Does the butcher-on-the-bus phenomenon require a dual-process explanation? A signal detection analysis

Richard J. Tunney*, Timothy L. Mullett, Claudia J. Moross and Anna Gardner

School of Psychology, University of Nottingham, Nottingham, UK

The butcher-on-the-bus is a rhetorical device or hypothetical phenomenon that is often used to illustrate how recognition decisions can be based on different memory processes (Mandler, 1980). The phenomenon describes a scenario in which a person is recognized but the recognition is accompanied by a sense of familiarity or knowing characterized by an absence of contextual details such as the person's identity. We report two recognition memory experiments that use signal detection analyses to determine whether this phenomenon is evidence for a recollection plus familiarity model of recognition or is better explained by a univariate signal detection model. We conclude that there is an interaction between confidence estimates and remember-know judgments which is not explained fully by either single-process signal detection or traditional dual-process models.

Keywords: episodic memory, recognition, signal detection, context, faces

INTRODUCTION

The *butcher-on-the-bus* is a rhetorical device or hypothetical phenomenon that is often used to illustrate how recognition decisions can be based on different memory processes (Mandler, 1980). The phenomenon describes a scenario in which a person is recognized but the recognition is accompanied by a sense of familiarity or knowing characterized by an absence of contextual details such as the person's identity as a butcher. A great many studies have examined how context facilitates recognition, and a great many studies have examined the subjective experience of remembering (Gardiner and Richardson-Klavehn, 2000) and whether or not this provides evidence that recognition is composed of multiple processes (Dunn, 2004). A few studies have examined the subjective experience of remembering in face recognition (Brandt et al., 2003). A few studies have even considered whether context affects the subjective experience of recognition in different ways and thus dissociates recollection and familiarity (Macken, 2002), but only one previous study has done so in face recognition (Gruppuso et al., 2007). That is, to our knowledge there is only one published experimental report that the butcher-on-the-bus as a rhetorical device might actually exist as an experimental phenomenon. Although this report did demonstrate that context influences the subjective experience of recognizing a face the results were equivocal in the sense that a number of different models of recognition memory could explain the data, including single-process models. In the first of the two experiments reported here we aim to first replicate the effect of context on the face recognition and the associated reports of remembering. In order to discriminate between the single and dual-process explanations of this effect we apply a signal detection analysis that was not used by Gruppuso et al. (2007). In the second experiment we test the claim in the recognition memory literature that the two proposed memory processes differ in that recollection encodes context but familiarity does not. Here we provide data that suggests that the butcher-on-the-bus phenomenon may well be indicative of two underlying memory

systems, but that this can only be explained by two continuous underlying signals, not by threshold models of recollection.

Tulving (1985) argued that if different memory processes give rise to different phenomenal experiences it is reasonable that the relative contribution of those memory processes to a decision could be estimated by simply asking people to report their subjective experience of remembering. A large literature subsequently developed that showed dissociations between estimates of recollection and familiarity based on the proportions of recognition judgments (and recalled items) that participants reported were accompanied by the experience of remembering or of knowing. For example, low frequency words are more likely to be recognized than high frequency words and this is more likely to be accompanied as an experience of recollection (Gregg et al., 2006). A similar effect is observed in face recognition when distinctive faces elicit more remember than know responses (Brandt et al., 2003). That is, we are more likely to recollect a distinctive face than an indistinct one that may feel familiar. Unsurprisingly the effect is also true of distinctive forenames (Brandt et al., 2006).

The crux of the butcher-on-the-bus as a rhetorical device lies in the predicted effect of context on the subjective experience of remembering and by inference on recollection and familiarity. Context effects have typically been studied in word recognition. Words presented in the same or similar context as the study episode are more likely to be recognized or recalled than words tested in a novel context (Godden and Baddeley, 1980; Rutherford, 2004). Moreover, word recognition in a context different to that of the study episode reduces the contribution of recollection to the recognition judgments (Macken, 2002).

Of course context can take a number of different forms (McGeoch, 1932), so in order to control experimentally the association between face and context a number of researchers have turned to associative recognition as the paradigm of choice. In an early study Watkins et al. (1976) presented pairs of faces at study and at test. Recognition accuracy was reduced at test when

the target faces were presented with a novel context face (see also Winograd and Rivers-Bukeley, 1977). Watkins et al. observed the same effect when faces were paired with brief personal descriptions of the person. Similar effects are also observed by changing the backgrounds behind the faces (Davies and Milne, 1982). A striking example of this was reported by Rainis (2001) who found that emotionally arousing contexts can lead to a reduction in accuracy for negative contexts (e.g., a concentration camp) or an increase in accuracy for positive contexts (e.g., paradise island). Some studies have even shown differences in event related potentials (ERPs) when faces are recognized and their paired contextual details are also retrieved compared to when faces are recognized but the contextual details are not retrieved (Yovel and Paller, 2004). These different ERP signals are indicative of different retrieval processes but do not necessarily imply differences in the subjective experience of retrieval or indeed underlying processes.

To date however only one study has directly tested whether context affects the subjective experience of remembering in face recognition (Gruppuso et al., 2007). In this experiment participants saw a series of faces paired with scenes. At test the target faces were presented with either the same context scene as at study, a switched scene, or a novel scene. The distractor faces were presented with either novel or old scenes. As expected recognition accuracy was more reliable when old faces were presented at test with the same context scene as at study, relative to when the context scene was switched or new. The key question is how context affected the subjective experience that accompanied the recognition judgments. Gruppuso et al. reported that the recollection-based memory was more accurate when the context was the same compared to when it was switched or new, but there was no such effect on familiarity-based memory. These results do indeed suggest that context facilitates face recognition and influences the subjective experience associated with recognition. The corollary of this effect is that if a face is recollected then the context in which the face was originally stored in memory (i.e., the source) should also be available for retrieval, but if a face merely feels familiar then source memory should be less accurate or absent altogether.

According to most dual-process models recollection is distinct from familiarity not only in terms of phenomenology but also in that it encodes context. Indeed studies that show that a reinstatement of study context at test facilitates memory do so presumably because in those models the context cues retrieval. Alternatively a signal detection model might assume that this merely increases the signal strength. An alternative way to conceptualize the butcher-on-the-bus phenomenon is that it is an example of strong familiarity-based recognition in the absence of the retrieval of source information, namely context.

In a series of five word recognition studies Perfect et al. (1996) examined whether remember responses are associated with better source memory than know responses. The source contexts included temporal order, list identity, spatial location, and visual form (i.e., font and font size). In only one of these experiments was source memory reliably greater than chance for items reported as know responses. More recently Dudukovic and Knowlton (2006) used a paired-associate procedure. After a 10-min retention interval there was a recognition test for one of the words in each pair and participants were asked to report their experience of remembering

for each decision. The participants were re-tested after a 7-day retention interval and asked to report contextual details such as the location of the word in the pair or the color of the image. Remember responses were associated with the retrieval of contextual details but know responses were not. In a later study that used the same materials the participants were able to indicate the color of the pictures that accompanied the target words and whether the target had appeared as the left-hand or the right-hand item of the study pair (Eldridge et al., 2005). Moreover, although responses were more accurate for remember responses, accuracy was well above chance for both remembering and knowing.

Wais et al. (2008) report an experiment in which participants studied nouns presented on screen either in blue or in red, and either above or below the center of the computer screen. They also reported that context accuracy for know responses was greater than chance suggesting that familiarity is not distinguished from recollection in terms of its encoding of context (see also Wixted and Mickes, 2010). Although we know of no study that has examined whether context or source memory for faces can distinguish between recollection and familiarity a recent series of studies Bell and Buchner (Buchner et al., 2009; Bell and Buchner, 2010, 2011) have shown that source memory (i.e., context information), but not recognition accuracy for faces is influenced by their emotional valence. Unfortunately these studies did not determine whether source memory occurred only for remember judgments or for know judgments as well.

Signal detection theory provides an elegant model of memory and also provides analytic techniques to determine if a decision is based on more than one source of information. That is, signal detection theory can be used to estimate the relative contributions of recollection and familiarity to recognition memory independently of subjective reports of remembering (Yonelinas, 1994); and can also be used as an alternative model that does not require two underlying memory processes (Donaldson, 1996). Signal detection theory is therefore an ideal paradigm to test the assumption that the butcher-on-the-bus phenomenon, should it exist as a real laboratory phenomenon and not merely a rhetorical device, really does discriminate between single and dual-process models of recognition memory. To test the signal detection model the participants are asked to report how confident they are in each decision. These confidence ratings are then used to plot receiver operating characteristics (ROC) curves. The coordinates of the ROC curves can be transformed into z-scores, and a regression line fitted. The regression line describes the form of the zROC and this discriminates between a dual-process account of recognition and a single-process account. For instance, a zROC with a regression line with a slope close to 1 implies that the recognition judgments were based on a single underlying dimension in which the signal and noise distributions had similar variance. If, as is often the case, the slope of the regression line deviates from 1 then this implies that the ROC is asymmetric, and that the signal distribution has a narrower variance than the noise distribution. However, this only implies that the signal distribution has more than one component if there is also a quadratic component to the regression line (i.e., two slopes). Thus an independent test of whether context facilitates recollection in face recognition is to look for a larger quadratic component in the zROC curves when

faces are presented with the same context as the study episode compared to when they are presented with different or new contexts (for a fuller description see Tunney and Bezzina, 2007; Tunney, 2010).

Wixted and Mickes (2010) have recently proposed an extension to the signal detection model adding recollection as an additional orthogonal process, but one which is also based upon a continuous signal detection scale. Importantly, it assumes that confidence in old-new judgments is always predicated upon the sum of the two signals. This means it maintains the same predictions as the univariate signal detection model described above but predicts a different mechanism for remember and know judgments. The model assumes that a remember response is made when the signal strength from the recollection process passes a criterion point. This means that the model still accounts for instances where the summative signal is high, but the recollection signal alone is low, or at least not high enough to pass the remember criterion. These instances result in high confidence familiarity responses: the classic butcher-on-the-bus phenomenon.

In the experiments that follow we explore the butcher on the bus and ask whether it is truly a phenomenon that discriminates between models of recognition memory. Experiment 1 replicates the procedure reported by Gruppuso et al. (2007) that demonstrated context effects on the subjective experience of recognizing faces. The only adjustment to the paradigm that is required to use SDT as a test of the dual-process account is to ask the participants to report how confident they are in each recognition decision. In Experiment 2 we ask whether recollection and familiarity are characterized by differences in the retrieval of contextual (source) details.

EXPERIMENT 1
MATERIALS AND METHODS
Participants
Forty-nine members of the University of Nottingham community volunteered to take part in this experiment in return for course credit. Their mean age was 21 years (SD = 0.64). Thirty-one were female and 18 were male. All had normal or corrected to normal vision.

Stimuli
The stimuli consisted of color photographs of face-scene pairs (see **Figure 1**). The faces were 96 naturalistic portrait photographs of different people collected by an Internet search. Half of these were females and half were males. These were cropped to exclude as much contextual information such as clothing and background locations as possible. The estimated age of the faces were 42 years old (SD = 13.56) and for 39 years (SD = 11.50) for males and females, respectively. The oldest were around 65 years and the youngest 20 years. The majority (81%) were judged to be of European-origin, 7% Asian-, 6% Chinese-, and 5% African-origin. Ninety-six photographs of a variety of scenes such as landscapes (43), building interiors (22), and exteriors (23) were also collected using the Internet. The remaining eight scenes such as a tennis court or market stall didn't fall into any obvious category. Each face-context pair was then randomly created. There were no explicit exclusion or inclusion criteria for the faces other than that the photographs were of a sufficient resolution. Similarly there
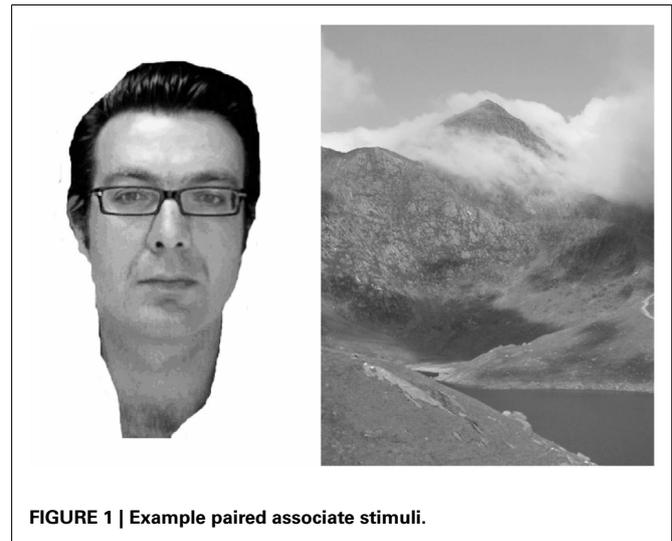
FIGURE 1 | Example paired associate stimuli.

were no criteria with respect to the scenes other than that they were unfamiliar in the sense that although they could be named by super-ordinate category (e.g., mountain), they could not be named as a specific instance by the experimenters (e.g., Snowdon).

Forty-eight faces were randomly selected as study items. Half of these were male and half were female. These were paired with a unique scene. There were six test conditions: the stimuli in the *old face–old scene* (OO) condition consisted of 12 of the study item. In the *old face–switched scene* (OS) a different set of 12 faces from the study list were paired with a set of 12 scenes that had appeared in the study list but had been paired with a different face. The *old face–new scene* (ON) condition included the remaining 24 faces from the study list paired with a scene that had not appeared in the study list. Forty-eight foils were created by pairing faces that had not appeared in the study list with either a scene that had appeared in the study list (*new face–old scene*, NS) or one that had not (*new face–new scene*, NN).

Procedure
The participants were told they would be shown a series of face-scene pairs and that they would be asked to rate how associated they believed the face-scene pair to be on a six-point scale ranging from strongly unassociated to strongly associated. The participants were not informed that they would have to recognize the faces in the test that followed. There was a 15-min retention interval between the study and test periods. Before the test phase began, the participants were told they would be shown another set of face-scene pairs, and that they are to make judgments only on whether they recognize the faces, not the scenes. The recognition judgments were based on a six-point scale ranging from *sure-new* to *sure-old*. Whenever the participants responded with one of the three old buttons they were then asked also to indicate whether their recognition decision was based on recollection or familiarity by clicking buttons marked *remember* or *know*, respectively. The order in which the test items appeared was randomized. To ensure that the participants understood the distinction between remember and know responses we used a modified version of the "standard instructions" (Gardiner and Richardson-Klavehn, 2000) as follows:

"In this part of the experiment you will see some more faces and scenes. Some of the faces and scenes have already appeared in the study part of the experiment. You are now asked to make judgments on whether you have seen the face before. You are to make judgments about the faces only and not the scene. On the screen below you will see six buttons marked *sure-old, fairly sure-old, guess-old, guess-new, fairly sure-new*, and *sure-new*. Please click one of the buttons to indicate whether you think that you have seen the face during the study part of the experiment and how confident you are in that judgment. Recognition memory is associated with two different kinds of awareness. Quite often recognition brings back to mind something you recollect about what it is that you recognize, as when, for example, you recognize someone's face, and perhaps remember talking to this person at a party the previous night. At other times recognition brings nothing back to mind about what it is you recognize, as when, for example, you are confident that you know you recognize them, because of strong feelings of familiarity, but you have no recollection of seeing this person before. You do not remember anything about them. These kinds of awareness are associated with recognizing the faces you saw earlier. Sometimes when you recognize one of the faces in the experiment, recognition will bring back to mind something you remember thinking about when the face appeared then. You recollect something you consciously experienced at the time. But sometimes recognizing a face will not bring back to mind anything you remember about seeing it then. Instead the face will seem familiar, so that you feel confident it was one that you have seen before, even though you don't recollect anything you experienced when you saw it then. For each face that you recognize, you will be asked to indicate your experience of remembering. Please then click the *remember* button, if recognition is accompanied by some recollective experience, or the *know* button, if recognition is accompanied by strong feelings of familiarity in the absence of any recollective experience. Click OK when you are ready to proceed."

RESULTS

The proportions of items in each condition endorsed as old are shown in **Figure 2**. Endorsements to old (hits) and to new items (false alarms) were entered into separate repeated measures ANOVAs. There was a main effect of context on hits, $F(2, 96) = 39.36$, MSE < 0.01, $p < 0.01$, $\eta_p^2 = 0.45$, due to an increase in endorsements to OO items compared to *OS* items, $F(1, 48) = 65.27$, MSE < 0.01, $p < 0.01$, $\eta_p^2 = 0.58$. There was no difference in endorsements to *OS* compared with *ON* items, $F(1, 48) = 0.11$, MSE < 0.01, $p = 0.75$, $\eta_p^2 < 0.01$. There was no effect of context on false alarms, $F(2, 96) = 0.37$, MSE < 0.01, $p = 0.55$, $\eta_p^2 < 0.01$. These data clearly show that face recognition is more accurate when the test items are presented with the same context as the study period. However, there appeared to be no increase in recognition accuracy when the context was switched compared to when the context was new.
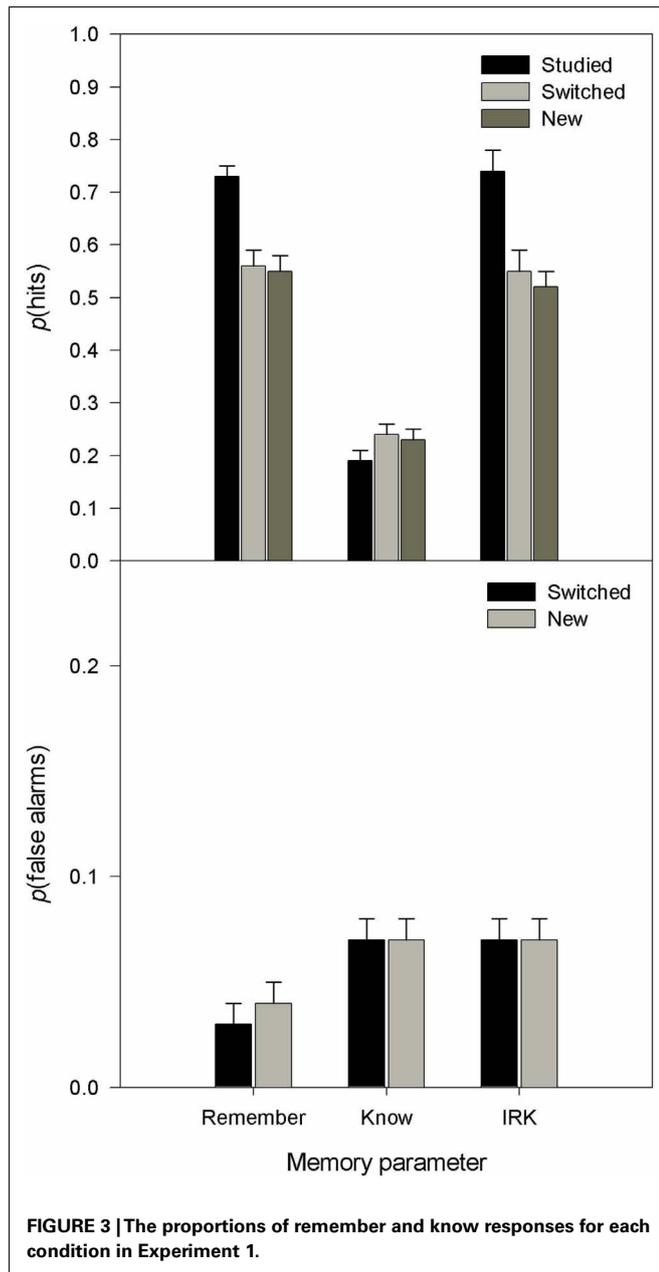
THE SUBJECTIVE EXPERIENCE OF REMEMBERING

Does context affect the subjective experience of remembering? The proportions of remember and know responses are shown in

FIGURE 2 | The proportions of items in each condition endorsed as old in Experiment 1.
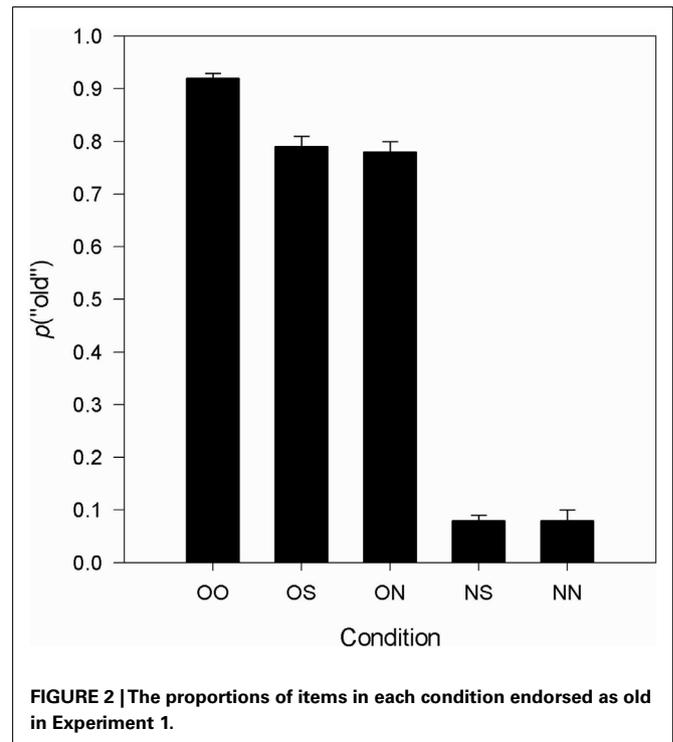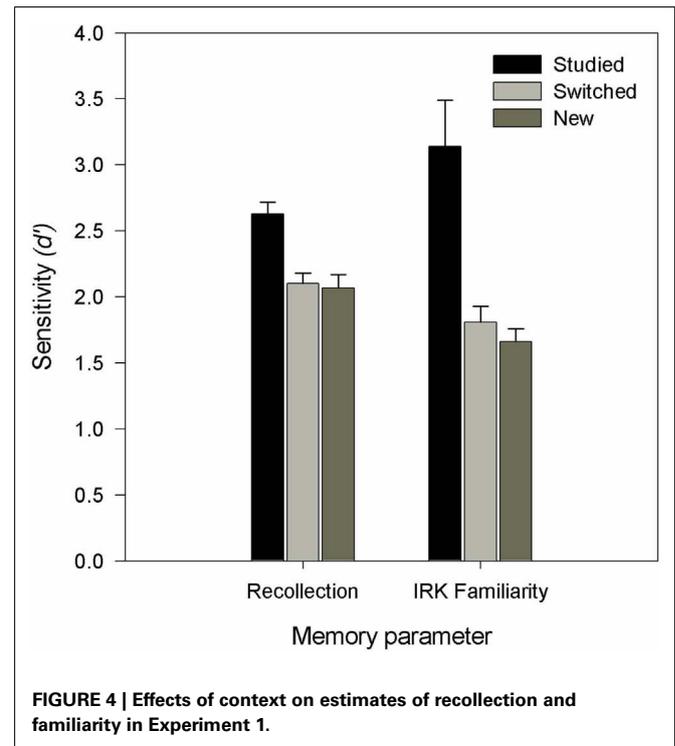
Figure 3 for each condition. The upper panel shows hits and the lower panel shows false alarms. Also shown are familiarity estimates made using the independence assumption (Jacoby et al., 1997). These data were entered into separate ANOVAs. There was a main effect of context on the proportion of remember responses to studied faces, $F(2, 96) = 49.15$, MSE $= 0.01$, $p < 0.01$, $\eta_p^2 = 0.51$. This was because old faces with their studied context scene elicited more remember responses than old faces with a switched context, $F(1, 48) = 75.46$, MSE < 0.01, $p < 0.01$, $\eta_p^2 = 0.61$. However, old faces seen with a switched context did not elicit reliably more remember responses than old faces seen with a new context, $F(1, 48) = 0.18$, MSE < 0.01, $p = 0.67$, $\eta_p^2 < 0.01$. There was no effect of context on remember false alarms (*new faces–switched scene* vs. *new faces–new scene*: $F(1, 48) = 0.96$, MSE < 0.01, $p = 0.33$, $\eta_p^2 = 0.02$). This pattern of results nicely demonstrates the butcher-on-the-bus phenomenon in the laboratory. That is, seeing faces in their original context elicits both more accurate recognition and a feeling of recollection than seeing faces in different or novel contexts.

We next examined the effects of context on familiarity assuming both exclusivity and independence of processes. There was a main effect of context on the proportion of correct know responses, $F(2, 96) = 4.60$, MSE < 0.01, $p = 0.01$, $\eta_p^2 < 0.09$. This was because fewer know responses were made to old faces seen with their studied scenes than to old faces seen with switched scenes, $F(1, 48) = 6.56$, MSE < 0.02, $p < 0.02$, $\eta_p^2 < 0.12$. There was no reliable difference in the proportion of correct know responses made to old faces seen with switched scenes and old faces seen with novel scenes, $F(1, 48) < 0.01$, MSE $= 0.01$, $p = 0.94$, $\eta_p^2 < 0.01$. Context had no reliable effect on the proportions of know false alarms, $F(1, 48) = 0.05$, MSE < 0.01, $p < 0.02$, $\eta_p^2 < 0.01$. The effects of context on estimates of familiarity assuming independence revealed a

FIGURE 3 | The proportions of remember and know responses for each condition in Experiment 1.

FIGURE 4 | Effects of context on estimates of recollection and familiarity in Experiment 1.

butcher-on-the-bus phenomena describes a situation in which a familiar person is seen out of context and thus, although retrieval fails, they nonetheless feel familiar. This does not apparently occur under our laboratory conditions for either studied or unstudied faces.

We next estimated the effects of context on the sensitivity of different memory "processes" using the statistic d' (see **Figure 4**). To do so the hit rates for faces presented with either studied or switched contexts were compared with the false alarm rates for new faces presented with switched contexts (*old faces–old scenes* vs. *new faces–switched scenes*, and *old faces–switched scenes* vs. *new faces–switched scenes*), and the hit rates for faces presented with new scenes were compared to new faces presented with new scenes (*old faces–new scenes* vs. *new faces–new scenes*). We used the standard formulas (Macmillan and Creelman, 1991) to compute sensitivity [$d' = z(\text{hits}) - z(\text{false alarms})$], and criterion placement [$c = -0.5 \times z(\text{hits}) + z(\text{false alarms})$]. To prevent values of 1 and 0 in the hit and false alarm rates we used the Snodgrass and Corwin (1988) correction in which a constant of 0.5 is added to each cell frequency and is then divided by $n + 1$. There was an effect of context on the sensitivity of recollection, $F(2, 96) = 38.63$, MSE $= 0.13$, $p < 0.01$, $\eta_p^2 = 0.45$, this was because recollection was more sensitive when old faces were presented with the same context scene as at study compared to when the context was switched, $F(1, 48) = 76.19$, MSE < 0.18, $p < 0.01$, $\eta_p^2 = 0.61$. Faces seen with switched contexts did not result in an increase in the sensitivity of recollection compared to old faces seen with new contexts, $F(1, 48) = 0.21$, MSE < 0.36, $p = 0.65$, $\eta_p^2 < 0.01$. There was a similar pattern of data for the effects of context of the sensitivity of the IRK estimates of familiarity, $F(2, 96) = 15.07$, MSE $= 2.17$, $p < 0.01$, $\eta_p^2 = 0.24$, that was due to an increase in sensitivity

different pattern of results. There was a main effect of context, $F(2, 96) = 18.90$, MSE < 0.04, $p = 0.01$, $\eta_p^2 < 0.28$, but in contrast to the pattern observed assuming exclusivity, this was due to an increase in familiarity when old faces were seen with their original scene, $F(1, 48) = 22.49$, MSE $= 0.08$, $p = 0.01$, $\eta_p^2 < 0.32$. There was no difference in familiarity when old faces were seen with switched contexts compared to when they were seen in new contexts, $F(1, 48) = 0.57$, MSE $= 0.07$, $p = 0.46$, $\eta_p^2 = 0.01$. Although these two patterns differ, and one might question which assumption to believe, the real problem for the dual-process interpretation of the butcher-on-the-bus phenomenon is that the feeling of familiarity doesn't appear to increase when faces are seen in switched contexts relative to when they are seen in novel contexts. The
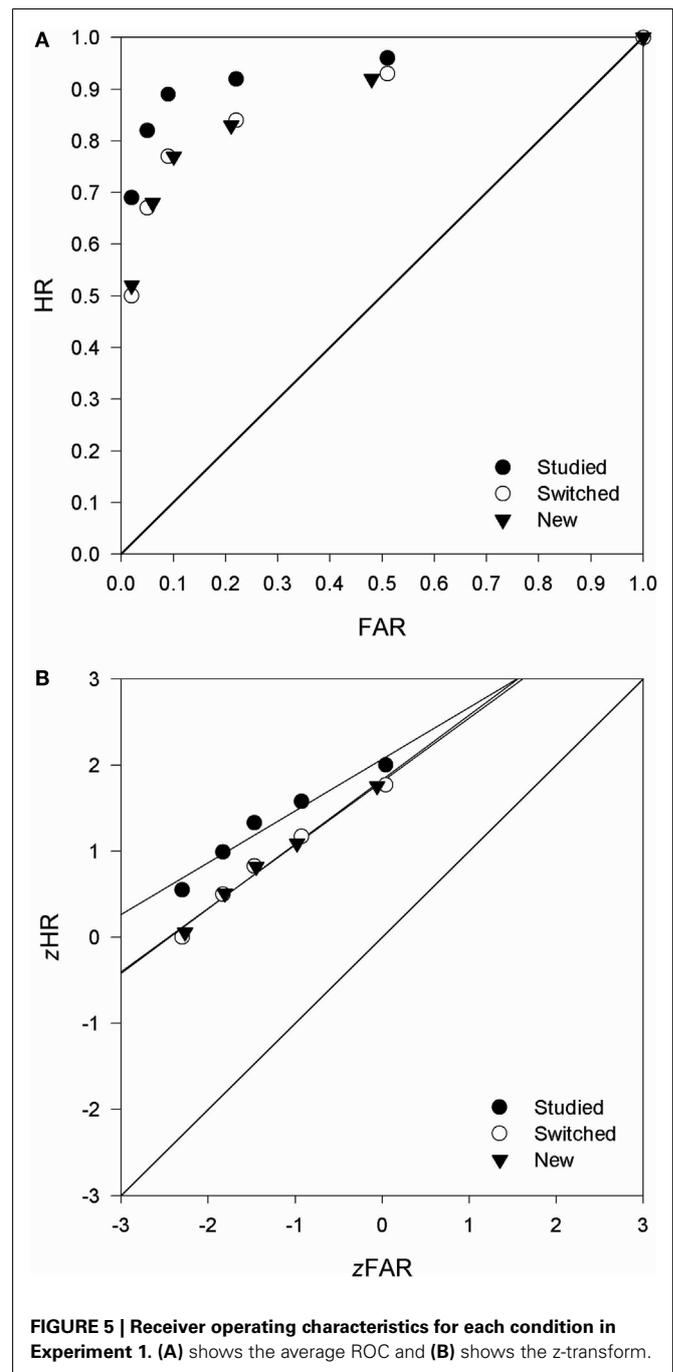
when old faces were presented with their studied context scenes compared to when they were presented with switched contexts, $F(1, 48) = 14.61$, MSE $= 5.97$, $p < 0.01$, $\eta_p^2 < 0.23$. Sensitivity was not reliably higher when old faces were seen with switched contexts than when they were seen with new contexts, $F(1, 48) = 1.33$, MSE $= 0.79$, $p = 0.25$, $\eta_p^2 = 0.03$. This latter pattern of results differs from that reported by Gruppuso et al. (2007) who found no increase in the sensitivity of familiarity for faces seen in studied contexts.

Context also had a reliable effect on criterion placement (c) for remember responses, $F(2, 96) = 45.95$, MSE $= 0.25$, $p < 0.01$, $\eta_p^2 = 0.49$. Participants were adopted a more liberal criterion for remember responses when old faces were presented with the same context scene as at study ($M = 0.49$, SD $= 0.25$) compared to when the context was switched ($M = 0.76$, SD $= 0.25$), $F(1, 48) = 76.20$, MSE $= 0.22$, $p < 0.01$, $\eta_p^2 = 0.61$. There was no difference in the remember criterion placement when old faces were seen with switched contexts compared to old faces seen with new contexts ($M = 0.76$, SD $= 0.29$), $F(1, 48) < 1.0$. A smaller effect of context was observed on criterion placement for know responses $F(2, 96) = 3.49$, MSE $= 0.38$, $p < 0.05$, $\eta_p^2 = 0.07$. Participants were more conservative in the criterion placement for know response when old faces were presented with the same context scene as at study ($M = 1.25$, SD $= 0.33$) compared to when the context was switched ($M = 1.17$, SD $= 0.36$), $F(1, 48) = 5.55$, MSE $= 0.03$, $p < 0.05$, $\eta_p^2 = 0.10$, but not when old faces were seen with switched contexts compared to old faces seen with new contexts ($M = 1.15$, SD $= 0.34$), $F(1, 48) < 1.0$.

RECEIVER OPERATING CHARACTERISTICS

If the butcher-on-the-bus phenomenon represents a dissociation between recollection and familiarity-based memory then this should be apparent not just in the subjective experience of remembering, but also in the ROC averaged over each participant. The precise form of the ROC discriminates between the possible interpretations of the effect of context on recognition. If recognizing a face in a studied context simply increases the overall signal strength of those faces then we should see an increase in the asymmetry of the ROC curve as the variance in the signal distribution increases. On the other hand, if context serves to cue the retrieval of the face from recollection (as opposed to increasing its signal strength) then we should see a quadratic component in the slope of the zROC curve (Glanzer et al., 1999). To test this we first constructed ROC for each participant based on their confidence ratings, and a corresponding zROC. The x and y coordinates for each comparison were the same as the remember and know responses. As a measure of symmetry we found the standardized regression coefficient (β) of the zROC for each participant. We then looked for a quadratic constant (bx) in the regression. The average ROC and zROC curves for each condition are shown in **Figure 5**. The average standardized regression coefficients (β) and the average quadratic components (bx) for each condition are shown in **Table 1**.

The standardized regression coefficients for all three zROCs were reliably less than 1. But none of the quadratic constants were greater than 0. When old faces were seen with studied contexts the

FIGURE 5 | Receiver operating characteristics for each condition in Experiment 1. (A) shows the average ROC and **(B)** shows the z-transform.
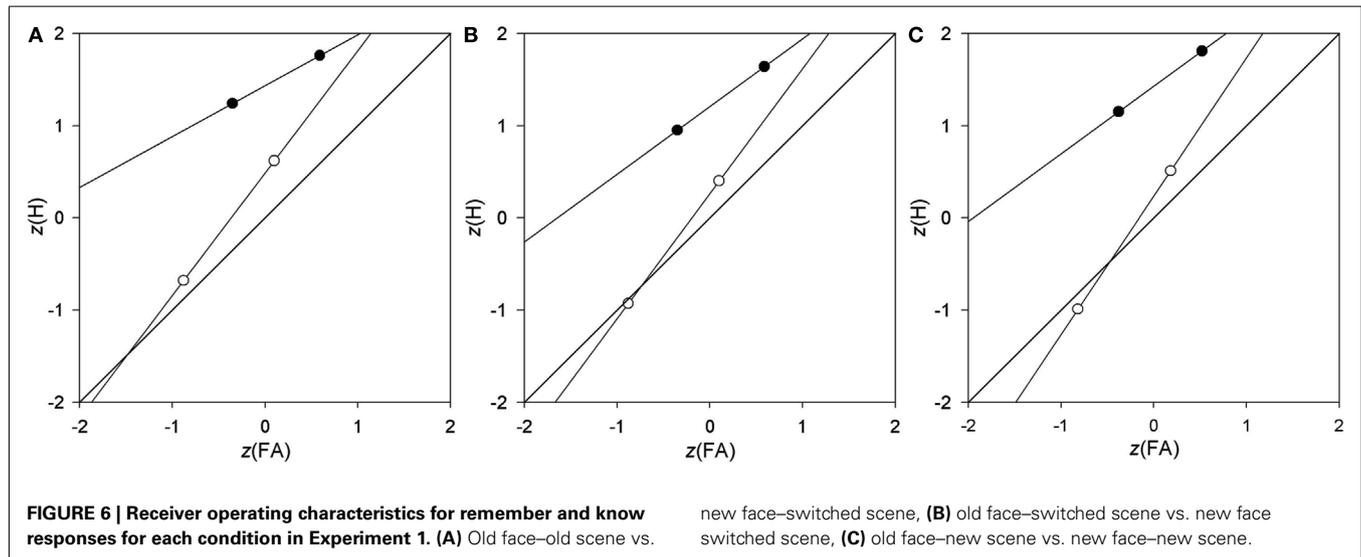
asymmetry was greater than when old faces were seen with new contexts, $t(48) = -2.24$, $p = 0.03$, but none of the other comparisons approached significance. This pattern suggests that context serves to increase the inequity in variance of the signal and noise distributions that causes the asymmetry in the zROC and that it has a continuous effect across the scale of confidence judgments. This matches both the univariate signal detection model and Wixted and Mickes' dual-process extension of it.

In many respects the Butcher-on-the-Bus is defined by a strong feeling of familiarity without the retrieval of any contextual detail

Table 1 | Distribution of responses for each level of confidence and each item type in Experiment 1.

	Remember			Know			New		
	Sure	Fairly sure	Unsure	Sure	Fairly sure	Unsure	Sure	Fairly sure	Unsure
Old face–old scene	408	20	5	24	52	31	15	19	14
Old face–switched scene	290	36	3	18	66	52	36	48	39
Old face–new scene	582	54	13	45	139	90	85	107	61
New face–switched scene	12	8	1	1	19	50	595	348	142
New face–new scene	16	7	4	4	26	42	623	334	120

FIGURE 6 | Receiver operating characteristics for remember and know responses for each condition in Experiment 1. (A) Old face–old scene vs. new face–switched scene, **(B)** old face–switched scene vs. new face switched scene, **(C)** old face–new scene vs. new face–new scene.

(Wixted and Mickes, 2010). In the laboratory this would be measured as know responses made with high confidence. The raw frequencies of remember and know responses for each level of confidence and for each condition are shown in **Table 1**. These data show that in actuality the participants made relatively few high confidence know responses, and instead tended to report high confidence responses as remember responses. Nonetheless even if this phenomenon is less common than the received wisdom would have us believe, these few responses might nonetheless be accurate. To test this we plotted the ROC (see **Figure 6**) for remember and know responses separately for each condition. This enables us the compute a d' value for each level of confidence (see **Table 2**). The results show that high confidence know responses fall on the diagonal indicating that these recognition decisions are no more accurate than chance. In contrast less confident know responses are more accurate than chance. It seems that given the choice of reporting a high confidence recognition decision as recollection or as familiarity the participants in this experiment opted for recollection.

EXPERIMENT 2

An obvious question is whether recollection and familiarity differ in how they encode context. We examined this issue in Experiment 2 by asking participants to identify the contexts associated with recognized faces. If recollection encodes context but familiarity

does not then we would expect that know responses would result in poor accuracy in identifying study contexts. On the other hand if both recollection and familiarity differ only in terms of signal strength then the identification of contexts should be good for both responses. We also ask whether the retrieval of context is associated with different levels of confidence. Specifically whether know responses made with high confidence are associated with more accurate context judgments than lower confidence judgments. Such a result would be problematic for the standard dual-process model in which context is associated only with recollection, but is predicted by the recent model described by Wixted and Mickes (2010).

MATERIALS AND METHODS

Participants

Forty-eight members of the University of Nottingham community volunteered for this experiment. Twenty-nine were female and 19 were male. Their average age was 21.35 years (SD = 1.84). All had normal or corrected vision.

Procedure

The procedure for the study phase was identical to that of Experiment 1. There was no interval between study and test. During the test period participants were presented with a face without any context item. They then responded whether it was old or new by

Table 2 | Mean sensitivity for sure and fairly sure confidence responses for remember and know responses and for each condition in Experiment 1.

	Old face–old scene vs. new face–switched scene				Old face–switched scene vs. new face–switched scene				Old face–new scene vs. new face new scene			
	Remember		Know		Remember		Know		Remember		Know	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Sure	1.59	0.08	0.20	0.08	1.31	0.08	−0.04	0.09	1.53	0.08	−0.16	0.10
Fairly sure	1.17	0.06	0.52	0.12	1.05	0.05	0.30	0.14	1.29	0.07	0.33	0.10

Table 3 | Showing the mean proportions of hits and false alarms for each subjective report of remembering in Experiment 2.

	Remember		Know		Total	
	M	SE	M	SE	M	SE
Hits	0.48	0.03	0.32	0.02	0.79	0.01
False alarms	0.02	0.01	0.10	0.01	0.11	0.01
Sensitivity (d')	2.10	0.08	0.96	0.10	2.29	0.10
Criterion (c)	1.09	0.05	1.05	0.07	0.31	0.05

clicking on one of 12 buttons. Each was labeled 1–6 counting out from "don't know" in the center to "sure-new" and "sure-old" at either side of the screen. The levels of confidence response was increased from three in Experiment 1 so that separate ROC plots could be calculated for remember and know judgments with sufficient power and reliability. If participants responded using one of the six old buttons they were then asked to provide a remember-know judgment. Four scenes were then presented: the correct old scene with three novel scenes for old trials and four novel scenes for foil trials.

Stimuli

The study items consisted of the same faces and scenes used in Experiment 1. During the study phase they were presented as 48 study face-scene pairs. The test items consisted of faces from the study period (old items) and an additional 48 novel faces (new items). The 4AFC items were composed of the study scenes (for old items) and novel scenes.

RESULTS

The average response rates for each item type and the resulting measures of sensitivity and criterion are shown in **Table 3**. The d' values computed over responses reported as remember judgments were more sensitive than those computed over know responses, $t(47) = 7.55$, SD = 1.04, $p < 0.01$, but there was no difference in criterion (c) placement, $t(47) < 1.0$.

The slope of zROC plots for recognition responses did not differ from 1, $b_{all} = 0.76$, SE = 0.09, $t(47) = −0.25$, $p > 0.05$ and there was no quadratic component $\hat{b}^2_{all} = −0.36$, SE = 0.23, $t(47) = −0.15$, $p > 0.05$. The ROC plots for remember and know responses were then calculated separately and are shown in **Figure 7**. The slope of the regression coefficient for "remember" responses was reliably less than 1, $b_{remember} = 0.71$, SE = 0.07,

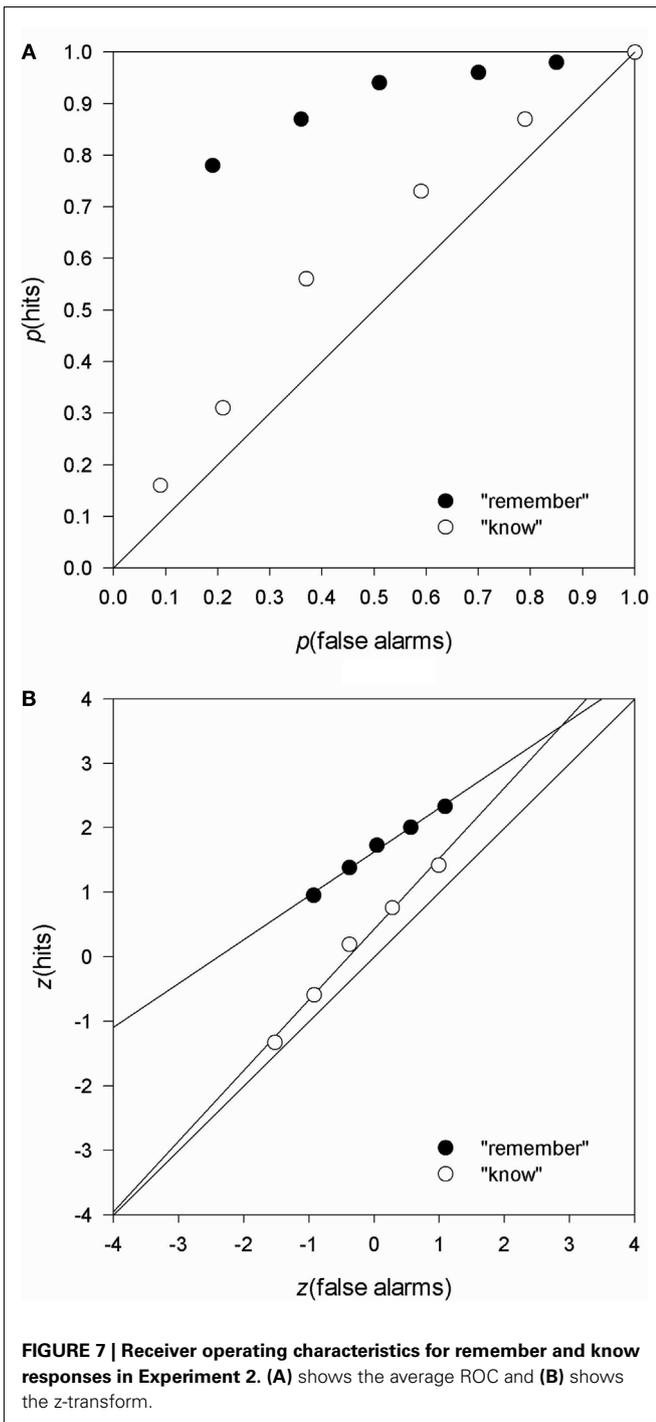$t(47) = 4.49$, $p < 0.01$, but not for "know" responses, $b_{know} = 1.13$, SE = 0.08, $t(47) = 1.70$, $p > 0.05$. The two slopes were reliably different from one another, $t(47) = 4.04$, SE = 0.11, $p < 0.01$, suggesting that remember-know judgments are representing something independent of confidence. Neither curve showed a quadratic component that differed reliably from 0, $\hat{b}^2_{remember} = −0.07$, SE = 0.06, $t(47) = 1.05$, $p > 0.05$; $\hat{b}^2_{know} = −0.17$, SE = 0.11, $t(47) = 1.59$, $p > 0.05$.

Now we turn to the question of whether recollection and familiarity differ in terms of their encoding of context. If recollection encodes context but familiarity does not then we expect that 4AFC accuracy for context should be close to chance for recognition judgments associated with "know" responses and highly accurate for judgments given a remember "response." However, although 4AFC accuracy was reliably greater for "remember" than "know" responses, $M = 0.87$, SE = 0.02 vs. $M = 0.71$, SE = 0.02, respectively: $t(46) = 6.90$, $p < 0.01$, 4AFC accuracy was reliably greater than chance even for familiarity, $t(46) = 19.60$, $p < 0.01$.
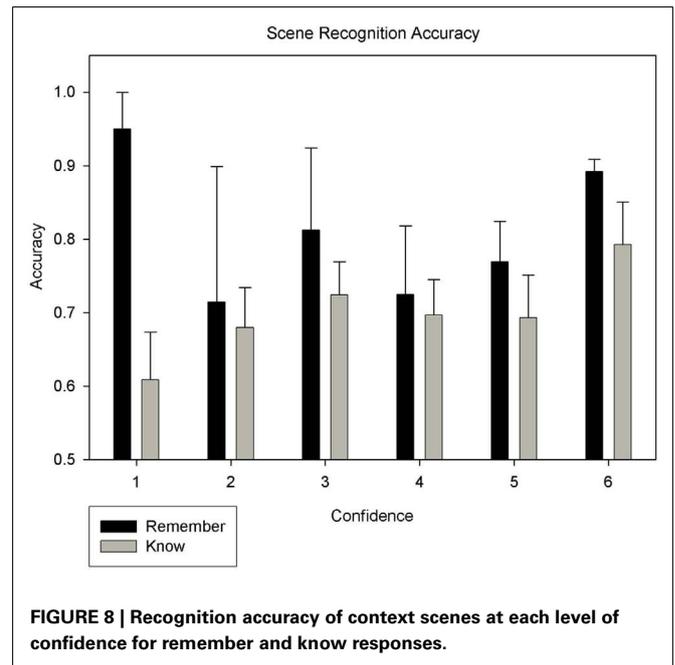
We therefore examined the accuracy of context judgments for each level of confidence and each subjective report of remembering (see **Table 4** and **Figure 8**). The data reveal that context judgments are significantly above chance for each level of confidence for both remember and know responses. This pattern demonstrates that there is a relationship between confidence and context recollection even when no remember response is made. A traditional threshold dual-process model cannot explain this and although the Wixted and Mickes model can accommodate it this requires the assumption of partially correlated familiarity and recollection systems.

DISCUSSION

The butcher-on-the-bus has long been used as a rhetorical device to illustrate how context can dissociate recognition based on familiarity and recognition based on recollection. We show that the rhetorical device can be reproduced as a real laboratory phenomenon using the stimuli on which it is based. However, our analyses show that the effect cannot be fully explained by either traditional threshold dual-process accounts or univariate signal detection. That is, the phenomenon does not neatly discriminate between models of recognition memory. Patterns of target recognition in both experiments pose a problem for the dual-process account as it does not predict low confidence remember responses, nor the reliable linear relationship found between remember hits and false alarms in zROCs. This is particularly problematic in Experiment

FIGURE 7 | Receiver operating characteristics for remember and know responses in Experiment 2. **(A)** shows the average ROC and **(B)** shows the z-transform.

FIGURE 8 | Recognition accuracy of context scenes at each level of confidence for remember and know responses.

Context accuracy analysis in Experiment 2 leads to a similar conclusion whereby neither SDT nor threshold models are supported. Threshold models can potentially explain above chance source memory for know responses, as presentation of the scene itself means the task arguably becomes a secondary recognition task rather than true recollection resulting purely from seeing the paired face. This would enable a familiarity system to increase correct responding. However, these models cannot explain the linear correlation between confidence and context accuracy for remember responses. In addition, the univariate account cannot explain the differences in accuracy for remember and know responses which are independent of item recognition confidence.

On balance it seems that the data presented here supports neither the traditional threshold based dual-process models nor traditional univariate signal detection. After 30 years the butcher-on-the-bus phenomenon can still reveal more about the nature of human memory. We feel that it reveals the existence of two orthogonal signal detection systems that have a summative relationship with recognition decisions. Although multiple memory system models are notoriously controversial there has nonetheless been a recent swell of evidence and support in the literature (Rotello et al., 2004; Wixted and Stretch, 2004; Wixted, 2007; Wixted and Mickes, 2010).

An important limiting factor on the interpretation of out experiments is the number of trials that were available to obtain parameter estimates of the ROC. Typically far more items are used in recognition memory experiments using words as stimuli than we were able to use in our experiments. Indeed Yonelinas and Parks (2007) note that between 50 and 60 items per condition are needed to reliable parameter estimates. Experimental preparations differ slightly for experiments involving either face recognition or subjective reports of remembering. Face recognition experiments often use fewer items than are typically used in

1 where reinstating context significantly alters the slope of the corresponding zROC without any quadratic component being evident. This suggests a continuous underlying system(s) such as signal detection, where the additional contextual information acts as a cue which increases signal strength across the scale's entire range. However, when separate plots are created for remember and know in Experiment 2, the univariate account fails as it cannot explain the reliable difference in slopes that show an underlying dichotomy along the whole confidence scale.

Table 4 | Distributions of responses for each level of confidence and each item type in Experiment 2.

Confidence	Lure			Target			Correct context		Incorrect context	
	Remember	Know	New	Remember	Know	New	Remember	Know	Remember	Know
1	2	45	47	10	84	94	9	53	1	31
2	2	45	47	11	113	144	9	73	2	40
3	8	44	52	22	125	228	18	86	4	39
4	2	30	32	50	173	321	38	127	12	46
5	4	23	27	87	119	283	66	88	21	31
6	8	6	14	927	107	1015	834	85	93	22

word recognition (e.g., Chan et al., 2011) and estimates of recollection and familiarity from subjective reports of remembering are themselves sensitive to the number of items to be remembered (Cary and Reder, 2003). Thus when in our Experiment 1 we sought to replicate a study that combined subjective reports of remembering and face recognition (Gruppuso et al., 2007) we used fewer items than might otherwise have been desirable to obtain ROC parameter estimates. If we had applied this criterion to our experiment there would have been a minimum of 480 trials which would not have been a replication of the experiment that we intended. Nonetheless this weakness would only seem to apply to the absence of detectable quadratic components in the zROC, and not at all to the effects of context on the subjective reports of remembering, or on the calculations of sensitivity across the different levels of confidence.

The recognition memory experiments overwhelmingly use words as stimuli because the characteristics of words that affect encoding and retrieval such as frequency and concreteness are well documented and this allows experiments to be conducted with carefully controlled stimuli. Indeed research that addresses very similar issues to those asked here has been conducted using words as stimuli (Wixted and Mickes, 2010; Ingram et al., 2012). Indeed the dual-process vs. single-process theoretical framework is derived almost entirely from word recognition (for reviews see Yonelinas, 2002; Yonelinas and Parks, 2007). The research reported here is intended to complement this body of knowledge by demonstrating that its findings are relevant in the domain that originally formed the basis of the field itself.

One final important issue is how we should interpret recognition responses that in our Experiment 2 participants reported as being based on familiarity (i.e., knowing), but which are also accompanied with a the retrieval of accurate contextual details. This is an issue at the heart of the recognition memory literature and in theoretical terms at least, defines the distinction between recollection and familiarity-based memory. The traditional view is that recollection encodes context (or is defined by the retrieval of contextual details) and familiarity does not. There are therefore three possible interpretations of our data. One possibility is that know responses were "contaminated" by recollection-based memory, presumably because the participants failed to follow or understand the instructions. We feel that this explanation is unlikely because we used a modification of the standard instructions and was precisely the hypothesis that we set out to test. Indeed it only makes sense to talk about contamination of familiarity with recollection if one adopts a dual-process perspective. The second is that the two processes are not separate processes at all. This interpretation too seems unlikely on the basis of the weight and diversity of evidence for some separation of processes. The third is that recollection and familiarity-based memory processes are not exclusively defined by the encoding of contextual information. Indeed, similar studies point to processes that are distinguished separate dimensions or signals that contribute to a recognition response (Wixted and Mickes, 2010; Ingram et al., 2012). Alternatively we may have observed the accurate context memory for items that had been attributed as familiarity-based is because we used a forced-choice procedure and so may have elicited knowledge that may not have been revealed in less sensitive tests of source memory such as recall.

CONCLUSION

In summary it seems that the weight of evidence from this study as well as other behavioral and neuroscience investigations points to two underlying processes. However, the data does not fit threshold models or those which assume exclusive prioritizing of one process over the other. Instead it seems these results can only be explained by orthogonal signal detection systems, which have an integrative relationship toward the eventual recognition response. A model that has properties of the sort described by Wixted and Mickes (2010) appears to explain the relevant empirical data and the data reported here, although this requires an assumption of partial collinearity between processes to do so.

REFERENCES

Bell, R., and Buchner, A. (2010). Valence modulates source memory for faces. *Mem. Cognit.* 38, 29–41.

Bell, R., and Buchner, A. (2011). Source memory for faces is determined by their emotional evaluation. *Emotion* 11, 249–261.

Brandt, K. R., Gardiner, J. M., and Macrae, C. N. (2006). The distinctiveness effect in forenames: the role of subjective experiences and recognition memory. *Br. J. Psychol.* 97, 269–280.

Brandt, K. R., Macrae, C. N., Schloerscheidt, A. M., and Milne, A. B. (2003). Remembering or knowing others? Person recognition and recollective experience. *Memory* 11, 89–100.

Buchner, A., Bell, R., Mehl, B., and Musch, J. (2009). No enhanced recognition memory, but better source memory for faces of cheaters. *Evol. Hum. Behav.* 30, 212–224.

Cary, M., and Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *J. Mem. Lang.* 49, 231–248.

Chan, J. P. K., Kamino, D., Binns, M. A., and Ryan, J. D. (2011). Can changes in eye movement scanning alter the age-related deficit in recognition memory? *Front. Psychol.* 2:92. doi:10.3389/fpsyg.2011.00092

Davies, G., and Milne, A. B. (1982). Recognizing faces in and out of context. *Curr. Psychol. Rev.* 2, 235–246.

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

Dudukovic, N., and Knowlton, B. J. (2006). Remember-know judgments and the retrieval of contextual details. *Acta Psychol. (Amst.)* 122, 160–173.

Dunn, J. C. (2004). Remember-know: a matter of confidence. *Psychol. Rev.* 95, 524–542.

Eldridge, L. L., Engel, S. A., Zeineh, M. M., Bookheimer, S. Y., and Knowlton, B. J. (2005). A dissociation of encoding and retreival processes in the human hippocampus. *J. Neurosci.* 25, 3280–3286.

Gardiner, J. M., and Richardson-Klavehn, A. (2000). "Remembering and knowing," in *The Oxford Handbook of Memory*, eds E. Tulving and F. I. M. Craik (New York: Oxford University Press), 229–244.

Glanzer, M., Kim, K., Hilford, A., and Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 500–513.

Godden, D. R., and Baddeley, A. D. (1980). When does context affect recognition memory? *Br. J. Psychol.* 71, 99–104.

Gregg, V. H., Gardiner, J. M., and Konstaninou, I. (2006). Recognition memory and awareness: a high frequency advantage in the accuracy of knowing. *Memory* 14, 265–275.

Gruppuso, V., Lindsay, D. S., and Masson, M. E. J. (2007). I'd know that face anywhere. *Psychon. Bull. Rev.* 14, 1085–1089.

Ingram, K. M., Mickes, L., and Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 325–339.

Jacoby, L. L., Yonelinas, A. P., and Jennings, J. R. (1997). "The relation between conscious and unconscious (automatic) influences: a declaration of independence," in *Scientific Approaches to Consciousness*, eds J. D. Cohen and J. W. Schooler (Mahwah: Lawrence Erlbaum Associates), 13–47.

Macken, W. (2002). Environmental context and recognition: the role of recollection and familiarity. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 153–161.

Macmillan, N. A., and Creelman, C. D. (1991). *Detection Theory: A User's Guide*. Cambridge: Cambridge University Press.

Mandler, G. (1980). Recognizing: the judgement of previous occurrence. *Psychol. Rev.* 87, 252–271.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychol. Rev.* 39, 252–370.

Perfect, T. J., Mayes, A. R., Downes, J. J., and Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *Q. J. Exp. Psychol.* 49A, 797–813.

Rainis, N. (2001). Semantic contexts and face recognition. *Appl. Cogn. Psychol.* 15, 173–186.

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

Rutherford, A. (2004). Environmental context-dependent recognition memory effects. *Q. J. Exp. Psychol.* 57A, 107–127.

Snodgrass, J. G., and Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J. Exp. Psychol. Gen.* 117, 34–50.

Tulving, E. (1985). "On the classification problem in learning and memory," in *Perspectives in Learning and Memory*, eds L. G. Nilsson and T. Archer (Hillsdale, NJ: Erlbaum), 67–94.

Tunney, R. J. (2010). Do changes in the subjective experience of recognition over time suggest independent processes? *Br. J. Math. Stat. Psychol.* 63, 43–62.

Tunney, R. J., and Bezzina, G. (2007). Effects of retention intervals on receiver operating characteristics in artificial grammar learning. *Acta Psychol. (Amst.)* 125, 37–50.

Wais, P. E., Mickes, L., and Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *J. Cogn. Neurosci.* 20, 400–405.

Watkins, M. J., Ho, E., and Tulving, E. (1976). Context effects on recognition memory for faces. *J. Verb. Learn. Verb. Behav.* 15, 505–517.

Winograd, E., and Rivers-Bukeley, N. T. (1977). Effects of changing context on remembering faces. *J. Exp. Psychol. Hum. Learn. Mem.* 3, 397–405.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychol. Rev.* 114, 152–176.

Wixted, J. T., and Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychol. Rev.* 117, 1025–1054.

Wixted, J. T., and Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychon. Bull. Rev.* 11, 616–641.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 1341–1354.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46, 441–517.

Yonelinas, A. P., and Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: a review. *Psychol. Bull.* 133, 800–832.

Yovel, G., and Paller, K. A. (2004). The neural basis of the butcher-on-the-bus phenomenon: when a face seems familiar but is not remembered. *Neuroimage* 21, 789–800.