and increasing the number of data points on which they are based. In the next experiment, we present full size pictures for study, but at test we will compare recognition based on a partial original photograph, in comparison with a full presentation. People in the partial view condition are expected to appeal to their memory retrieval processes to a greater degree than those in the full view condition.

EXPERIMENT 3

Method

There were 56 participants, randomly allocated into two groups of 29 and 27. The method was the same as in the previous two experiments, except for the number of study lists and the test conditions. There were three lists of 60 photographs each in which a set of 30 items were chosen randomly for each participant. In this experiment, each picture was presented for 300 ms three times in each study cycle, and separated from the successive stimulus by a 100-ms black screen. The order of presentation was randomly chosen. There were two conditions at test. In the full picture condition participants were presented with the complete image that they saw at study, whereas in the partial condition people saw only the 20% central part (see Figure 3). The recognition test proceeded as in the previous experiments. In summary, participants studied three lists with 30 stimuli each presented for a total of 90 trials, and received 60 test stimuli with each list (30 studied plus 30 non-studied).

Results and discussion

The recognition scores are summarised in Table 5, which shows the mean proportions of responses, as well as the discrimination indices.

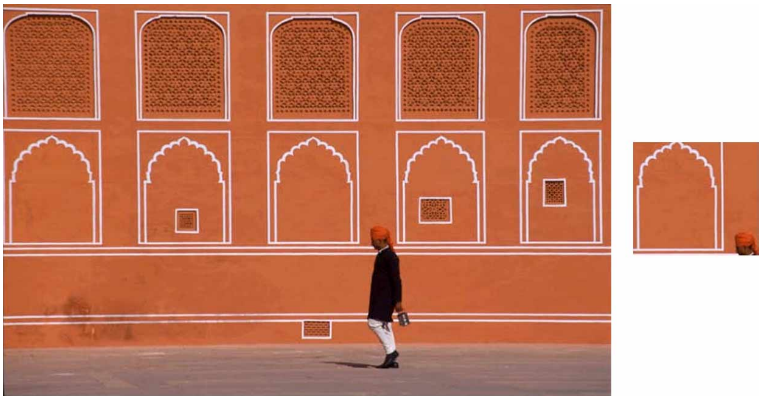

Figure 3. Examples of stimuli presented in Experiment 3: The study item common to all conditions (left), and the complete (left), and partial picture (right) presented at test. Original in colour.

TABLE 5
Mean proportions (and standard errors) for recognition as a function of context condition for Experiment 3

	Remember		*Know*		
Condition	*Hits*	*False Alarms*	*Hits*	*False Alarms*	*d′*
Partial picture	.39 (.02)	.17 (.02)	.28 (.02)	.25 (.02)	0.67 (.05)
Full picture	.44 (.03)	.06 (.01)	.28 (.03)	.12 (.02)	1.52 (.07)

TABLE 6
Statistics (mean and standard error) for the ROCs and *z*-ROCs of Experiment 3

	ROC		*z*-ROC			
Condition	*quadratic constant*	*R2*	*linear slope*	*R2*	*quadratic constant*	*R2*
Partial	−0.50 (0.06)	.997	0.80 (0.03)	.996	0.04 (0.02)	.999
Full	−1.68 (0.22)	.973	0.75 (0.02)	.998	0.01 (0.04)	.998

Discrimination and remember–know data. Full picture recognition led to more hits and fewer false alarms, $t(54) = 2.29$, $p = .026$ versus $t(54) = 7.45$, $p < .01$. No significant difference was found in either remember or know hits, $t(54) = 1.50$, $p = .14$; $t < 1$, respectively. However, recognition based on partial pictures produced more false alarms in remember as well as know judgements, $t(54) = 4.51$, $p < .01$; $t(54) = 4.48$, $p < .01$, respectively. There was a significant difference in d', $t(54) = 10.30$, $p < .01$, indicating logically that full picture recognition was easier.

Remember responses were made with higher confidence than know responses in both partial and full conditions, $t(28) = 8.75$; $t(26) = 5.11$, $p < .01$, respectively. Similarly, the remember false alarms were also made with higher confidence than know false alarms, $t(28) = 6.77$; $t(26) = 4.59$, $p < .01$, respectively, as well as the remember false alarms in relation to know hits, $t(28) = 5.49$, $p < .01$; $t(26) = 2.48$, $p < .05$, respectively. The correlation across participants between the remember hit rate and the remember false alarm rate was significant in the partial condition ($r = .55$, $p < .01$) but not in the full condition (although it was positive, $r = .27$). However there was a positive correlation across participants between know hits and know false alarms considering both conditions together ($r = .32$, $p < .05$).

ROC and z-ROC analyses. The two ROC quadratic constants differed from zero (see Table 6), $t(28) = 8.99$, $p < .01$ (partial picture); $t(26) = 7.69$, $p < .01$ (full picture), indicating that the curves were convex. The *z*-ROC quadratic constant of the partial picture condition was significantly different from zero, $t(28) = 2.10$, $p = .045$, whereas the *z*-ROC quadratic constant of the full picture condition was not ($t < 1$). With regard to the *z*-ROC linear slopes, both of them differed from 1, $t(28) = 6.99$, $p < .01$; $t(26) = 13.91$, $p < .01$, for the partial and full picture condition, respectively.

The *z*-ROCr slopes were 0.96 and 1.10 which indicate that their values may be taken as different from the values of 0.84 and 0.75 estimated from the confidence data, for the partial and full picture conditions, respectively.

GENERAL DISCUSSION

We designed the three experiments that we report here with the idea of promoting differentially the utilisation of contextual and/or item evidence in the test phase of a recognition task. Our assumption was that the use of both is not well balanced in the typical recognition experiment published in the literature, and is dependent on the experimental conditions. Therefore this was a prerequisite to any further consideration of the virtues of the remember–know methodology as a tool to shed light on the one- versus two-process view of recognition. The first two experiments had as a control a condition typically run in a standard recognition experiment: just a series of pictorial elements presented routinely for study. We expected that participants would have little contextual evidence at test in comparison with the remaining conditions. On the other hand, the congruent and incongruent manipulations had additional ''contextual'' information to facilitate or interfere with the task of identifying the correct stimuli. The third experiment tried to achieve the same goal but this time presenting a partial or full stimulus at test. In this case, and beyond the differences in difficulty of both conditions, we expected that participants would use recollection more extensively in the partial condition.

In the first two experiments, the raw data indicated as expected that, in general, the level of hits was greater in those conditions promoting recollection of information positively related to the target, but interfered (greater false alarms) in those positively related to the lures (the incongruent conditions). In the third experiment, given that the ''partial'' condition was more difficult, the percentage of hits was lower and the false alarms greater than in the full picture condition. Whereas in the first two experiments the information provided by the item at test remained constant across conditions and recollection varied accordingly, in the third experiment we varied both dimensions orthogonally. That is, when people saw the whole picture at test, they had maximum item and minimum recollective evidence, whereas the opposite was true in the partial view test.

Overall, the data suggest that people search their memory for specific evidence related to the current test image, trying to retrieve the rest of the picture or any other contextual evidence to confirm the test item as studied or not. The strategy produced good relative results in the congruent groups because it increased the proportion of hits and led to a small and non-significant increase in false alarms (perceived as ''know'') in comparison with the control group. However, it produced the opposite results in the incongruent group, increasing false alarms in the remember and know response categories. The interesting aspect of the data on the incongruent group is that the participants explicitly and implicitly falsely believed that some new items never presented had been studied (remember false alarms and know false alarms), respectively. As part of this strategy, people in the congruent group falsely accepted a given test item, as their ''know'' false alarm rates were also slightly greater than the controls. In the present case, the old and new lists are thematically matched, and this factor may help to explain the slight increase in false alarms in the congruent group. These results are closely related to those obtained in false recall research using words (Roediger & McDermott, 1995) or pictures (Miller & Gazzaniga, 1998) as stimuli. People are induced to believe that new items have been presented on the basis of their associative relation to previously studied ones. Here we have a stronger case of false recognition due to the multiple item and context presentations and their very strong connections. We believe, in agreement with Geraci and McCabe (2006), that the source of this increase in false alarms is not related to a specific processing of the lures, but to the recollection of contextual elements presented at study.

Recent experiments by Davenport and Potter (2004) also support the previous interpretations. They used natural scenes in which a background and a foreground object were defined in a consistent or inconsistent manner (e.g., a football player or a priest against a stadium as a background). The results showed that after very brief presentations (80 ms) people recognised the foreground object more accurately when the context was congruent (0.82 versus 0.68), but the presentation of the isolated object was more accurate than its combination with a background of any type. When asked to recognise both foreground and background, the latter suffered and was less accurate. Besides the parallelism, this experiment points towards the predominant role that the item has in any recognition experiment, and the small role played by the recollective evidence (see also Dobbins et al., 1998).

However, the first question of interest, given the previous raw results, is whether those data give any clues on the involvement of one or two processes in recognition. Following the recent arguments by Wixted and Stretch (2004) we will discuss our results taking into account the relations between subjective responses (remember–know) and confidence, then the mirror effect, and finally the receiver-operating characteristic analysis.

Our experiments indicate that, across participants, confidence is greater for remember than know responses, and for remember false alarms than for know and know false alarms. Additionally, there was a positive correlation between the remember false alarm and know false rates across the seven conditions ($r = .76$, $p < .05$). The signal detection view of the remember–know task interprets these relations as the result of criterion changes, and considers that the critical data for past two-process interpretations has to do with the significance of remember false alarms. In most experiments, false alarms are discarded because of their insignificance, or are considered as guesses (as in Yonelinas's model). How is it possible to explain from this point of view that remember false alarms are associated with greater confidence than know hits? Obviously, if false alarms are equated with guesses this is not possible. Our view is that remember false alarms are false recollections in which the important aspect is more the recollection (illusory, see Geraci & McCabe, 2006) of the study context, than the processing of the lure itself. In our experiments the increase in remember false alarms as a function of the presentation of an ''incongruent'' context (Experiments 1 & 2) fits this explanation. Further support is in the fact that the only condition in which remember false alarms are not associated with greater confidence than know hits is the ''absent'' condition. This ''absent'' condition is defined as that in which the item itself carries poor information (the annulae) and no explicit ''context'' was presented.

In these experiments, as in many others in which different discriminability conditions are studied, a mirror effect is observed. That is, the combination of higher hits and lower false alarm rates in higher-discriminability in relation to lower-discriminability conditions. But of more interest is the possible presence of a remember mirror effect. In terms of correlations, the remember mirror effect is observed as a negative correlation across conditions between remember hits and false alarms rates ($r = -.73$, $p = .06$ for the conditions of the three experiments reported here). The interpretation of the mirror effect from the point of view of the signal detection approach is linked to changes in criteria across participants. When in an experiment liberal and conservative participants are measured, a positive correlation between hits and false alarms is observed, but across conditions the opposite is obtained because increases in d' (changes in the target distribution) are followed by changes in the old/new criterion, leading to a simultaneous increase in hits and decrease in false alarms.

Wixted and Stretch (2004) proposed a signal detection explanation of the remember mirror effect based in the simultaneous shift of the know and remember criterion (see their figure 8, p. 633). If remember responses are simply high-confidence responses, a remember mirror effect across conditions should be observed. Our data show a remember mirror effect when a weak condition is compared with a stronger one. However, we would also expect a decrease in know hits and know false alarms if a shift in criterion is the correct explanation of the remember mirror effect. Although it is not a necessity to assume that the know and remember criteria move in a correlated fashion, from this assumption follows naturally the simultaneous prediction of a remember mirror effect and the decrease of know hits and know false alarms when two conditions of different strengths are compared (positive correlation between know hits and false alarms). We do not observe this result in our data: a remember mirror effect but no decrease in know hits, and/or increase in know false alarms comparing the incongruent and absent, absent and congruent (Experiment 1), and the partial versus full picture (Experiment 3). The main reason for this discrepancy seems to lie in the interpretation of false alarms. If we think that the study episode gives place to an increase in false recollective evidence and/or familiarity, depending on conditions, on the lures then all these results can be perfectly assimilated by a dual model. Similar explanations have been advanced by Diana et al. (2006) and by Geraci and McCabe (2006) in relation to data obtained with different paradigms.

With regard to the ROC analysis, the z-ROC quadratic constants calculated in the first experiment showed that the ''no context'' and the ''incongruent'' context were significantly different from zero, whereas the ''congruent'' one was not. This difference seems paradoxical. However, the ''no context'' condition is similar to most recognition experiments in which participants are able to generate a mental context for each presented stimulus. That is, people can form their own images and connections, or gather other evidence and recollect them later. We were able to replicate these results in the second experiment, after lengthening the interstimulus interval to preclude the possible unitisation of context and item. This conclusion was more firmly established in the third experiment.

Previous experiments in item recognition using similar analysis (Glanzer et al., 1999; Hilford et al., 2002) have not found curvilinear z-ROCs as predicted by dual models, although the evidence is more arguable in associative recognition (e.g., see Kelley & Wixted, 2001; Yonelinas, 1997). In source recognition, experiments also show a mixture of results (Quamme, Frederick, Kroll, Yonelinas, & Dobbins, 2002; Hilford et al., 2002) that have led to a modification of the unequal variance signal detection model to account for the presence of curvilinear z-ROCs (but see Slotnick & Dodson, 2004). It seems obvious to us that there are situations or individual tests in which only familiarity underlies the response, and others in which there is a mixture of familiarity and recollection. The case where recollection plays a major role is less frequent in the experimental literature. In our experiments, all constants obtained by fitting a quadratic equation to the ROC data differ from zero and are negative, except for the incongruent condition of Experiment 2, explainable because of the low accuracy. This pattern of data indicates normality of the underlying distributions of new and old items (Hilford et al., 2002).

The z-ROC slopes showed a clear pattern with values in the vicinity of .80, and approaching 1 as a function of accuracy, as expected from the previous literature (see also Glanzer, Adams, Iverson, & Kim, 1993; Glanzer et al., 1999; Heathcote, 2003; Hilford et al., 2002; Rotello et al., 2004). As the accuracy decreased, so the z-ROC slope approached 1. However in the third experiment, even when accuracy was quite low, still the z-slope was lower than 1. It is apparent that the study episode is affecting the variances of the old distribution, or the new one (in the incongruent conditions). However, it is also apparent that in some of the conditions of the three experiments, recollection was absent, and still the z-ROC slopes were considerable lower than 1, and the z-ROC functions were linear, an unacceptable combination for the Yonelinas model. It has been remarked elsewhere (Heathcote, 2003) that the Yonelinas model could solve this problem allowing unequal variances in the signal detection component of the theory. However, given the continuity of remember and know judgements (e.g., Perfect et al., 1996) we think that a different dual model without the threshold restriction would be more reasonable. Finally, the slopes computed on the basis of the remember confidence data can be taken as different from those computed from the global confidence judgements in the absent (Experiment 1), incongruent (Experiment 2), and in the two conditions of Experiment 3, possibly. Moreover, the correlation across conditions for the three experiments between both sets of slopes is 0.43 ($p > .05$) and it is considerably lower than the values calculated by Wixted and Stretch (2004). This set of analyses provides a different picture from that shown by Wixted and Stretch (2004, Table 3, p. 626) in their analysis of the published literature, indicating a greater discrepancy between both set of estimations.

In the three experiments we have described, we have forced participants to recollect evidence and explicitly use it in recognition. These experiments were run under what Rotello et al. (2005) call ‘‘standard instructions’’. We have reached an opposite conclusion to theirs in that, when given appropriate opportunities to recollect, ROC parameters and R/K judgements basically agree (see also Bastin & Van der Linden, 2003; Prull, Dawes, Martin III, Rossenberg, & Light, 2006). In the two first experiments, the congruent context is helping to achieve greater hit rates than the controls at a slight cost of incrementing false alarms. This increment in false alarms is what to a greater extent is happening in the incongruent condition in which there is a strong cost in hits (decrease) and false alarms (increase). The third experiment, from the point of view of overall performance, shows that the partial image presented at test made recognition much more difficult. Nevertheless, recollection helps to take a decision, as confirmed by the roc analysis.

Although recent papers (e.g., Rotello et al., 2005) have questioned the validity of the remember–know methodology in contrast to previous claims by its proponents (Gardiner & Richardson-Klavehn, 2000), the overall conclusion we obtain from our experiments is mostly positive. There is little doubt that the procedure may be subject to bias, meaning clearly that it is not process pure, but given the appropriate conditions it is responsive to the influence of contextual variables of experimental interest and in ways not so obvious from the objective performance. For example, the fact that people ‘‘remember’’ at times and at some others falsely ‘‘know’’ an item as presented is something only foreseen when this type of response is taken into account, in a similar way to when people are not able to verbalise any contingency but an effect on recognition is found (Manier, Apetroiaia, Pappas, & Hirst, 2004).

In conclusion, our experiments have been designed to allow measurement of recognition in a variety of changing conditions, requiring different levels of recollection of the study episode. We have collected confidence and remember–know responses that have been analysed thoroughly to discern whether one or two process views help to explain them better. We think that no single position is able to account uniquely for all aspects of the results. In fact, an explanation can be deduced from each position that is able to deal with most of the data. However, we tend to favour a two-process explanation if we add to the behavioural analysis the conclusions we can extract looking at the neurophysiological and neuropsychological literature. We mentioned in the introduction that the neurophysiological literature is more in agreement with the idea that there are two brain areas involved in recognition (for a recent review see Aggleton & Brown, 2006). The event-related potentials studies (ERP) and the neuroimaging literature (see a relevant exception to this general conclusion in Gold et al., 2006) seem to indicate that the hippocampus is mainly involved in recollection, whereas several surrounding areas like the perirhinal and other parahippocampal structures are possibly involved in familiarity, in addition to the involvement of the prefrontal cortex. In this area of research, the remember–know paradigm is routinely used to index familiarity and recollection and linked to different pattern of neural activity in the brain (e.g., Duarte, Ranganath, Winward, Hayward, & Knight, 2004). Finally, the neuropsychological and ageing literature also provides support to the two-process position. Older people (Bastin & Van der Linden, 2003; Howard, Bessette-Symons, Zhang, & Hoyer, 2006) show impairment in more recollective-oriented recognition tests (yes–no tests) than in familiarity (forced choice). Finally, patients diagnosed with mild cognitive impairment, a prodromal stage of Alzheimer's disease, show intact or very mild impairment in familiarity estimates but strong deficits in recollection, although Alzheimer patients show strong deficits in both components (Westenberg, Paller, Holdstock, Mayes, & Reber, 2006). In conclusion, we think that, globally considered, the evidence favours a dual-process view of recognition, in which both processes can be behaviourally and neurophysiologically characterised, which the human cognitive system uses differently according to encoding and testing conditions.

Manuscript received 19 June 2006
Manuscript accepted 2 March 2007
First published online 26 April 2007

REFERENCES

Aggleton, J. P., & Brown, M. W. (2006). Interleaving brain systems for episodic and recognition memory. *Trends in Cognitive Sciences*, *10*, 455–463.

Algarabel, S., Pitarque, A., & Gotor, A. (2006). Effect of the retention interval on the simultaneous cognate–noncognate and remember–know mirror effects. *Memory*, *14*, 79–86.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol 1. Learning, memory & thinking* (pp. 243–293). San Francisco: Freeman.

Baddeley, A. D. (1982). Domains of recollection. *Psychological Review*, *89*, 708–729.

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology*, *17*, 14–24.

Boldini, A., Russo, R., & Avons, S. E. (2004). One process is not enough! A speed–accuracy tradeoff study of recognition memory. *Psychonomic Bulletin & Review*, *11*, 353–361.

Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & Cognition*, *22*, 273–280.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*, 559–564.

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

Dobbins, I. G., Kroll, N. E. A., & Liu, Q. (1998). Confidence–accuracy inversions in scene recognition: A remember–know analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1306–1315.

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

Duarte, A., Ranganath, C., Winward, L., Hayward, D., & Knight, R. T. (2004). Dissociable neural correlates for familiarity and recollection during the encoding and retrieval of pictures. *Cognitive Brain Research*, *18*, 255–272.

Dudukovic, N. M., & Knowlton, B. J. (2006). Remember–know judgements and retrieval of contextual details. *Acta Psychologica*, *122*, 160–173.

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

Gardiner, J., & Java, R. I. (1993). Recollective experience in word and nonword recognition. *Memory & Cognition*, *18*, 23–30.

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

Geraci, L., & McCabe, D. P. (2006). Examining the basis for illusory recollection: The role of remember/know instructions. *Psychonomic Bulletin & Review*, *13*, 466–473.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1–67.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*, 546–567.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 500–513.

Gold, J. J., Smith, C. N., Bayley, P. J., Srager, Y., Brewer, J. B., Stark, G. E. L., et al. (2006). Item memory, source memory, and the medial temporal lobe: Concordant findings from fMRI and memory impaired patients. *Proceedings of the National Academy of Sciences*, *103*, 9351–9356.

Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1210–1230.

Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *28*, 503–508.

Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L. T. (2002). Regularities of source recognition: ROC Analysis. *Journal of Experimental Psychology: General*, *131*, 494–510.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behaviour Research Methods, Instruments and Computers*, *16*, 96–101.

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

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember–know paradigm. *Memory & Cognition*, *25*, 345–351.

Howard, M. W., Bessette-Symons, B., Zhang, Y., & Hoyer, W. J. (2006). Aging selectively impairs recollection in recognition memory for pictures: Evidence from modeling and receiver operating characteristic curves. *Psychology and Aging*, *21*, 96–106.

Inoue, C., & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgements. *Memory & Cognition*, *26*, 299–308.

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

Java, R. I., Gregg, V. H., & Gardiner, J. M. (1997). What do people actually remember (and know) in "Remember/know" experiments? *European Journal of Cognitive Psychology*, *9*, 187–197.

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 701–722.

Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 153–161.

Mandler, G (1980). Recognizing: The judgement of previous occurrence. *Psychological Review*, *87*, 252–271.

Manier, D., Apetroaia, I., Pappas, Z., & Hirst, W. (2004). Implicit contributions of context to recognition. *Consciousness and Cognition*, *13*, 471–483.

Miller, M. M., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia*, *36*, 513–520.

Murdock, B. B. Jr. (1974). *Human memory: Theory and data*. Potomac, MD: Lawrence Erlbaum Associates Inc.

Murdock, B. B. Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609–626.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*, 403–415.

Perfect, T. J., Mayes, A. R., Downes, J. J., & Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *Quarterly Journal of Experimental Psychology*, *49A*, 797–813.

Pike, R. (1984). A comparison of convolution and matrix distributed memory systems. *Psychological Review*, *91*, 281–294.

Prull, M. W., Dawes, L. L. C., Martin, A. M. III, Rossenberg, H. F., & Light, L. L. (2006). Recollection and familiarity in recognition memory: Adult age differences and neuropsychological test correlates. *Psychology and Aging*, *21*, 107–118.

Quamme, J. R., Frederick, C., Kroll, N. E. A., Yonelinas, A. P., & Dobbins, I. G. (2002). Recognition memory for source and occurrence: The importance of recollection. *Memory & Cognition*, *30*, 893–907.

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes on picture recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 365–377.

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*, 518–535.

Reder, L. M., Donavos, D. K., & Erickson, M. A. (2002). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition*, *30*, 312–323.

Roediger, H. L. III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.

Rotello, C. M., & Macmillan, N. A. (2006). Remember–know models as decision strategies in two experimental paradigms. *Journal of Memory and Language*, *55*, 479–494.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal detection model. *Psychological Review*, *111*, 588–616.

Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review*, *12*, 865–873.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburg: Psychology Software Tools Inc.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166.

Slotnick, S. D., & Dodson, C. S. (2004, November). *The natural shape of the source memory ROC supports a continuous (single-process) model of memory*. Paper presented at the 45th meeting of the Psychonomic Society, Minneapolis, Minnesota.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1–17.

Westenberg, C. E., Paller, K. E., Holdstock, J. S., Mayes, A. R., & Reber, P. J. (2006). When memory does not fail: Familiarity-based recognition in mild cognitive impairment and Alzheimer's disease. *Neuropsychology*, *20*, 193–205.

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

Xu, M., & Bellezza, F. S. (2001). A comparison of the multimemory and detection theories of know and remember recognition judgements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1197–1210.

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

Yonelinas, A. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, *25*, 747–763.

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

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition*, *5*, 418–441.

APPENDIX

Frequency of responses per confidence (from sure old, 1, to sure new, 6) and experimental condition for Experiment 1

	Confidence					
Context	*1*	*2*	*3*	*4*	*5*	*6*
Absent						
Old	826	285	201	167	355	86
New	48	125	161	304	767	515
Congruent						
Old	1053	298	196	140	180	52
New	54	138	207	347	663	511
Incongruent						
Old	623	351	219	248	395	84
New	202	266	245	287	624	294

Frequency of responses per confidence (from sure old, 1, to sure new, 6) and experimental condition for Experiment 2

	Confidence					
Context	*1*	*2*	*3*	*4*	*5*	*6*
Congruent						
Old	876	315	216	183	247	82
New	40	153	202	425	680	419
Incongruent						
Old	522	275	254	287	427	153
New	165	242	290	324	590	309

Frequency of responses per confidence (from sure old, 1, to sure new, 6) and experimental condition for Experiment 3

	Confidence					
Condition	*1*	*2*	*3*	*4*	*5*	*6*
Partial						
Old	783	441	506	466	356	58
New	195	342	549	755	611	158
Full						
Old	1115	351	286	235	303	140
New	100	136	223	409	790	772