I think I'd change the phrase 'subjective experience' to 'recollective experience'.

The Relationship Between Source, Confidence, and Subjective Experience When Two Judgments Are Made Post-Recognition

Helen L. Williams[a1], Martin A. Conway[a2], & Chris J.A. Moulin[a3]

[a] Affiliation where the work was carried out: Leeds Memory Group, Institute of Psychological Sciences, University of Leeds, Leeds, LS2 9JT, UK.

[1] Corresponding author. Present address: Department of Psychology, P. O. Box 3050, University of Victoria, Victoria, British Columbia, Canada, V8W 3P5.

Email: helenwilliams098@gmail.com. Phone: +1 250 721 7554. Fax: +1 250 721 8929.

[2] Present address: Department of Psychology, City University, Northampton Square, London, EC1V 0HB, UK. Email: martinconway1@me.com. Telephone: +44 (0)20 7040 8500.

[3] Present address: LEAD - CNRS UMR 5022, Université de Bourgogne, Pôle AAFE, Esplanade Erasme, BP 26513, 21065 Dijon CEDEX.

Email: chris.moulin@me.com. Telephone: +33 (0)3 80 39 39 68.

Abstract

In research on recognition memory processes, concerns have been raised that multiple post-recognition judgments may influence each other (e.g., Bruno & Rutherford, 2010). To test for influence, acrossthis in three experiments judgments of subjective experience, confidence, and source were compared and with order of judgments was counterbalanced across items. Comparison of responses across four categories of subjective experience (Remember, Know, Familiar, and Guess) and four levels of confidence (High, Medium, Low, and None) showed that confidence judgments were faster, more lenient, less sensitive to source accuracy, and influenced to a greater degree by manipulation of source, than were judgments of subjective experience. Overall, confidence and subjective experience judgments were not utilised in the same manner; however, both response distributions and response time (RT) data demonstrated that order of judgments did not influence responding. Crucially for understanding subjective experience judgments, Know and Familiar responses differed in recognition accuracy, source accuracy, confidence, and recognition RT. In contrast, Remember and Know judgments were only differentiated by source accuracy. Results suggest that subjective experience, source, and confidence judgments can safely accompany one another post-recognition, and subjective experience can go beyond the traditional Remember-Know(-Guess) procedure in simple episodic tasks. THIS LAST BIT NEED TO BE MORE SPECIFIC.

Keywords

Remember-Know, recognition, recollection, familiarity, source-monitoring, dual-process.

Acknowledgements

This research was undertaken as part of the first author's PhD research at the University of Leeds funded by a University Research Scholarship. Portions of this data were presented at the British Psychological Society Postgraduate Affairs Group Conference (PSYPAG; July 2009), the British Psychological Society Cognitive Section Conference (September, 2009), and the Experimental Psychology Society London Meeting (January 2011).

"As remarked by an anonymous reviewer, whether confidence ratings can safely accompany an R-K(-G) judgement, and whether they should follow or precede such judgement, is as much an empirical as it is a theoretical question." (Bruno & Rutherford, 2010, footnote 2, p. 127).

In recognition memory, three type of judgment have been used to examine the processes people use to make their recognition decisions: judgments of confidence, subjective experience, and source. The current experiments examine relationships between these judgment types.

Confidence judgments ask participants to rate the level of confidence they hold for their recognition decision and vary from binary judgments (e.g., sure/unsure; Gardiner & Java, 1990) to 20-point rating scales (e.g., Wixted & Mickes, 2010). Subjective experience is typically measured using the Remember-Know (RK) procedure where participants make a judgment about their subjective experience for each item retrieved from memory by making a 'Remember' or 'Know' response. An item should be classified as Remember when the participant retrieved accompanying detail about something thought or experienced at the time of encoding and categorised as Know when the participant was aware that the item had been on the study list but could not recall anything experienced for the item (Gardiner & Richardson-Klavehn, 2000; Tulving, 1985). When the RK procedure was first introduced Remember and Know responses were posited as reflecting autonoetic and noetic consciousness based on retrieval from episodic and semantic memory respectively (Tulving, 1985). However, subsequently the RK procedure has been primarily utilised in the study of episodic recognition processes (Bruno & Rutherford, 2010; Yonelinas, 2002). In this sphere, while some view the subjective states of Remembering and Knowing as orthogonal to the underlying processes of recollection and familiarity (for example see Gardiner, 2000, and Wixted & Mickes, 2010), the

majority of studies using the RK procedure assume that a Remember response is an index of recollection while a Know response is an index of familiarity (Yonelinas, 2002).

Source judgments are somewhat similar to subjective experience judgments as they are also based on retrieval (or not) of information from encoding. In source monitoring procedures, participants make a judgment about some aspect of the context in which an item was originally presented. Examples are whether an item was written in a particular font/colour/size, was shown on the left or right of the screen, was spoken in a particular voice, or a combination of these contexts (e.g., Meiser & Bröder, 2002). A key difference between source judgments and judgments of subjective experience is that source judgments involve a stricter criterion. The only form of recollection which 'counts' in this procedure is recollection which allows the individual to determine which list the item was shown on. Noncritical recollection, for example recalling that you pictured your favourite coffee shop when the word 'café' was shown, is not helpful unless it supports source discrimination.

In this paper, three experiments are presented that examine the influence that judgments of confidence, source, and subjective experience have on each other. Analysis focuses on distribution of responses across confidence levels, subjective experience categories, and source accuracy, dependent on whether judgments were made first or second following a recognition decision. Reaction times to make Old/New recognition responses are analysed dependent on subsequent judgment and, in a novel approach to this field, the time taken to make the actual judgments is also explored.

The relationship between confidence and subjective experience is of critical importance to the debate between single- versus dual-process accounts of memory. Dual-process models assume that two distinct processes or forms of memory underlie recognition and that successful

recognition is determined by the relative contributions of both processes. Conversely, single-process accounts propose that recognition relies on only one continuous dimension of familiarity or memory strength, and successful recognition is determined by the strength of this single dimension (for comprehensive reviews of the literature summarizing the opposing viewpoints see Diana, Reder, Arndt, & Park, 2006; Dunn, 2004, 2008; Gardiner, 2008; Parks & Yonelinas, 2007; Wixted & Stretch, 2004; and Yonelinas, 2002). Through the current experiments we aim to better understand recognition processes theoretically by comparing the more subjective measures of RK and confidence with the relatively more objective measure of source. Here we are not concerned with providing evidence which supports one or other of the increasingly complex single- or dual-process models (Gardiner, 2008); rather the experiments presented here take post-recognition judgments at face value and examine how people make and understand the three types of judgment under study. The move 'back' towards understanding Remembering and Knowing at a more conceptual level was advocated by both Dunn (2004) and Gardiner (2008).

As stated in the opening quotation, whether judgments influence each other is of interest both empirically and theoretically. However, many experimental paradigms combine judgments without the issue of influence having been explored. Experiments that have compared combinations of two of these three judgments are discussed in subsequent sections. Furthermore, since Tulving (1985) introduced the RK paradigm researchers have used a number of variations on the paradigm. Methodological differences pertinent to the design of the current experiments are discussed. In addition to focusing on response patterns for subjective experience, source, and confidence judgments when different judgments are compared, experiments that have explored reaction times associated with different responses are also

reviewed. Reaction time to make judgments was a primary analysis of interest in the current experiments as, to our knowledge, no one has yet compared how long it takes participants to make these three different types of judgment.

Methodological issues for measuring subjective experience

One methodological issue in the RK paradigm is whether use of one-step or two-step RK procedures influences recognition accuracy or makes Remember responding more lenient. Two-step procedures involve separate judgments of recognition (old/new) and subjective experience. Participants initially judge whether they recognise an item, and only if they do recognise the item are they asked to judge their state of awareness regarding that recognition. One-step procedures combine the recognition judgment with the judgment of subjective experience: participants are asked to judge whether an item is Remembered, Known, or New. Comparisons suggest that no differences in overall recognition accuracy are found using one-step and two-step procedures. However, if interested in the experiential states underlying recognition and use of Remember and Know responses across different experimental conditions, the two methods produce different patterns of responding, primarily for assignment of hits and false alarms to Know (Bruno & Rutherford, 2010; Eldridge, Sarfatti, & Knowlton, 2002; Hicks & Marsh, 1999). Moreover, if interested in reaction time (RT) measures, separation of the recognition decision from the judgment of subjective experience – as occurs in a two-step procedure – is critical for isolating the time taken to perform these separate processes. Examination of source, confidence, and subjective experience judgments using one-step and two-step procedures has produced some divergent results and methodological differences are drawn upon in later discussion of experimental findings.

A second methodological issue is whether the original two categories of Remembering and Knowing are sufficient or whether other states of awareness can be identified. Whilst the addition of a Guess response option is now common in Remember-Know experiments, and is included in order to remove experimental 'noise' caused by participants assigning guesses to the Know category (Eldridge et al., 2002; Gardiner, 2008), how Know responses are conceptualised varies greatly across studies. In their review of the Remember-Know literature, Gardiner and Richardson-Klavehn (2000) identified interpretation of Know responses as "the most vexatious problem in the remember/know paradigm" (p. 238). This problem has arisen because researchers' definitions of Knowing vary in terms of the emphasis put on familiarity or certainty (Geraci, McCabe, & Guillory, 2009); and relatedly, whether Know responses are interpreted as reflecting an underlying process of familiarity or a state of knowing.

In the current experiments participants are asked to assign responses to one of four categories of subjective experience: Remember (R), Know (K), Familiar (F), and Guess (G). Full definitions provided in Table 1. The separation of the subjective experiences of Know and Familiar was first performed by Conway, Gardiner, Perfect, Anderson, and Cohen (1997). In their study of student learning, Conway et al. found that for some multiple-choice questions (MCQ) students reported that their subjective state for an answer was one of neither recollection or familiarity, instead they felt that they 'just knew' the answer. Across learning, Conway et al. demonstrated evidence of a shift from Remembering to Knowing amongst higher performing students. Conway et al. discuss this 'R-to-K shift' as reflecting a change in knowledge representation from episodic to semantic memory brought about by loss of episodic details from memory and the emergence of conceptual organisation. Evidence of an R-to-K shift using the four categories of subjective experience (R, K, F, G) has been demonstrated in similar studies of

student learning by Barber, Rajaram, and Marsh (2008) and Herbert and Burt (2001, 2003, 2004) and a study exploring learning of rare word definitions by Dewhurst, Conway, and Brandt (2009). Separated Know and Familiar categories have also been utilised with complex materials in studies of face recognition (Wright & Sladden, 2003) and eyewitness identification (Sauerland & Sporer, 2009).

[Insert Table 1 about here]

One might argue that the separation of Know and Familiar would only apply to learning of conceptually rich materials over long time periods (e.g., Conway et al., 1997; Herbert & Burt, 2001, 2003, 2004). In laboratory recognition tasks with lists of unrelated words and short study-test intervals, the experiential state of 'Knowing' may not appear to be applicable as integration of the studied information into any body of knowledge might not be able to occur. However, Dewhurst et al. (2009) used definitions of rare words precisely because these were unrelated facts that would be more difficult to integrate into a wider schema. Dewhurst et al.'s (2009) participants assigned 10% of items to Know at first test – analogous to a single-time-point recognition experiment. Although this is lower than the 22% reported by Conway et al., it demonstrates that even for less meaningful materials, participants judged the Know response to reflect their experiential state for some items only five minutes post-study with no opportunity for integration into wider conceptual organisation[1]. These 10% of unrelated definitions were 'just known' without recollection or feelings of familiarity about the study episode only a short time after said study episode. This suggests that separate categories of Know and Familiar might well be applicable to the subjective experiences in typical episodic recognition paradigms where the materials have lower intrinsic meaning (e.g., lists of unrelated words).

Other researchers have also utilised modified RK procedures. Montaldi, Spencer, Roberts, and Mayes (2006) trained participants to distinguish between recollection and three levels of familiarity: very weak familiarity, moderate familiarity, and strong familiarity. Similarly, Yonelinas, Otten, Shaw, and Rugg (2005) instructed participants to assign recollected items to Remember and non-recollected items to four levels of confidence. While these procedures do not map directly on to the four categories of subjective experience used in the present experiments, they do demonstrate that other researchers are exploring alternatives to the dichotomous RK paradigm and thus support the underlying theoretical suggestion that there are additional states of subjective awareness accessible for study (McCabe, Geraci, Boman, Sensenig, & Rhodes, 2011; see also Palmer, Brewer, McKinnon, & Weber, 2010; (e.g., Sauerland & Sporer, 2009; Strong, 1913; Wright & Sladden, 2003). A key question in the current experiments is whether the separated categories of Know and Familiar have any utility in standard episodic recognition memory paradigms, and indeed whether differences exist between Know and Familiar in the same way that differences exist between Remember and Know. In particular, the separation of Know and Familiar subjective experiences is pertinent to the issue of whether items assigned to Know can be recognised with accurate source. Research comparing subjective experience and confidence judgments with judgments of source is discussed in relation to Experiments 2 and 3 which include source judgments.

Comparing subjective experience and confidence judgments

Traditionally confidence was operationalized as a two-category scale of Sure-Unsure and proportion of items assigned to these two categories were compared against proportion assigned to Remember and Know. Gardiner and Java (1990) thus compared subjective experience and confidence to words and non-words and found that more Remember judgments were given to

words while more Know judgments were given to non-words. This interaction between word type and response was not found when Sure-Unsure judgments were made. Similar patterns were demonstrated by by Rajaram (1993), Rajaram et al. (2002), and Tunney and Fernie (2007; with the addition of a Guess category). Overall, high confidence or 'Sure' responses were made more often than Remember judgments, indicating more leniency in confidence compared to subjective experience judgments. Furthermore, patterns across experimental manipulations demonstrated that while judgments of confidence and subjective experience may be interrelated, the two judgment types are not "experimentally interchangeable" (Rajaram et al., 2002, p. 234).

Remember-Know and confidence have also been compared using a larger scale to measure confidence and responses plotted using Receiver-Operating Characteristic (ROC) curves. The consistent finding from these studies has been that higher confidence is associated with Remember responses compared to Know responses (Rotello, Macmillan, & Reeder, 2004; Rotello, Macmillan, Reeder, & Wong,,2005; Slotnick, 2010; Wixted & Stretch, 2004; Wixted & Mickes, 2010; Yonelinas, 2001a; Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996; see also Tulving, 1985). However, distribution of Remember and Know responses across confidence levels is influenced by how conservative the definitions of Remembering are (Rotello et al., 2005; Slotnick, 2010), suggesting that subjective experience and confidence are not equivalent measures. Whilst patterns of ROC curves are interesting for the single- versus dual-process debate, the manner in which confidence and subjective experience are operationalised in ROC studies means that participants' response distributions for the two cannot be directly compared as was done when Sure-Unsure were compared against Remember-Know.

In the current experiments, in order to directly compare participants' use of subjective experience and confidence judgments in a manner akin to the early studies of Gardiner and Java (1990) and Rajaram (1993), the four categories of subjective experience (Remember, Know, Familiar, and Guess) were operationalized alongside four levels of confidence (High, Medium, Low, and None). As we were interested in the influence of one judgment on another judgment, this operationalization allows the influence of one response on a subsequent response to be examined when the parameters of each judgment type are equal.

The influence of one judgment on another judgment

To date, no studies have explored the influence that confidence and subjective experience judgments may have on each other when the two judgments are reported post recognition. The traditional viewpoint on this subject is summarised by Holmes, Waters, and Rajaram (1998) who stated that "…multiple judgments might affect performance in unforeseen ways" (p. 1031) and Humphreys et al. (2003) who thought that "When two judgments are made on the same trial, the knowledge that a second judgment will be required could distort the first judgment" (p. 806). This issue has also recently been considered by Bruno and Rutherford (2010):

> It is our belief that collecting confidence ratings when also R–K(–G) judgements are provided is problematic because these measures may not be independent of each other. R–K(–G) judgements may directly influence confidence ratings (e.g., 'if I *remember* an event, then I must assign a higher rating') and, vice versa, confidence ratings may exert an influence on the R–K(–G) judgement (e.g., 'if I give a high rating, then it must be *remember'*). (italics in original; Bruno & Rutherford, 2010, p. 127)

These comments highlight that this belief has existed in the Remember-Know paradigm across at least the last two decades but has not yet been subject to empirical testing.

In line with the traditional viewpoint of Holmes et al. (1998), previous experiments comparing subjective experience and confidence judgments have employed measures to ensure that the two judgment types did not influence one another. In the original subjective experience and confidence judgment comparisons by Gardiner and Java (1990) and Rajaram (1993), judgment type was compared across two separate experiments. Rajaram et al. (2002) later replicated these findings using a within-subjects design, but even in this design testing sessions employing the two types of judgments took place one week apart and subjective experience was always tested first, "…to ensure that confidence judgments did not contaminate the remember-know judgments with a carryover effect to the second session" (p. 230). Although the same participants were making both types of judgment, Rajaram et al. remained concerned about how one judgment might influence the other.

Reaction time data associated with confidence and subjective experience judgments

As with the influence of judgments on each other, reaction times to make post-recognition judgments have not before been tested experimentally. In two-step procedures, when examining reaction times to items recognised with differing levels of confidence or subjective experience the standard analysis is to compare recognition decision RTs for items categorised by their subsequent judgment category. In one-step procedures it is the judgment of subjective experience or confidence which is timed, but this judgment is made in conjunction with the recognition decision so the timings of the two decision processes are indistinguishable. However, reliable differences have been demonstrated for both judgment types using both these types of RT analysis.

Recognition is faster when confidence is higher[2]. Mandler and Boeck (1974) used a two-step Yes/No then 3-point confidence scale procedure and found recognition decisions to highest confidence items to be approximately 1200ms faster than recognition decisions to items given the lowest confidence ratings. Using a one-step 6-point confidence recognition scale this pattern was replicated by Ratcliff and Murdock (1976) who demonstrated a 1000ms (approx.) difference between highest and lowest confidence RTs. Recent evidence from neuroimaging has also produced similar findings. For example, Henson, Rugg, Shallice, and Dolan (2000) demonstrated an advantage in RT of approximately 600ms for high confidence compared to low confidence recognition judgments.

Recognition RTs are also faster for Remember items than for Know items (Dewhurst & Conway, 1994; Dewhurst, Hitch, & Barry, 1998, Dewhurst, Holmes, Brandt, & Dean, 2006; Henson, Rugg, Shallice, Josephs, & Dolan, 1999; Vilberg & Rugg, 2007; Wixted & Stretch, 2004). Dewhurst et al. (2006) demonstrated this finding to be true when both one-step and two-step procedures are used, and also found reliable differences in recognition RTs when stimulus characteristics are varied; for example a larger RT advantage for Remember responses for words compared to non-words. However, differences in Remember and Know recognition RTs can be much smaller than those observed for confidence, for example Wixted and Stretch (2004) reported a difference of only 80ms to be reliable. In comparison, other researchers have found as large an RT advantage for Remember over Know responses as that obtained for high compared to low confidence; for example a 730ms advantage was reported by Dewhurst et al. (2006) and an 830ms advantage by Henson et al. (1999).

Examining subjective experience and confidence together Rotello and Zeng (2008) found recognition decisions for Remember items to be faster than for Know items. However, when

Remember and Know responses were equated for confidence the RT advantage for Remember responses was reduced, leading Rotello and Zeng to conclude that Remember and Know responses result from a single underlying memory process. Further interpretations of RT data in terms of theory are reserved until the General Discussion.

So why have judgment RTs not been explored before? One methodological reason is that the participant often has to wait till a prompt appears on the screen before they can make their judgment (Wilding & Rugg, 1996). The assumption perhaps being that the participant would have decided on their response prior to the prompt and therefore no difference in RTs would be observed. In the current experiments this methodological issue was avoided through counterbalancing of judgment order and the inclusion of 'catch' trials where participants were asked to make other non-critical judgments for the item. On these trials, the first judgment made by participants was one critical to the experiment (e.g., either confidence or subjective experience in Experiment 1); however, the second judgment was either a judgment of age of acquisition (AoA) or of pleasantness (rather than the other 'critical' judgment of subjective experience, confidence, or source – Experiments 2 and 3 only). As participants would not have known which judgment they would be required to make first post-recognition, and they certainly would not know what type of judgment would be coming second, the issue of pre-prompt decisions being made for judgments is reduced. Participants could not systematically predict which judgment they would be making.

As the experiments presented here compared two judgments made post-recognition, a methodological reason for analysing judgment RTs in addition to recognition RTs classified by subsequent judgment is that both judgments made for each item are dependent on the same recognition decision and therefore this RT could not be used to compare judgment types across

judgment order. A theoretical reason for analysing judgment RTs is that whilst recollection and high confidence have each been shown to increase the speed at which a separate recognition decision is made, research has not yet examined whether these can influence RT for decisions which are independent of recognition. This separation of recognition decision time and judgment decision time also enables comparison of judgment types when employed in identical paradigms as in the literature RTs vary between studies. Here RT to make subjective experience and confidence judgments can be directly compared for the first time. It is predicted that confidence judgments will be made more quickly than judgments of experience as it is suggested that the consideration of whether anything is recollected about the encoding of the item will take longer than simply thinking about the confidence associated with recognition. RT for judgments is also interesting from an influence point of view. In addition to examining whether making a Remember judgment always leads to a secondary judgment of high confidence, for example (Bruno & Rutherford, 2010), RTs for judgments can explore whether making a Remember judgment also influences the speed at which that later confidence judgment is made, and vice versa.

The current experiments were designed to explore whether when two judgments were made post-recognition the response to one judgment influenced the other judgment. The examination of influence of subjective experience and confidence judgments on each other is entirely novel, as is examination of RTs to make judgments that occur after, and are independent of, recognition. The three experiments presented here are identical in procedure except that the two post-recognition judgments differ across experiments. In Experiment 1, confidence judgments are compared with judgments of subjective experience.

Experiment 1

In this experiment participants studied medium-frequency words and at test made an Old-New recognition judgment followed by two post-recognition judgments. The two critical judgments made in this experiment were confidence and subjective experience. For both judgment types four response options were provided: Remember, Know, Familiar, and Guess for subjective experience judgments, and 3 (High), 2 (Medium), 1 (Low), and 0 (None) for confidence judgments. The order in which participants made these judgments was counterbalanced across items.

On the basis of previous findings it was predicted that participants would use confidence judgments more leniently than subjective experience judgments resulting in a greater proportion of High confidence responses than Remember responses (Gardiner & Java, 1990; Rajaram, 1993; Rajaram et al., 2002; Tunney & Fernie, 200). However, as previous comparisons of subjective experience and confidence have not utilised four categories of each judgment type specific predictions comparing other categories of subjective experience and confidence are not made. It is assumed that the two judgment types are not "experimentally interchangeable" (Rajaram et al., 2002, p. 234) and to examine this, distribution patterns, accuracy, and RT data for the four categories of subjective experience and confidence were explored. If confidence and subjective experience are independent then it would be expected that a fully counterbalanced design would reveal dissociations between the two types of judgment on these key dependent variables.

With regard to the influence of one judgment on the other, the order of judgments was examined. It is possible that Bruno and Rutherford's (2010) suggestion is correct and that after making a Remember judgment participants will be more inclined to make a judgment of High

confidence, for example; but comparing the order of judgments allows the reverse pattern to also be explored – whether more Remember judgments are made after a judgment of High confidence. Influence is also able to be explored in the RT analysis. As it is considered that subjective experience gives rise to confidence, not the other way around (Gardiner, 2001), it is suggested that confidence judgments that are made following a judgment of subjective experience may be speeded as assessment of recollection and familiarity has already occurred when the subjective experience judgment was made.

Method

Participants and Design

Participants were 44 undergraduate students (38 female), mean age 19.25 years (range 18 to 23), from the University of Leeds, who received either participant credit or £5 for taking part. Participants were tested in groups of between three and fourteen at individual PCs. Data from two further participants could not be analysed due to computer problems during testing. The experiment employed a within-subjects design and for the study and test phases all instructions and stimuli were presented, and data collected, using E-Prime version 1.2.

Materials

Information concerning materials and procedures are presented here in detail for all experiments in this paper. Critical methodological differences are then discussed for Experiments 2 and 3. The target words were 56 medium-frequency words[3] (mean familiarity rating of 424; range 350-480) limited to between five and eight letters in length. Targets were matched with 56 words to be used as lure items in the recognition test, and 16 filler words. In total participants studied 64 words (56 targets plus 8 fillers, 4 shown at the start and 4 at the end of the study phase) and performed the recognition test on 128 words[4]. This 128 included the 8

studied fillers and a further 8 lure fillers; all fillers were shown at the start of the test phase to acquaint the participants with the procedures for making recognition decisions and judgments. All fillers were excluded from analysis.

Participants made four types of judgment in this experiment: subjective experience judgments (Remember, Know, Familiar, Guess), confidence judgments, age-of-acquisition (AoA) judgments, and pleasantness judgments. Participants were always required to make either a subjective experience judgment or a confidence judgment first, followed by another type of judgment (of the three not already made). The critical trials were those where participants made a combination of a subjective experience judgment and a confidence judgment, either subjective experience then confidence or confidence then subjective experience. The AoA and pleasantness trials were included in an attempt to avoid participants guessing the manipulation of interest and so that participants were not able to predict what judgment would follow the first. Assigned to the 64 studied words (56 targets, 8 fillers) were 16 trials of each judgment pairing, either: subjective experience + confidence, confidence + subjective experience, subjective experience + AoA/pleasantness (eight of each), and confidence + AoA/pleasantness (eight of each). Words were randomly assigned to judgment pairings, though it was ensured that filler words were always followed by a non-critical judgment pairing (involving AoA or pleasantness). Lure words were assigned to judgment pairings in an identical manner to target words. AoA and pleasantness responses were not analysed.

Four word lists were utilised. Two sets of 64 words were counterbalanced as lure and target lists across participants. Two versions of each were constructed where each word was matched with a different judgment pairing and which also had differing inter-trial-intervals (ITI)

and inter-judgment-intervals (IJI) for each word in the test phase. Allocation of participants to list was random.

Procedure

In the study phase, target words were presented in random order individually in the centre of the computer screen. Participants were instructed that they should study the words and that later they would undergo an Old-New recognition test. It was emphasized that it was a long list of words and that participants should concentrate on the screen throughout. Participants studied each word for four seconds separated by a fixation point shown for 750 milliseconds. The eight filler words were shown in random order at the start and the end of the target word list. Prior to undergoing the recognition test, participants completed pattern comparison and letter comparison speed of processing tasks (e.g., Salthouse, 1991). Although not formally timed, participants took approximately five minutes to complete this.

In the test phase, participants were instructed that for each word recognised as Old they would then be asked to make two judgments from the list of four judgments. The four judgments were then explained on individual screens (with full definitions of Remember (R), Know (K), Familiar (F), and Guess (G) provided on paper; see Table 1). Each explanation screen included an example of the scale that the participant would use for each type of judgment. The subjective experience judgment scale consisted of four response boxes labelled (from left to right) R, K, F, and G, with the question 'What is your EXPERIENCE of recognising this word?'. The confidence scale consisted of four boxes labelled (from left to right) 0, 1, 2, 3, with 0 accompanied by the label 'Not confident at all' and 3 accompanied by 'Extremely confident', and the question 'How CONFIDENT are you that you correctly recognised this word?'. For ease of description in discussing analyses, these numerical

confidence levels are translated into verbal labels: 3 = High, 2 = Medium, 1 = Low, 0 = None. The labelling of the subjective experience scale 'R, K, F, G' from left to right and the confidence scale '0, 1, 2, 3' from left to right was purposefully done to ensure that on trials involving both these judgments, if assigning the item to both Remember and High confidence for example, participants were not moving the mouse to exactly the same position to make both responses. Between judgments the mouse pointer was repositioned to the centre of the screen. The AoA judgment scale consisted of four boxes labelled 0-4, 5-8, 9-12, and 13+, and the question 'What AGE were you when you first learnt this word?'. Finally the pleasantness scale consisted of the question 'How PLEASANT is this word?' and four boxes labelled 1, 2, 3, and 4, with 1 accompanied by the label 'Not pleasant' and 4 accompanied by 'Very pleasant'. For each judgment type, the response boxes were the same size, shape and positioned identically.

Commencing with the 16 filler words, participants then underwent the recognition test. At the start of a trial a fixation point was shown for 750ms. Each word was then presented in the centre of the screen with the cues 'New' and 'Old' presented below and towards the left and right of the screen respectively, reflecting the number keys which participants were required to use to make their Old-New judgment: 1 for New, 2 for Old. Participants were instructed to press these keys with their left hand. Reminders for which numbers corresponded to Old and New were shown on all trials. When participants indicated that they recognised a word as Old the first judgment screen was shown. The word re-appeared in the centre of the screen with the response boxes and judgment question below. The mouse pointer appeared in the centre of the screen and participants made all judgments using the mouse to click on one of the response boxes. Speed and accuracy of responses were emphasised in the instructions for the test phase and response times for all recognition decisions and judgments were recorded. In between the

two judgment screens and prior to a new trial commencing a blank screen was shown as an IJI or ITI (for a duration which ranged from 250 to 1250ms). Every twelve trials during the test phase participants were allowed to take a break if they were fatigued. A 'take a break' slide appeared and remained until the participant pressed the spacebar. After completion of the test phase participants completed a debrief questionnaire to ensure that they had understood the different types of judgments made; no participants were excluded on the basis of their responses.

Results and Discussion

Memory performance was examined, followed by the proportion of correct responses assigned to each of the confidence levels and subjective experience categories calculated using both *a priori* and *a posteriori* methods. The influence of First Judgment on Second Judgment was then explored using *a priori* and *a posteriori* methods. Reaction times for recognition decisions and judgments were analysed by confidence level or subjective experience category. Prior to analysis, data from any responses that had been made faster than 300ms or slower than 8000ms were excluded from the dataset. Data from Guess and None responses was excluded from all ANOVA to avoid violating the assumption of independence in analysis of *a priori* proportional data and to increase listwise N in analysis of RT and *a posteriori* accuracy data (though data from Guess and None are included on figures). Greenhouse-Geisser values are reported when ANOVA did not meet assumptions of sphericity. The central findings from each analysis are summarised within the following sections and more substantive theoretical discussion is carried out in the General Discussion.

Memory measures

Memory performance measures are shown in Table 2. Overall recognition performance (M = .54) was significantly higher than chance (zero), $t(43) = 17.86$, $p < .001$. As participants made FAs to only 8% of lures, analysis of lure data was not performed.

[Insert Table 2 about here]

Subjective experience and confidence measures

The level of confidence or type of subjective experience participants had based their recognition decisions on was examined next. Firstly *a priori* proportions were calculated – the proportion of correct responses assigned to each of the confidence levels or subjective experience categories; see the upper panel of Figure 1. A 2(judgment type) x 2(judgment order) x 3(response category) within-subjects ANOVA demonstrated no main effect of judgment type, $F(1,43) = 1.48$, $p = .23$; or judgment order, $F < 1$. The main effect of response category was significant, $F(1.55,66.55) = 41.34$, $p < .001$; and there was a significant interaction between judgment type and response category, $F(1.43,61.61) = 13.61$, $p < .001$. As shown in Figure 1, nearly 60% of responses were assigned to High confidence, while just over 40% of responses were assigned to Remember. Approximately equal proportions of responses were assigned to Know and Familiar, whilst slightly more responses were assigned to Medium compared to Low confidence.

[Insert Figure 1 about here]

There was no interaction between judgment order and response category $F(1.68,72.41) = 1.61$, $p = .21$; or judgment type and judgment order, $F < 1$; and the three-way interaction was not significant, $F < 1$. As there was no main effect of judgment order, judgment order did not interact with response category, and there was no three-way interaction, further analysis was

conducted with data for First and Second Judgments combined. Confidence and subjective experience do not appear to differ as a function of which judgment type is made first.

To examine the significant interaction between judgment type and response category, separate 3(response category) ANOVAs for subjective experience and confidence were conducted with responses aggregated across judgment order. Both ANOVAs demonstrated significant main effects of response category: subjective experience, $F(1.63,69.94) = 7.77$, $p = .002$; and confidence, $F(1.35,58.10) = 75.64$, $p < .001$. For subjective experience, comparisons between the different response categories demonstrated no significant difference in the proportion of responses assigned to Know and Familiar, $t < 1$. All other comparisons were significant, all at least $p < .007$; more responses were assigned to Remember than to Know or Familiar. For confidence judgments all comparisons were significantly different, all at least $p < .006$. More responses were assigned to High compared to Medium or Low confidence, and more were assigned to Medium than to Low confidence. Comparing across judgment type, more responses were assigned to High confidence than to Remember, $t(43) = 4.87$, $p < .001$, and fewer responses were assigned to Low confidence than were assigned to Familiar, $t(43) = 6.35$, $p < .001$.

The observed patterns suggest that while order does not influence the relationship between one judgment type and another, there are not straightforward mappings between confidence and subjective experience judgments. For instance, Remember judgments do not merely reflect the highest level of confidence; more items were assigned to High confidence than to Remember.

To explore the level of accuracy associated with each response category *a posteriori* probabilities were calculated. These probabilities were based on First Judgments only as for

each item both First and Second Judgments were conditional on the same recognition response. Analysis was conducted separately for subjective experience and confidence judgments as the main comparison of interest was accuracy of response categories within judgment type.

Separate 3(response category) ANOVAs demonstrated significant main effects of response category for subjective experience judgments, $F(1.60,62.54) = 23.79$, $p < .001$; and confidence judgments, $F(1.57,59.54) = 11.26$, $p < .001$; data shown in the top panel of Figure 2. For subjective experience judgments, no significant difference was found in the accuracy of responses assigned to Remember or Know, $t < 1$, and both Remember and Know responses were significantly more likely to be correct than were responses assigned to Familiar, both $p < .001$. Contrastingly, for confidence judgments, responses assigned to High confidence were significantly more likely to have been correct than responses assigned to Medium or Low confidence, both at least $p = .001$; but Medium confidence responses were no more accurate than Low confidence responses, $t(38) = 1.36$, $p = .18$.

[Insert Figure 2 about here]

In sum, patterns of distributions of responses and accuracy of responses revealed that participants did not use confidence and subjective experience judgments in the same way. Participants were more lenient in their use of confidence judgments than subjective experience judgments and more responses were assigned to High confidence than to Remember. This fits with previous findings (Gardiner & Java, 1990; Rajaram, 1993; Rajaram et al., 2002; Tunney & Fernie, 2007) but demonstrates this without the inclusion of a manipulation such as a word/non-word comparison or masked priming. In the current experiment, confidence and subjective experience judgments produced different patterns on identical stimuli. Accuracy analysis further demonstrates the different ways participants used the two judgments. A linear

relationship was demonstrated between confidence and accuracy, whereas the relationship was not linear for subjective experience judgments as Know recognition decisions were just as accurate as Remember recognition decisions. This latter finding replicates Dewhurst et al. (2009) and Sauerland & Sporer (2009). With regards to the order in which the judgments were made, the *a priori* analysis provides initial evidence that the order of the two judgment types did not influence distribution patterns across response categories.

The influence of one judgment on the other

To fully explore whether making two judgments per item led to one judgment influencing the other judgment, conditional probabilities were calculated on correct recognition decisions in both temporal orders. Firstly, the proportion of items assigned to Remember, Know, Familiar, and Guess at Second Judgment was calculated split by which confidence level the item had been assigned to at First judgment; i.e., *'of the items assigned to High confidence at First judgment, what proportion were then assigned to each of the categories of subjective experience at Second Judgment?'.* As shown in the upper panel of Figure 3, if an item had been assigned to High confidence at First Judgment then .61 of these items were then assigned to Remember at Second Judgment, .36 were assigned to Know, and .03 were assigned to Familiar. Contrastingly, if an item was assigned to Medium confidence at First Judgment then at Second Judgment .24 were assigned to Remember, .29 to Know, and .47 to Familiar; and so on.

These proportions were also calculated in the opposite direction. In the counterbalanced design, participants either judged confidence followed by subjective experience, or the order of these two judgments was reversed. Data were next examined by calculating the proportion of items assigned to Remember, Know, Familiar, and Guess at First Judgment split by which confidence level the item had later been assigned to at Second judgment; i.e., *'of the items that*

were assigned to High confidence at Second Judgment, what proportion of these had come from a Remember First Judgment?'. Mean proportions are shown in the lower panel of Figure 3.

Comparing the upper and lower panels of Figure 3 it is evident that the order in which judgments were made did not influence the relationship between confidence and subjective experience; the similarities across the two panels are plainly evident. For example, in the lower panel, if an item had been assigned to High confidence at Second Judgment then .64 of these items had previously been assigned to Remember at First Judgment, compared to .61 when the judgments were in the opposing order as shown in the upper panel. Comparing across panels, within each level of confidence the patterns of means are paralleled for each category of subjective experience.

Separate 2(judgment order) x 3(response category) ANOVAs for items assigned to each of the confidence levels compared the distribution of subjective experience responses at each confidence level across the panels of Figure 3; i.e., how correct recognition responses that were assigned to High confidence were distributed across subjective experience categories whether judgment order was confidence then subjective experience or the reverse. At each level of confidence, ANOVA demonstrated no main effect of judgment order, all $F < 1$; a significant main effect of response category: High, $F(1.08,45.37) = 39.36$, $p < .001$, Medium, $F(2,58) = 2.60$, $p = .083$ – though this effect only approached significance, and Low, $F(1.22,30.60) = 90.48$, $p < .001$; and no interaction between judgment order and response category, all $F < 1$. ANOVA was not performed on responses with a confidence level of None as the listwise N was 4; less than 5% of responses had been assigned to None.

[Insert Figures 3 and 4 about here]

In sum, as there was no interaction between judgment order and subjective experience response at any level of confidence, these data support the conclusion that the relationship between confidence and subjective experience was not influenced by the order in which the two types of judgment were made. However, conditional proportions were also able to be calculated in the opposing way. Data for proportion of items assigned to each level of confidence was split by the category of subjective experience to which the item had been assigned.

Firstly, the proportion of items assigned to High, Medium, Low, and None at Second Judgment was calculated split by which category of subjective experience the item had been assigned to at First judgment; i.e., *'of the items assigned to Familiar at First judgment, what proportion were assigned to each of the confidence levels at Second Judgment?'.* Means are shown in the upper panel of Figure 4. These proportions were also calculated in the opposite direction. The proportion of items assigned to High, Medium, Low, and None at First Judgment was calculated split by which category of subjective experience the item had later been assigned to at Second judgment. Means are shown in the lower panel of Figure 4. Comparison of the upper and lower panels of Figure 4 again demonstrates that the order in which judgments were made did not influence the relationship between confidence and subjective experience. However, comparison of Figure 4 and Figure 3 shows that very different patterns are produced by the two different methods of calculating conditional probabilities.

Separate 2(judgment order) x 3(response category) ANOVAs for items assigned to each of the four categories of subjective experience compared the distribution of confidence judgments within each category of subjective experience across the upper and lower panels of Figure 4.

For Remember responses, ANOVA demonstrated no main effect of judgment order, $F(1,37) = 1.00$, $p = .32$; a significant main effect of response category, $F(1.00,37.09) = 98.95$, $p < .001$; and no interaction between judgment order and response category, $F < 1$. For Know responses, ANOVA demonstrated no main effect of judgment order, $F < 1$; a significant main effect of response category, $F(1.04,30.14) = 37.98$, $p < .001$; and no interaction between judgment order and response category, $F < 1$. For Familiar responses, ANOVA demonstrated a main effect of judgment order which approached significance, $F(1,31) = 3.68$, $p = .064$; a significant main effect of response category, $F(1.67,51.80) = 12.84$, $p < .001$; and no interaction between judgment order and response category, $F < 1$. ANOVA was not performed on Guess responses as the listwise N was 6; less than 5% of responses had been assigned to Guess.

In sum, both methods of calculating these conditional probabilities reveal identical patterns within judgment type and no influence of judgment order; however, comparing across which judgment type was used to split the analysis demonstrates that participants did not simply map the four levels of confidence onto the four categories of subjective experience, they used the two judgment types in very different ways. Comparison of Figures 3 and 4 shows that while .61 of items that were initially assigned to High confidence went on to be assigned to Remember, when order of judgments was reversed, .83 of items initially assigned to Remember later went onto be assigned to High confidence. Participants were not 100% confident about an item they Remembered; and if they did have High confidence for an item it did not mean they Remembered it. Comparisons of Familiar subjective experience and Low confidence demonstrate that it was not only Remember and High confidence which did not map onto each other. If an item was judged to be Familiar at First Judgment, .45 and .41 were later assigned to Medium and Low levels of confidence respectively (Figure 4, upper panel). Conversely, if an

item was first assigned to Low confidence, .74 then went on to be assigned to Familiar at Second Judgment and .21 went on to Guess (Figure 3, upper panel). Items assigned to Low were often then assigned to Familiar, but not exclusively; and items initially assigned to Familiar were equally likely to be assigned to either Low or Medium levels of confidence.

The above figures are illustrative of the different relationships and influence between successive post-recognition judgments. One further way to analyse this data was to calculate mean confidence ratings for items assigned to Remember, Know, Familiar, and Guess. This analysis utilised the original numerical confidence levels of 3, 2, 1, and 0 that participants used to make their responses instead of the verbal labels High, Medium, Low, and None. As order of judgment had been shown not to influence patterns of responses above, mean confidence ratings were calculated irrespective of judgment order. A 3(response category) ANOVA comparing confidence ratings to Remember, Know, and Familiar items demonstrated a significant main effect of response category, $F(2,76) = 101.41$, $p < .001$. Items assigned to Remember ($M = 2.75$, $SD = .44$) or Know ($M = 2.65$, $SD = .38$) were both given higher confidence ratings than items assigned to Familiar ($M = 1.52$, $SD = .51$): $t(38) = 12.41$, $p < .001$ and $t(38) = 12.73$, $p < .001$ respectively. However, there was no significant difference between the confidence ratings given to Remember or Know items, $t < 1$. Though not included in ANOVA, mean confidence for Guess items was 0.81 ($SD = .57$).

In sum, exploration of the influence of one judgment type on the other further demonstrated that participants did not simply map the four levels of confidence onto the four categories of subjective experience. For example, over 80% of Remember responses were also assigned to High confidence, but when proportions were calculated based on confidence response, only approximately 60% of items which were assigned to High confidence were also

Remembered. The two main points to take from this analysis are that participants did not simply assign all Remember items to High confidence, Know items to Medium confidence, and Familiar items to Low confidence and vice versa; however, the patterns obtained for assignment of responses were matched across judgment order. While all items that were given Remember responses at First Judgment were not assigned to High confidence at Second judgment, the patterns obtained in that judgment order matched the patterns of responding when the order of judgments was reversed. Whilst subjective experience and confidence judgments are no doubt related, concerns such as those voiced by Bruno and Rutherford (2010) regarding the influence of one judgment on the other do not have credence. Here responses at Second Judgment were not influenced by what response had been made at First Judgment.

Regarding categories of subjective experience, analysis of mean confidence demonstrates that items that were assigned to Remember and Know did not differ in terms of confidence; adding to the previous finding that Remember and Know did not differ in terms of accuracy. However, both accuracy and confidence ratings differentiated Know from Familiar responses. The issue of the separation of Know and Familiar subjective experiences is returned to in the General Discussion.

Reaction times for recognition decisions

As discussed earlier, the typical form of analysis to explore reaction time is to analyse RTs for recognition decisions split by which subjective experience category the item was later assigned to. In the current experiment this analysis was performed on correct recognition RTs classified by the confidence or subjective experience response category the item was then assigned to at First Judgment. Mean correct recognition decision RTs are shown in the upper panel of Figure 5.

[Insert Figure 5 about here]

A 2(judgment type) x 3(response category) ANOVA demonstrated no main effect of judgment type, $F(1,32) = 1.93$, $p = .17$; a significant main effect of response category, $F(1.38,44.24) = 38.97$, $p < .001$; and no interaction between the two, $F(1.53,48.88) = 1.06$, $p = .35$. Separate 3(response category) ANOVAs for confidence judgments and subjective experience judgments demonstrated significant effects of response category for both: subjective experience, $F(2,76) = 23.04$, $p < .001$; and confidence, $F(1.70,59.66) = 16.42$, $p < .001$. For subjective experience judgments, correct recognition decisions to items that were categorised as either Remember or Know were made significantly faster than those for items later categorised as Familiar, both $p < .001$. There was no significant difference in the speed of Remember and Know recognition decisions, $t < 1$. For confidence judgments, High and Medium confidence recognition decisions were both made significantly faster than Low confidence recognition decisions, both at least $p = .001$. Additionally, recognition decisions made with High confidence were made more quickly than those of Medium confidence, this difference approached significance, $t(35) = 1.89$, $p = .067$. Whilst High and Medium confidence responses were not matched on accuracy as Remember and Know responses were, for recognition RTs patterns are matched across the two judgment types.

In sum, recognition RT analysis demonstrated a case where confidence and subjective experience responses did not differ. There was no difference in speed of recognition when items were later categorised as Remember or Know, and recognition of these items was faster than items later assigned to Familiar or Guess. Similarly, recognition of items later classed as High or Medium confidence was equally fast and was faster than for items assigned to Low or None. Taken together with the findings regarding accuracy, where a significant difference between

High and Medium confidence responses was demonstrated, this again highlights that subjective experience and confidence judgments behave in different ways: High and Medium confidence recognition decisions are differentiated by accuracy but not RT, while Remember and Know recognition decisions are not differentiated by accuracy or RT.

Reaction times for subjective experience and confidence judgments

A novel approach to RT analysis here was how quickly participants made their actual judgments of confidence and subjective experience. Firstly, the overall mean RTs for confidence and subjective experience judgments to correct recognition decisions were compared irrespective of response category; data shown in Table 3. A 2(judgment type) x 2(judgment order) ANOVA demonstrated a significant main effect of judgment type, $F(1,43) = 25.51$, $p < .001$; confidence judgments were made more quickly than judgments of subjective experience. The main effect of judgment order approached significance, $F(1,43) = 3.04$, $p = .09$, as for both judgment types Second Judgments were slightly faster than First Judgments. The interaction between judgment type and judgment order was not significant, $F < 1$.

[Insert Table 3 about here]

Further analysis examined whether differences in RT were observed for different response categories within judgment types. As there was no effect of judgment order in the 2x2 ANOVA, data were aggregated across order. Mean judgment RTs split by response category are shown in the lower panel of Figure 5. Separate 3(response category) ANOVAs for confidence and subjective experience revealed a significant main effect of response for confidence judgments, $F(1.25,50.06) = 6.61$, $p = .009$; but no main effect of response for subjective experience, $F < 1$. As shown in Figure 5, time to make subjective experience

judgments was stable at around 2000ms, whilst for confidence judgments speed increased as confidence increased.

For confidence, High confidence judgments were made significantly faster than both Medium and Low confidence judgments, at least $p = .003$. The difference in speed between Medium and Low confidence judgments was not significant, $t(40) = 1.34$, $p = .19$. For subjective experience, there were no significant differences between the time to make Remember, Know, or Familiar judgments, all at least $p > .26$.The novel analysis of comparing RT to make judgments independent of recognition RT confirmed the prediction that, on the whole, confidence judgments would be made more quickly than judgments of subjective experience. It is suggested that it is simply easier to assess memory strength than whether or not you can recollect details from the study episode and this is demonstrated via reaction time. No interaction with judgment order was demonstrated however: confidence judgments were not speeded when made following a judgment of recollective experience.

Contrasting patterns of RT were demonstrated within judgment types. Confidence judgments were made more quickly the more confident they were – paralleling the speed advantage for High confidence recognition decisions – but, as predicted, no differences in RT were demonstrated for subjective experience judgments. It is suggested that this is related to recollection being a threshold process (e.g., Yonelinas & Parks, 2007). Though Remembered items have quicker recognition decision RTs, when making the subjective experience judgment itself, if no recollection of the study episode comes to mind search is terminated and a different subjective experience response is made. The time to make a judgment of any category of subjective experience is therefore related to the time spent undertaking additional metacognitive processing. This appears to be quite a long process compared to making a confidence judgment.

So why might judgments of subjective experience take so long? If making judgments which require recalling information from the encoding phase is a slow and effortful process involving retrieval reaching the threshold of activation for recollection (e.g., Yonelinas & Parks, 2007) or mental time travel (e.g., Tulving, 1985), it might be expected that making a judgment of source will take just as long as making a judgment of subjective experience. On this basis, Experiments 2 and 3 were designed with the aim of following up the slower reaction times for subjective experience judgments demonstrated in Experiment 1. However, designing these two experiments also enabled the examination of a number of previously un-researched relationships such as the relationship between source accuracy and confidence and source judgments and Remember, Know, Familiar and Guess subjective experiences.

Experiments 2 and 3

Comparing subjective experience and source judgments

In contrast to confidence and subjective experience judgments which cannot be objectively verified, source memory judgments can be objectively scored as correct or incorrect. Many researchers have therefore combined judgments of source with confidence and/or subjective experience in attempts to tap both subjective and objective measures of recollection. The central theoretical concern that has arisen from research comparing judgments of source and subjective experience is whether accurate source should be able to be retrieved from memory when an item is assigned to Know. Previous research has produced mixed results regarding the accuracy of source for Know items. Some studies have demonstrated source accuracy to only be at chance levels for Know items, which reflects the interpretation of Know responses being recognition unaccompanied by any contextual details from encoding (Dewhurst & Hitch, 1999; Dudukovic & Knowlton, 2006; Perfect, Mayes, Downes, & Van Eijk, 1996,

except for Experiment 3). However, other studies have found source accuracy to be above chance for items assigned to Know (Conway & Dewhurst, 1995b; Hicks, Marsh, & Ritschel, 2002; Meiser & Bröder, 2002; Meiser & Sattler, 2007; Starns & Hicks, 2005). This suggests that in some cases enough contextual details from the study phase are retrieved at recognition for source to be judged correctly but not enough details are recollected for the item to be assigned to Remember.

In the above studies there were a number of methodological differences that are relevant to the current experiments. Five of the eight studies used a standard Old-New recognition paradigm with judgments of source and/or subjective experience being made to only items judged Old, whereas three used a one-step procedure by including a 'New' category with the first judgment made. Perfect et al. (1996) and Starns and Hicks (2005) asked participants to make a Remember-Know-New judgment followed by a source judgment and Hicks et al. (2002) had participants make a Left source-Right source-New judgment followed by a judgment of subjective experience. These methodological differences do not map onto the differences in findings however; the five studies which found Know items to have above chance source accuracy are not the five that incorporated a separate Old-New judgment. While one-step and two-step procedures have been shown to influence patterns of responding for Remember-Know judgments (cf., Bruno & Rutherford, 2010; Eldridge et al., 2002; Hicks & Marsh, 1999; see General Introduction) and source judgments (Dodson & Johnson, 1993; Marsh & Hicks, 1998) individually, the above studies that employed both types of judgment do not show any clear patterns regarding judgment order.

Comparing source and confidence judgments

Confidence and source judgments have not been directly compared in the same way as confidence and subjective experience: *do these judgments produce the same patterns of responding?* (Gardiner & Java, 1990; Rajaram, 1993; Rajaram et al., 2002), or in the same way as source and subjective experience have: *can source for Know judgments be accurate?* (Conway & Dewhurst, 1995b; Hicks et al., 2002; Meiser & Bröder, 2002; Meiser & Sattler, 2007; Perfect et al., 1996; Starns & Hicks, 2005). Instead, studies have typically combined the two judgments into a source-confidence judgment. For example, in four experiments Yonelinas (1999) compared recognition-confidence and source-confidence judgments. These were both measured on 6-point scales from '*1 = confident new'* to '*6 = confident old'*, and from *'1 = sure it was spoken by the female voice',* to *'6 = sure it was spoken by the male voice'.* Results demonstrated that the higher the recognition-confidence or source-confidence judgment the higher the recognition or source accuracy. In addition, the ROC curves differed for recognition-confidence and source-confidence and the shapes of the curves obtained fit with the prediction that recognition confidence can rely on both familiarity and recollection while source confidence relies primarily on recollection (Yonelinas, 1999).

Other researchers have used similar approaches, combining source-confidence measures with recognition-confidence ratings (and judgments of subjective experience) and obtained similar findings with recognition-confidence ratings being higher for accurate, and higher confidence, source judgments (e.g., Rotello et al., 2005; Slotnick, 2010). However, the shapes of their ROC curves led these authors to suggest different interpretations of recollection and familiarity. Although these findings are relevant to the ongoing debates surrounding recollection and familiarity processes, the current experiments operationalize and examine

source, confidence, and subjective experience judgments in a different manner not directly comparable to these ROC studies.

In relation to the subjective experience and confidence judgments made in Experiments 2 and 3 it is predicted that distribution patterns, accuracy, and RT will parallel those observed in Experiment 1. Distribution patterns, accuracy, and RT for these two judgment types are also then able to be explored split by whether the accompanying source judgment is correct or incorrect. As an accurate judgment of source requires retrieval of contextual information from the study episode it is predicted that correct source items will be associated with higher levels of confidence and a greater proportion of Remember responses than incorrect source items. In line with the majority of previous research it is also predicted that while source accuracy will be greatest for items assigned to Remember, a large proportion of items assigned to Know will also be accompanied by accurate source judgments (Conway & Dewhurst, 1995b; Hicks et al., 2002; Meiser & Bröder, 2002; Meiser & Sattler, 2007; Starns & Hicks, 2005). Furthermore, the use of separate categories of Know and Familiar in the current experiments permits the accuracy of source judgments for these two subjective experiences to be compared. If source accuracy differs for Know and Familiar this will add to the validity of separating these two categories of subjective experience.

The influence of one judgment on another

As with the influence of confidence and subjective experience judgments on each other, to date no direct manipulation of judgment order for source and subjective experience judgments has been published. As discussed earlier, experiments comparing source accuracy and subjective experience have employed a variety of different procedures and findings have been mixed (cf. Dewhurst & Hitch, 1999; Perfect et al., 1996). With regard to the order of

source and subjective experience judgments when made after Old-New recognition all five studies which used this procedure asked for subjective experience first and then source (though in Conway and Dewhurst (1995b) participants actually made separate RK judgments for their recognition decision and their source judgment). Although no published study has manipulated judgment order for source and subjective experience judgments, one Master's dissertation from Jason Hicks' lab has examined this manipulation.

Martin (2007, *unpublished thesis*) manipulated order of source and subjective experience judgments between participants. At study, words were presented either on the left or the right of the computer screen and at test half the participants made a Remember-Know-New judgment[5] followed by a Left-Right source judgment, and the other half made a Left-Right-New judgment followed by Remember-Know. Many of the general patterns of results replicated previous findings. Source accuracy to Remember items was higher than source accuracy for Know items; and source accuracy for Know items was higher than chance (Conway & Dewhurst, 1995b; Hicks et al., 2002; Meiser & Bröder, 2002; Meiser & Sattler, 2007; Starns & Hicks, 2005). However, findings regarding the effect of judgment order were mixed. Participants who made subjective experience judgments first gave more Remember responses to hits than participants who made this judgment following a source judgment, though this result was only marginal. For source accuracy, no main effect of judgment order was observed when source accuracy was analysed collapsed across subjective experience, though puzzlingly, when response category (Remember-Know) was included in analysis a main effect was reported with higher source accuracy when source was the first judgment made. Additionally, the discussion states that the source-first group showed better source memory for Remember items than the subjective experience-first group, however an interaction between judgment order and

subjective experience nor any separate analysis for Remember items was reported. These discrepancies are not explained and it cannot be known which of these opposing findings may be an error in analysis, or an error in reporting. While no study has explored the influence of source and confidence judgments on each other, and these very mixed findings from an unpublished dissertation should be considered with caution, it is of interest that the influence of one judgment on the other has begun to be explored.

The influence of one judgment on the other is again of key interest in Experiments 2 and 3. While no effects of judgment order were obtained in Experiment 1, it is suggested that the objective nature of source judgments could lead them to have more influence over subsequent judgments. Perhaps when a source judgment is made first and one element of source context is retrieved this may influence the participant to make a Remember or High confidence response somewhat more leniently as was demonstrated, albeit only marginally, by Martin (2007). For example, if 'only' the required Left/Right source information is retrieved the item may still be assigned to Remember even if recognition is not accompanied by any feelings of mental time travel or recollection of thoughts that came to mind during study which would have normally be required for the item to exceed that individual's Remember threshold. Additionally, the accuracy of source judgments could be influenced by whether they precede or follow a judgment of confidence or subjective experience as was suggested by some of Martin's (2007) data. To examine these suggestions patterns of judgment distribution are again explored by judgment order.

Reaction time for judgments of source

In addition to high confidence and recollection speeding recognition decisions, some research has also shown recognition to be faster for items when source is accurate compared to

inaccurate. Using a one-step source procedure where participants categorised items as either 'Old - mentally imaged at study', or 'Old - read backwards at study', or 'New', Kahn, Davachi, and Wagner (2004) found item + source recognition (Old + correct source) decisions were approximately 250ms faster than item only recognition (Old + incorrect source), but only for items which had been mentally imaged at study. Similar results were obtained by Lundstrom, Ingvar, and Petersson (2005) who also used a one-step procedure. However, opposing results have been demonstrated using a two-step procedure. Wilding and Rugg (1996) presented items in male or female voices at study and correct recognition responses were categorised by whether the subsequent judgment of source was correct. No differences in recognition RTs for correct and incorrect source items were found.

In the current two experiments, analysis of RTs to make post-recognition judgments again allows completely novel analysis to be performed: the comparison of how long it takes to make judgments of source, confidence, and subjective experience. As source judgments require assessment of retrieved contextual information it is predicted that RT for accurate source judgments will be approximately equal to those for subjective experience, which were slower than confidence judgments in Experiment 1. Whether inaccurate source judgment RTs or their prior recognition decision RTs will be slower than those for accurate source judgments will also be explored as, as discussed above, some previous research on source recognition RTs has demonstrated differences while other research has not (Kahn et al., 2004; Lundstrom et al., 2005; Wilding & Rugg, 1996).

In Experiments 2 and 3, influence is also able to be explored in the RT analysis by examining whether the subjective experience category or confidence level an item was assigned to differentiated how long it took to make an accurate source judgment. For example, it could

be suggested that accurate judgments of source accompanied by High confidence or Remembering may be made more quickly than accurate judgments of source where confidence is not so high or the recognition is based on Familiarity. Final analysis in these experiments focuses on this novel method of exploring judgment RTs.

The current experiments

The materials and basic procedures employed in Experiments 2 and 3 were identical to those in Experiment 1: Participants studied medium-frequency words and at test made an Old-New recognition judgment followed by two post-recognition judgments. However, in order that retrieval of source could be tested at recognition, words were presented for study on either the left or the right of the screen with an encoding instruction to associate words on the left with one person and words on the right with another person (after Yonelinas, 1999). The two critical judgments made post-recognition were source and subjective experience in Experiment 2 and source and confidence in Experiment 3. Subjective experience and confidence judgments were made using the same four response options as in Experiment 1 and for source judgments participants selected whether they thought the word had been studied on the Left or the Right. The order in which participants made the two critical judgments was counterbalanced across items.

Methods

Participants and Designs

Thirty-two undergraduate students (30 female) participated in Experiment 2, mean age 18.88 years (range 18 to 21). Thirty-four undergraduate students (32 female) participated in Experiment 3, mean age 18.62 years (range 18 to 20). All participants were from the University of Leeds and received Participant Pool Credits for taking part. Participants were tested in

groups of between five and thirteen at individual PCs. In Experiment 2, data from two participants was excluded from analysis as they did not follow experimental instructions. Both experiments employed within-subjects designs and for the study and test phases all instructions and stimuli were presented, and data collected, using E-Prime version 1.2.

Materials and Procedures

The materials and procedures for Experiments 2 and 3 were identical to that of Experiment 1 except changes due to the source manipulation. In both experiments, in the study phase target words were presented individually to either the left or the right of the computer screen (placement randomised). Participants were instructed that as well as trying to remember all the words, they should also try to remember on which side of the screen the word was shown. To improve performance participants were instructed to associate words on the left with one person and words shown on the right with another person (Yonelinas, 1999). Participants were told they could associate the words with anyone, e.g., celebrities/family/friends, and were asked to write the names of their chosen people on their response booklet (to check they followed this instruction correctly). All other aspects of the study and distracter phases were identical to those of Experiment 1.

In the test phases of both experiments participants were instructed that for each word recognised as Old they would then be asked to make two judgments from the list of four judgments. The four judgments were then explained on individual screens. The subjective experience (Experiment 2 only), confidence (Experiment 3 only), AoA, and pleasantness judgments and instructions were identical to those in Experiment 1. For source judgments the screen consisted of two boxes labelled ‘Left’ and ‘Right’ with the question ‘Which side of the screen was this word shown on?’ below. These two response boxes were the same size and

shape as the four response boxes required for the subjective experience and confidence judgments. For each source judgment the two response boxes were positioned below the re-presented word, equidistant from the centre of the screen.

Participants underwent the recognition test in the same manner as in Experiment 1. Each word was presented in the centre of the screen with the cues 'New' and 'Old' presented below and towards the left and right of the screen respectively, reflecting the number keys which participants were required to use to make their Old-New judgment: 1 for New, 2 for Old. When participants indicated that they recognised a word as Old the first judgment screen was shown. In Experiment 2, participants were always required to make either a source judgment or a subjective experience judgment first, followed by another type of judgment (of the three not already made). The critical trials were those where participants made a combination of a source judgment and a subjective experience judgment, either source then subjective experience or subjective experience then source. In Experiment 3, participants were always required to make either a source judgment or a confidence judgment first, followed by another type of judgment. The critical trials were those where participants made either source then confidence or confidence then source.

Results and Discussion

Analysis was conducted in the same manner as in Experiment 1 with the addition of comparisons split by source accuracy. Memory performance was examined, followed by overall source accuracy and the proportion of correct responses assigned to confidence levels and subjective experience categories using both *a priori* and *a posteriori* methods. The influence of First Judgment on Second Judgment was then explored. Reaction times for recognition decisions and judgments were analysed by source accuracy, confidence level, or subjective

experience category. Prior to analysis, data from any responses which had been made faster than 300ms or slower than 8000ms were excluded from the dataset. Data from Guess and None responses was excluded from all ANOVA to avoid violating the assumption of independence in analysis of *a priori* proportional data and to increase listwise N in analysis of RT and *a posteriori* accuracy data (though data are included on figures). Greenhouse-Geisser values are reported when ANOVA did not meet assumptions of sphericity. The central findings from each analysis are summarised within the following sections and more substantive theoretical discussion is reserved until the General Discussion. Exploratory statistical analysis comparing experiments is carried out where appropriate.

Memory measures

Memory performance measures for Experiments 2 and 3 are shown in Table 2. In both experiments memory performance was higher than in Experiment 1. The source encoding manipulation appears to have increased participants' encoding proficiency leading to higher performance. Participants made FAs to under 10% of lure items and therefore further analysis of FA data was not performed.

Source accuracy

Judgments of whether an item appeared on the left or right of the screen for study were accurate over 80% of the time in both experiments. Mean proportion of source judgments correct in Experiment 2 was .85 (*SD* = .15) at First Judgment and .84 (*SD* = .13) at Second Judgment. In Experiment 3 proportion correct was .86 (*SD* = .10) at First Judgment and .88 (*SD* = .12) at Second Judgment. No effect of judgment order was found in either experiment, both $t < 1$. Source accuracy was high and remained stable whether the source judgment preceded or followed a judgment of confidence or subjective experience. In sum, both memory measures

and source accuracy demonstrated that performance was strong and comparative across Experiments 2 and 3.

Subjective experience and confidence measures

The level of confidence or type of subjective experience participants had based their recognition decisions on was examined next. Firstly *a priori* proportions were calculated – the proportion of correct responses assigned to each of the confidence levels or subjective experience categories. As shown in the lower panel of Figure 1, there were only slight differences between assignment of First and Second Judgments to responses, but patterns of responses across confidence levels (in Experiment 3) differed from patterns of responses across subjective experience categories (in Experiment 2). In Experiment 3 over 70% of responses were assigned to High confidence, while in Experiment 2 just over 40% of responses were assigned to Remember. Approximately equal proportions of responses were assigned to Know and Familiar, whilst slightly more responses were assigned to Medium compared to Low confidence.

Separate 2(judgment order) x 3(response category) ANOVAs were performed for Experiments 2 and 3. For subjective experience responses in Experiment 2 there was no main effect of judgment order, $F(1,31) = 2.25$, $p = .14$; a significant main effect of response category, $F(2,62) = 5.93$, $p = .004$; and no interaction, $F < 1$. Aggregated across judgment order, further comparisons demonstrated that more responses were assigned to Remember than to Know or Familiar, at least $p < .015$; and there was no significant difference in the proportion of responses assigned to Know and Familiar, $t < 1$. For confidence judgments in Experiment 3 there was no main effect of judgment order, $F < 1$; a significant main effect of response category, $F(1.25,41.31) = 185.21$, $p < .001$; and no interaction between judgment order and response

category, $F < 1$. Here all proportions were significantly different, all $p < .001$; more responses were assigned to High compared to Medium or Low confidence; and more were assigned to Medium compared to Low confidence.

To explore the level of accuracy associated with each response category *a posteriori* probabilities were calculated. Here proportions were based on First Judgments only as for each item both First and Second Judgments were conditional on the same recognition response. Data shown in the lower panel of Figure 2. Separate 3(response category) ANOVAs demonstrated significant main effects of response category for subjective experience judgments in Experiment 2, $F(1.43,31.49) = 20.51$, $p < .001$; and for confidence judgments in Experiment 3, $F(1.31,34.23) = 21.40$, $p < .001$. For subjective experience judgments, there was no significant difference in the accuracy of Remember or Know responses, $t(22) = 1.31$, $p = .20$, and both were significantly more likely to be correct than were Familiar responses, both $p < .001$. In contrast, for confidence judgments, all comparisons were significantly different from each other, all at least $p < .003$. High confidence responses were significantly more likely to have been correct than responses assigned to Medium or Low confidence, and Medium confidence responses were more accurate than Low confidence responses.

As predicted, distribution patterns, accuracy, and RT findings demonstrated for confidence and subjective experience judgments in Experiment 1 were replicated in Experiments 2 and 3 demonstrating that participants do not use confidence and subjective experience judgments in the same way. Participants were again more liberal when making confidence judgments compared to subjective experience judgments and accuracy analysis demonstrated that for subjective experience judgments Know judgments were as accurate as

Remember judgments, whereas only High confidence judgments were more likely to be correct than responses assigned to lower levels of confidence.

Comparisons across experiments

Comparing the upper and lower panels of Figure 1 it is evident that the distribution of subjective experience responses in Experiment 2 was very similar to that obtained in Experiment 1; however, for confidence judgments more responses were assigned to High confidence in Experiment 3 than in Experiment 1. To examine the observation that the source encoding manipulation introduced in these experiments appears to have led to differences in memory performance and confidence, but not reports of subjective experience, further analysis was performed. Recognition performance in Experiments 2 and 3 was over 65%, which was significantly higher than the 54% recognition accuracy obtained in Experiment 1: Experiment 2, $t(74) = 2.49$, $p = .015$, Experiment 3, $t(76) = 3.94$, $p < .001$, and across Experiments 2 and 3 there were no differences between memory performance, $t(64) = 1.34$, $p = .19$, or source accuracy, $t < 1$. Additionally, there were no differences in the proportion of items assigned to Remember, Know, Familiar, and Guess in Experiments 1 and 2, all $t < 1$. However, in Experiment 1 nearly 60% of correct items were assigned to High confidence whereas in Experiment 3 this figure was higher, at over 70% (compare upper and lower panels of Figure 1); significantly more responses were assigned to High confidence in Experiment 3 than in Experiment 1, $t(76) = 2.71$, $p = .008$ and the corresponding reduction was in the proportion of responses assigned to Low confidence where fewer were assigned in Experiment 3 than in Experiment 1, $t(76) = 2.94$, $p = .004$. This exploratory analysis shows that, while the source encoding instructions increased both memory performance and patterns of confidence judgments in the current experiments, they did not alter reports of subjective experience.

The influence of one judgment on the other – Experiment 2

To fully explore the possible influence of judgments on each other, conditional probabilities were calculated on correct recognition decisions in both temporal orders. For Experiment 2, firstly the proportion of items assigned to Remember, Know, Familiar, and Guess at Second Judgment was calculated split by whether a First judgment of source had been accurate or inaccurate; i.e., *'of the items where source was judged correctly at First judgment, what proportion were then assigned to each of the categories of subjective experience at Second Judgment?'.* These proportions were then calculated in the opposite direction. In the counterbalanced design participants either judged source followed by subjective experience, or the order of these two judgments was reversed. The proportion of items assigned to Remember, Know, Familiar, and Guess at First Judgment were separated by whether the source judgment at Second judgment had been accurate; i.e., *'of the items where source was accurate at Second Judgment, what proportion of these had come from a Remember First Judgment?'.* Data shown in the upper panel of Figure 6.

[Insert Figure 6 about here]

As is evident from Figure 6, opposing patterns were observed when source judgments were accurate compared to inaccurate. For items where source was correct, the majority (45%) of these items were associated with Remember judgments. Approximately 30% of items were associated with Know, 20% with Familiar, and less than 5% with Guess. In contrast, for incorrect source judgments (which made up less than 20% of source judgments) approximately 45% were assigned to Familiar, approximately 27% to Guess, 20% to Know, and 6 to 11% were assigned to Remember. Separate 2(judgment order) x 3(response category) ANOVAs were performed on correct and incorrect source judgment items. For correct source items, ANOVA

demonstrated no main effect of judgment order, $F(1,30) = 2.33$, $p = .14$; a significant main effect of response category, $F(1.45,43.49) = 10.53$, $p = .001$; but no interaction between judgment order and response category, $F < 1$. Further comparisons aggregated across judgment order demonstrated that a greater proportion of items were assigned to Remember than to Know or Familiar, both at least $p < .02$. More correct source items were also assigned to Know than to Familiar and this difference approached significance; $t(31) = 1.96$, $p = .059$.

For incorrect source items, ANOVA again demonstrated no main effect of judgment order, $F < 1$; a significant main effect of response category, $F(2,36) = 14.08$, $p < .001$; and no interaction between judgment order and response category, $F < 1$. A greater proportion of items were assigned to Familiar than to Remember or Know, both at least $p = .001$. No other comparisons were significant. Patterns of subjective experience varied systematically depending on whether the source judgment for an item was correct or incorrect. Participants' accurate judgments of source were not just random guesses as the subjective experience accompanying them was appropriate. The greatest proportion of items were assigned to Remember when source was correct whereas the greatest proportion of items were assigned to Familiar when source was incorrect.

These proportions were then calculated in the opposite direction. The proportion of items given an accurate source judgment was calculated split by whether the item was assigned to Remember, Know, Familiar, or Guess. As shown in the top panel of Figure 7, accuracy of source was highest for items assigned to Remember, slightly lower for items assigned to Know, then Familiar, then Guess. In the 2(judgment order) x 3(response category) ANOVA the main effect of judgment order approached significance, $F(1,19) = 4.12$, $p = .057$, suggesting that perhaps source judgments were more accurate when made as First Judgment (upper left panel,

Figure 7). However, as discussed previously, overall source accuracy – which took into account all recognition decision data including Guesses – demonstrated no significant main effect of judgment order. There was a significant main effect of response category, $F(2,38) = 15.12$, $p < .001$; and no interaction between judgment order and response category, $F(2,38) = 2.20$, $p = .13$.

[Insert Figure 7 about here]

Comparisons aggregated across judgment order demonstrated that Remembered items were associated with higher source accuracy than Know or Familiar items, and Know items in turn were associated with higher source accuracy than Familiar items, all at least $p = .021$. One-sample t-tests demonstrated that the likelihood of items assigned to Remember, Know, or Familiar being given an accurate source judgment was significantly higher than chance (.50), all $p < .001$. The probability of Guessed items' source judgments being accurate did not differ from chance, $t < 1$. In this experiment where source was a relatively easy 2AFC judgment and deep processing had been undertaken at study, Remember, Know, and Familiar items were all associated with above-chance source accuracy.

The influence of one judgment on the other – Experiment 3

The same analysis comparing the influence that judgments of source and confidence had on one another was also performed for Experiment 3. Firstly the proportion of items assigned to High, Medium, Low, and None levels of confidence at Second Judgment was calculated split by whether a First judgment of source had been accurate or inaccurate; i.e., *'of the items where source was judged correctly at First judgment, what proportion were then assigned to each of the levels of confidence at Second Judgment?'.* These proportions were then calculated in the opposite direction by calculating the proportion of items assigned to High, Medium, Low, and None levels of confidence at First Judgment split by whether the source judgment at Second

judgment had been accurate; i.e., *'of the items where source was accurate at Second Judgment, what proportion of these had come from a High confidence First Judgment?'.* Data are shown in the lower panel of Figure 6.

As evident from Figure 6, very different patterns of confidence were observed when source judgments were accurate compared to inaccurate. For correct source items, a High level of confidence was generally associated with an accurate judgment of source. Over 75% of items given a correct source judgment were also assigned to High confidence. In contrast, for incorrect source judgments (which made up less than 20% of source judgments) assignment of items to levels of confidence was more varied.

Separate 2(judgment order) x 3(response category) ANOVAs were performed on correct and incorrect source judgments. For correct source judgments, ANOVA demonstrated no main effect of order, $F < 1$; a significant main effect of response category, $F(1.23,41.11) = 273.94$, $p < .001$; but no interaction between judgment order and response category, $F < 1$. Aggregated across judgment order, further comparisons between confidence levels were significant, all $p < .001$. When source judgment was correct, a greater proportion of items were assigned to High than to Medium or Low confidence, and more were assigned to Medium than to Low. For incorrect source judgments there was no significant main effect of judgment order, $F < 1$; response category, $F < 1$; and no interaction between the two, $F(2,44) = 1.99$, $p = .15$. There was no difference between the proportion of incorrect source items assigned to High, Medium, or Low confidence, all $t < 1$. In sum, whilst accurate source items were primarily retrieved with the highest level of confidence, for the less than 20% of items that received incorrect source judgments the confidence judgments that accompanied these items did not vary systematically.

These proportions were then calculated in the opposite direction. The proportion of items given an accurate source judgment was calculated split by whether the item was assigned to a confidence level of High, Medium, Low, or None. As shown in the lower panel of Figure 7, patterns of source accuracy split by confidence level differed depending on judgment order, with a linear reduction in source accuracy as confidence level decreased when confidence judgments were made following source judgments (left panel), but a more mixed pattern shown when order of judgments was reversed (right panel). Accuracy of source was always highest for items assigned to High confidence and lowest for items assigned to None but source accuracy associated with Medium and Low levels of confidence was variable.

A lot of missing data was encountered in this analysis as participants had not assigned responses to all confidence levels. As discussed earlier, over 70% of correctly recognised items were assigned to High confidence. A 2(judgment order) x 3(response category) ANOVA on this data led to an N of only 13 due to listwise exclusion. Due to the large reduction in N the conditional probabilities in this data were not analysed. Instead, the mean confidence value assigned to correctly recognised items was calculated split by the accuracy of the source judgment assigned to the item. This analysis used the original numerical confidence levels of 3, 2, 1, and 0 that participants used to make their responses instead of the verbal labels High, Medium, Low, and None. Mean confidence for items given a correct source judgment was 2.63 (*SD* = .28) at First Judgment and 2.58 (*SD* = .28) at Second Judgment. Contrastingly, mean confidence for incorrect source items was 1.76 (*SD* = .82) at First Judgment and 1.88 (*SD* = .83) at Second Judgment. A 2(judgment order) x 2(source accuracy) ANOVA demonstrated no main effect of judgment order, $F < 1$; a significant main effect of source accuracy, $F(1,22) = 32.98$, $p < .001$, and no interaction between the two, $F < 1$. Confidence was significantly higher for

correct source items compared to incorrect source items, irrespective of the order in which judgments of confidence and source had been made.

Summary of influence data

No influence of judgment order was demonstrated in Experiments 2 and 3. Patterns of responses for conditional probabilities were identical whether source had been judged before or after subjective experience or confidence. This does not fit with the findings of Martin (2007) who reported that source judgments were more accurate when performed first and that higher levels of Remember were demonstrated when subjective experience was judged second. One possible explanation is that the procedure used by Martin (2007) conflated the first judgment participants made with the recognition judgment. At test half Martin's participants made a Remember-Know-New judgment followed by a Left-Right source judgment, and the other half made a Left-Right-New judgment followed by Remember-Know. However, one-step and two-step procedures have been demonstrated to produce different patterns of responding for both subjective experience (Bruno & Rutherford, 2010; Eldridge et al., 2002; Hicks & Marsh, 1999) and source judgments (Dodson & Johnson, 1993; Marsh & Hicks, 1998). It is suggested that using a one-step procedure for his first judgment led to the results reported by Martin (2007) and that when recognition decisions and judgments are recorded separately, as in the current study, judgments of source, confidence, and subjective experience, do not influence each other.

The conditional probabilities for subjective experience and confidence judgments separated by source accuracy did reveal very different relationships between confidence and source, and subjective experience and source. In Experiment 2, subjective experience varied systematically for both correct and incorrect source items. The greatest proportion of items was assigned to Remember when source was correct whereas the greatest proportion of items were

assigned to Familiar when source was incorrect. In contrast, for confidence judgments (Experiment 3), whilst accurate source items were primarily retrieved with the highest level of confidence, for items that received incorrect source judgments the confidence accompanying retrieval did not vary systematically. This difference in the relationship between confidence and source and subjective experience and source is returned to in the General Discussion.

When conditional probabilities were calculated in the opposing direction, in addition to Remembered items, both Know and Familiar items were associated with above-chance source accuracy. This fits with previous research that has demonstrated that accurate source can accompany retrieval of items assigned to Know (Conway & Dewhurst, 1995b; Hicks et al., 2002; Meiser & Bröder, 2002; Meiser & Sattler, 2007; Starns & Hicks, 2005). While it may appear that the finding of above-chance source accuracy for Familiar items does not fit with this literature, it should be noted that previous studies have conflated Know and Familiar into one judgment. The separation of Know and Familiar is returned to in the General Discussion. For confidence judgments, source accuracy was not able to be calculated for the different levels of confidence as participants had assigned over 70% of items to High confidence. However, as predicted, mean confidence ratings for accurate source items were significantly higher than for inaccurate source items (Rotello et al., 2005; Slotnick, 2010; Yonelinas, 1999).

Reaction times for recognition decisions

Correct recognition RTs were first examined classified by accuracy of source judgment (at First Judgment). Separate 2(source accuracy) ANOVAs for recognition RTs for Experiments 2 and 3 produced different results. In Experiment 2, the difference in RTs only approached significance, $F(1,26) = 2.55$, $p = .12$; correct source recognition decisions were marginally quicker (M = 1465ms, SD = 406ms) than incorrect source recognition decisions (M =

1644ms, *SD* = 698). However, in Experiment 3 the difference in RTs was significant, $F(1,29) = 5.73$, $p = .023$; correct source recognition decisions were significantly quicker (*M* = 1640ms, *SD* = 584ms) than incorrect source recognition decisions (*M* = 1852ms, *SD* = 765ms). As there seemed to be no influence of subjective experience on source in Experiment 2, one might assume that the lack of an effect in Experiment 2 was due to a lack of power. Since Experiments 2 and 3 were identical in terms of memory performance, source accuracy, stimuli, and all elements of methodology (except other critical post-recognition judgment), it is possible to increase power for this analysis by combining the tasks. A 2(source accuracy) ANOVA demonstrated that, aggregated across experiment, correct recognition decisions followed by correct source judgments (*M* = 1557ms, *SD* = 511ms) were significantly faster than those followed by incorrect source judgments (*M* = 1753ms, *SD* = 735ms), $F(1,56) = 7.87$, $p = .007$. Thus it could tentatively be concluded that RTs are faster for recognition decisions where source judgments are ultimately correct. This conclusion fits with some previous findings (Kahn et al., 2004; Lundstrom et al., 2005) but not others (Wilding & Rugg, 1996).

Correct recognition RTs were next examined classified by the confidence or subjective experience response category the item was assigned to at First Judgment; data shown in the upper panel of Figure 8. For Experiment 2, a 3(response category) ANOVA for recognition RTs categorised by later subjective experience demonstrated a significant main effect of response category, $F(1.37,30.10) = 4.90$, $p = .025$. Correct recognition decisions for items later categorised as Remember or Know were made significantly faster than decisions for items later categorised as Familiar, both at least $p = .031$. There was no difference in speed of correct recognition if later categorised as Remember or Know, $t < 1$. For Experiment 3, a 3(response category) ANOVA demonstrated a significant main effect of response category, $F(1.49,34.28) =$

10.95, $p = .001$. Correct recognition decisions for items later categorised as High or Medium confidence were significantly faster than decisions for items later categorised as Low confidence, both at least, $p = .004$. There was no difference in the speed of correct recognition if later categorised as High or Medium confidence, $t(23) = 1.04$, $p = .31$. These findings for speed of recognition decision categorised by later subjective experience or confidence replicated those obtained in Experiment 1.

[Insert Figure 8 about here]

Reaction times for subjective experience, confidence, and source judgments

These two experiments were again interested in how quickly participants had made their post-recognition judgments; data shown in Table 3. Firstly, source judgment RTs split by accuracy and judgment order were compared across experiments. Experiment was included as a factor in order to examine whether RTs for source were consistent across experiments or whether source was speeded when participants were also required to report subjective experience or confidence. A 2(experiment) x 2(source accuracy) x 2(judgment order) ANOVA demonstrated no main effect of experiment, $F < 1$; a significant main effect of source accuracy, $F(1,46) = 26.49$, $p < .001$; but no main effect of judgment order, $F < 1$; and no significant interactions, all at least $p > .30$. Correct source judgments were faster than incorrect source judgments in both experiments, both $p < .002$. In sum, there were no differences in RT to make source judgments whether they were made as a First or Second judgment; and correct source judgments were made more quickly than incorrect source judgments. Taken together with the recognition RT data these results show a clear picture: when accurate source information does not come to mind when an item is presented RTs are slower for both the initial recognition response and subsequent source judgment.

The next analysis of interest was comparisons across judgment type. Table 3 includes mean time to make confidence and subjective experience judgments in Experiments 2 and 3. These RTs were compared against RTs for correct source judgments[6]. For Experiment 2, a 2(judgment type) x 2(judgment order) ANOVA comparing subjective experience and source RTs demonstrated a significant main effect of judgment type, $F(1,31) = 17.17$, $p < .001$, no main effect of judgment order, $F < 1$, and no interaction between the two, $F < 1$. For Experiment 3, a 2(judgment type) x 2(judgment order) ANOVA comparing confidence and source judgment RTs demonstrated a significant main effect of judgment type, $F(1,33) = 18.26$, $p < .001$, no main effect of judgment order, $F(1,33) = 1.75$, $p = .20$, and no interaction between the two, $F < 1$. Source judgments were made significantly faster than both confidence judgments and subjective experience judgments and order of judgment did not influence time to make judgment. Mean RTs for confidence and subjective experience judgments in these experiments parallel those in Experiment 1.

The finding that source judgments were made more quickly than both confidence and subjective experience judgments was not predicted. It was thought that RTs for source judgments would be similar to those for subjective experience judgments as both require assessment of what contextual information is able to be retrieved from memory. A possible explanation for this finding is that in these experiments source judgments were able to be made quickly due to the ease of the judgment. Source judgments were only 2AFC (compared to some experimental designs where multiple types of source information are tested at once, e.g., Meiser & Bröder, 2002), and the encoding manipulation whereby participants were instructed to associate words on the left with one person and words on the right with another person led to improved recognition memory and source judgments being correct over 80% of the time. If the

task had been more difficult, source judgments may require more processing time and RTs may be slower.

Further analysis examined whether differences in RT were observed for different response categories within the confidence and subjective experience judgment types. As the effect of order in the 2x2 ANOVA had not reached significance, data were aggregated across judgment order. Mean judgment RTs split by response category are shown in the lower panel of Figure 8. Separate 3(response category) ANOVAs for confidence and subjective experience revealed a significant main effect of response for confidence judgments, $F(1.43,38.48) = 5.38$, $p = .016$; but no main effect of response for subjective experience, $F(1.50,37.50) = 1.84$, $p = .18$. High confidence judgments were made significantly faster than both Medium and Low confidence judgments, both at least $p = .009$. The difference between Medium and Low confidence judgments was not significant, $t(27) = 1.33$, $p = .20$. In Experiment 2, there was no difference between speed of Remember and Know judgments, $t(25) = 1.61$, $p = .12$, or Know and Familiar judgments, $t < 1$. The difference between the speed of Remember and Familiar judgments approached significance, $t(25) = 1.96$, $p = .06$. These patterns of RTs are equivalent to those obtained in Experiment 1.

The influence of one judgment on the other's reaction time

The finding that source judgments were made significantly faster than judgments of subjective experience and confidence was not predicted. To explore the relationship between source and the other two judgment types further, RTs to make correct source judgments were analysed split by what category of subjective experience or level of confidence the item was also assigned to. For example, *'if an item was given an accurate source judgment and assigned to High confidence, how quickly had that source judgment been made?'.* Source judgment RTs

were aggregated across judgment order and separate 3(response category) ANOVAs were performed for Experiment 2 and Experiment 3. As shown in Figure 9, for both experiments, RT to make a source judgment differed across subjective experience categories or confidence levels; the fastest source judgments being those where the item was recognised with High confidence or Remembering. In both experiments, ANOVA demonstrated a significant effect of response category: Experiment 2, $F(1.46,34.98) = 6.05$, $p = .01$, and Experiment 3, $F(1.43,31.45) = 6.24$, $p = .01$. For Experiment 2, source judgments to Remember items were made more quickly than those for items assigned to Know, $t(24) = 1.94$, $p = .065$, though this only approached significance[7]. Source judgments to both Remember and Know items were also made more quickly than source judgments to Familiar items, Remember: $t(24) = 3.18$, $p = .004$, and Know: $t(24) = 1.84$, $p = .078$, though this only approached significance. For Experiment 3, source judgments to High confidence items were made more quickly than those for Medium, $t(22) = 3.06$, $p = .006$, and Low, $t(22) = 3.73$, $p = .001$, confidence items. There was no significant difference between the time taken to make source judgments for items assigned to Medium or Low confidence, $t(22) = 1.00$, $p = .33$. These results demonstrate that the subjective experience or confidence associated with recognition of an item was related to how quickly a source judgment was made to that item.

[Insert Figure 9 about here]

As summarised earlier, analysis of conditional probabilities demonstrated that post-recognition judgments are not influenced whatever source, subjective experience, or confidence judgment accompanies them. However, for RTs to make source judgments an effect of influence was revealed. In both Experiment 2 and 3, RTs to make source judgments differed across subjective experience categories or confidence levels. Source judgments were made

more quickly for items recognised with High confidence or Remembering. This demonstrates that the subjective experience or confidence associated with recognition of an item influenced how quickly a source judgment was made to that item. This finding is of particular interest for the relationship between source and subjective experience in Experiment 2.

In both Experiment 1 and Experiment 2 no significant differences were observed in RTs for Remember, Know, or Familiar judgments, the subjective experience associated with recognition of the item did not influence how long it look for participants to make their judgment of subjective experience (lower panels Figures 5 and 9; although in Experiment 2 the difference in speed of Remember and Familiar judgments approached significance). It was suggested that this is due to the threshold nature of the recollection process and that memory is assessed until something is recollected or retrieval is terminated and RTs are therefore the same for these two outcomes. However, Experiment 2 demonstrates that, while the type of subjective experience associated with recognition of an item does not influence the time it takes to make a judgment of subjective experience, the type of subjective experience associated with recognition of an item does influence the time taken to make a source judgment about the item. This is particularly interesting as source judgments in the current experiments were extremely fast. This finding shows that within that mean time to make an accurate source judgment there was significant variation which was influenced by the subjective experience or confidence with which the item was recognised.

General Discussion

Comparing source, confidence, and subjective experience judgments when the order of judgments was counterbalanced post-recognition, three experiments demonstrated no differences in patterns of judgment, relationships with source, or reaction times across judgment

order. Therefore it appears that the fear of contamination of one judgment by the other was unfounded (Bruno & Rutherford, 2010; Holmes et al., 1998; Martin, 2007; Rajaram et al., 2002). Alternatively, it could be argued that all judgments were confounded by the judgment that accompanied them. Comparisons across experiments suggest that this is not the case.

Experiment 1 compared subjective experience judgments with confidence judgments in a within-subjects design with judgment order counterbalanced across items. Conversely, Experiments 2 and 3 only employed one of these judgment types and compared it against judgments of source. The patterns for subjective experience and confidence judgments obtained in these latter two experiments replicate those obtained when subjective experience and confidence judgments were made together (Experiment 1). More specifically, if subjective experience was influenced by being made alongside confidence judgments, then this influence must be identical to that exerted when subjective experience judgments are accompanied by source judgments, as the distribution of subjective experience responses and accuracy of these responses were the same in Experiments 1 and 2. This comparison of patterns alone strongly supports the assertion that making two judgments together does not lead to contamination of one judgment by the other.

The current experiments were the first to compare four categories of subjective experience: Remember, Know, Familiar, and Guess, against four levels of confidence, and combine both these measures with judgments of source. Although the order of post-recognition judgments demonstrated no influence of one judgment type on another, distribution of responses across subjective experience categories and confidence levels, their associated source accuracy, and their reaction times, revealed hitherto unknown relationships between these different measures.

Comparing subjective experience and confidence

Both distributions and accuracy of responses demonstrated that participants did not use confidence and subjective experience judgments in the same way. Participants were more lenient in their use of confidence judgments than subjective experience judgments: more responses were assigned to High confidence than to Remember. This fits with previous findings (Gardiner & Java, 1990; Rajaram, 1993; Rajaram et al., 2002; Tunney & Fernie, 2007), but the current experiments were the first to demonstrate this by contrasting the four categories of subjective experience against four levels of confidence. Accuracy analysis demonstrated that Medium confidence responses were less accurate than High confidence responses, whereas Know responses were as accurate as Remember responses. This latter finding replicates Dewhurst et al. (2009) and Sauerland and Sporer (2009) but the differentiation across confidence and subjective experience categories had not been shown before. In Experiment 1, mean confidence ratings also demonstrated no differences between Known and Remembered items.

Distribution of judgments conditional on previous response and which type of judgment had been made first or second further demonstrated that participants used subjective experience categories and confidence levels in different ways, while also supporting the conclusion that the order in which judgments were made did not influence response patterns. For example, there were no differences in the proportion of High confidence items assigned to Remember whether judgment order was confidence then subjective experience or subjective experience then confidence. However, there was a difference when comparing the proportion of Remember responses assigned to High confidence with the proportion of High confidence responses that were assigned to Remember (cf. Figures 3 and 4). The four categories of the two different types

of judgment did not simply map onto each other. Subjective experience and confidence are therefore not "experimentally interchangeable" (Rajaram et al., 2002, p. 234), either holistically as judgment types, or when comparing individual categories such as Remember and High confidence. These patterns are considered in terms of different models of memory later in this discussion.

Further dissociation of confidence and subjective experience is shown by the fact that the source encoding manipulation employed in Experiments 2 and 3 led to increased recognition performance and higher confidence in these experiments compared to Experiment 1, but patterns of subjective experience did not change. This finding runs contrary to previous manipulations which have shown that more elaborative or conceptual processing at encoding leads to more Remember responses. Such manipulations include deep versus shallow processing, generation or vocalization versus reading of items (Gardiner, 1988; Gregg & Gardiner, 1991), orthographically distinctive versus orthographically common words (Rajaram, 1998), repetition of items at study (Dewhurst & Anderson, 1999), and intentional versus incidental learning (Macken & Hampson, 1993). It is unclear why the source encoding manipulation employed in Experiments 2 and 3 led to increases in recognition performance and confidence but no increases in Remember responses. In interpreting this result, it is important to remember that this difference was observed across experiments, not within. But the fact that increases were shown in High confidence but not in Remember judgments supports the argument that the two processes are dissociable in some tasks and in response to some variables.

To fully explore how subjective experience and confidence judgments respond to source encoding instruction manipulations, source encoding would need to be performed at study and both subjective experience and confidence judgments made at test. We are currently pursuing

this line of enquiry. Though prior research has measured both subjective experience and confidence at test and included a source manipulation at encoding (e.g, Wixted & Mickes, 2010); to our knowledge, subjective experience and confidence judgments have not been operationalized in the same manner and been measured at the same time when a source manipulation has been included at encoding.

The separation of Know and Familiar

The current experiments were the first to employ the four categories of Remember, Know, Familiar, and Guess in simple word recognition. Previous research has only employed the separation of Know and Familiar with objectively more meaningful stimuli such as learning of rare word definitions, faces, and student course material (e.g., Conway et al., 1997; Dewhurst et al., 2009; Herbert & Burt, 2001, 2003, 2004; Sauerland & Sporer, 2009; Wright & Sladden, 2003). A key question for the current experiments was therefore: How would participants use the Know response in an episodic memory task with less meaningful material?

In terms of similarities, the current experiments suggest that Know and Familiar experiences are similar in how often they are reported by participants. Experiments 1 and 2 demonstrated very similar *a priori* usage of Know and Familiar responses by participants; approximately 25% of correct recognition responses were assigned each to Know and Familiar. However, this is where the similarity ends. Recognition decisions assigned to Know were demonstrated to be more accurate, more confident, faster, and have higher source accuracy than those assigned to Familiar. In fact, Know recognition did not differ from Remember recognition on all of the above factors except source accuracy. That there were no differences in accuracy or confidence between Remember and Know recognition responses when Know and Familiar are separated replicates the results of Dewhurst et al. (2009; and regarding accuracy Sauerland

& Sporer, 2009) using very different experimental stimuli. The current novel findings for source accuracy and RT add to these previous findings concerning differences between Remember, Know, and Familiar subjective experiences.

These results firmly establish that participants are able to use separate Know and Familiar responses in episodic memory tasks. A subjective experience of 'just knowing' that an item was presented previously without any associated recollective experience or feelings of familiarity was how a significant proportion of items were recognised in the current experiments. As suggested by Gardiner (2001) concerning the traditional Remember-Know paradigm, "Other classifications of subjective conscious experiences of memory are undoubtedly possible and may even prove more useful, in the longer run, than this one" (p. 1360; see also McCabe et al., 2012). The separation of Know and Familiar is suggested to be one of these more useful classifications and one that may benefit future research in, for example, aging, neuropsychology, education, or identifying the neurological bases of memory retrieval. Indeed, as discussed previously, some researchers interested in the neural correlates of recollection and familiarity have already used modified versions of the Remember-Know categories. Montaldi et al. (2006) trained participants to distinguish between recollection and three levels of familiarity and showed that certain neural activation increased or decreased linearly as familiarity increased but hippocampal activity was not modulated by changes in familiarity: the hippocampus was only activated for items that were recollected (see also Diana, Yonelinas, & Ranganath, 2010; Yonelinas et al., 2005). Use of the separated Know and Familiar categories – or at least use of more than one category of non-recollection retrieval – may enable identification of separate patterns of neural activation for different subjective experiences, and thus perhaps aiding clarification of the opposing patterns of conclusions that

currently confuse the literature (cf. for example, Diana et al., 2010; Squire, Wixted, & Clark, 2007; Wais, 2008; Wais, Wixted, Hopkins, & Squire 2006).

Source accuracy, subjective experience and confidence

Some previous studies have found source accuracy to only be at chance for Know items (Dewhurst & Hitch, 1999; Dudukovic & Knowlton, 2006; Perfect et al., 1996, except for Experiment 3) while other studies have shown source accuracy to be above chance for items assigned to Know (Conway & Dewhurst, 1995b; Hicks et al., 2002; Meiser & Bröder, 2002; Meiser & Sattler, 2007; Perfect et al., 1996 – Experiment 3; Starns & Hicks, 2005). In the current experiments source judgments were accurate over 80% of the time overall, and Remember, Know, and Familiar recognition decisions were accompanied by above-chance source accuracy[8]. That Know responses were accompanied by accurate source supports the latter group of studies listed above. For a substantial proportion of items that were Known in Experiment 2, enough contextual details from the study phase were retrieved at recognition for source to be judged correctly but not enough details were recollected, or were experienced in the right way, for the item to be assigned to Remember. For participants this might mean that retrieval of source is more semantic in nature; for example, using the encoding manipulation in the current task a participant might have the thought "I just know I associated *'silence'* with George Clooney, so it must have been shown on the left, but I can't recall what association I made". In this example, while source is accurate, subjective experience is Know rather than Remember. Importantly, source accuracy was the only factor that reliably differentiated Remember and Know responses. A similar dissociation was demonstrated by Wixted and Mickes (2010) who asked participants to make a Remember-Know judgment in conjunction with a 20-point confidence judgment and a judgment of source. They found that when only the

highest confidence Remember and Know judgments were compared they were equally accurate in terms of recognition sensitivity; however they were accompanied by different levels of source accuracy.

As regards the studies that previously found source accuracy for Know responses to only be at chance, it is suggested that methodological differences resulted in these findings. For example, Dudukovic and Knowlton (2006) only measured source in a re-test one week after initial testing. This delay led to a decrease in Remember responses and an increase in Know responses indicating an overall reduction in retrieval of contextual details. This forgetting of source information could have led to items being Known rather than Remembered and also explain why Know responses were not accompanied by accurate source judgments. Dewhurst and Hitch (1999) did not employ a delay but their participants only made a source judgment if they were confident about their recognition decision. They suggested that the low source accuracy for Know items may have been a result of the overall low response rate for source. For Perfect et al. (1996) it is the difficulty of the source judgment that is suggested to have led to at-chance source accuracy for Know items. Of their four experiments, the three that found chance levels of source accuracy for Know responses employed source manipulations that were much more complex (temporal order, undifferentiated lists, and sequential quadrant presentation) than that employed in their Experiment 3, which did find above-chance source accuracy for Know. Experiment 3 used simultaneous quadrant presentation which could have allowed participants to more easily create spatial associations between items thus aiding retrieval of location at test – even if other elements of the association were not recollected and the experience of retrieval was one of Knowing rather than Remembering.

The current experiments were the first to separate Know and Familiar and record source judgments and thus source accuracy for Familiar responses had not been explored previously. The finding that Familiar responses were also accompanied by above-chance source accuracy suggests that feelings of familiarity at recognition can include retrieval of elements of the encoding episode that allow a correct source judgment to be made. As concluded by Hicks et al. (2002) it appears that judgments of source "can be based on recollection but can also effectively use qualitative characteristics that lack clarity and sufficient amounts of details to give rise to the subjective feeling of remembering" (p. 503). Utilising the four categories of Remember, Know, Familiar, and Guess, further research could examine whether Familiar (and/or Know) responses lose their accurate source when the source judgment is made more difficult by increasing the delay between study and test (Dudukovic & Knowlton, 2006), requiring multidimensional source judgments (Meiser & Bröder, 2002), or eliminating the deep encoding instructions that were used in the present experiments. If overall source accuracy is reduced it might be expected that Familiar recognition would no longer have accurate source. For example, previous research has demonstrated that the specificity of the source judgment differentiates between whether Know responses have accurate source or not (in a traditional Remember-Know procedure). Meiser and Sattler (2007) found that specific source memory for the voice of an individual speaker was more accurate for Remember items than for Known items. However, partial source memory for the more global attribute of gender did not differentiate between Remember and Know items. Further work is required to explore the relationship between the difficulty and informativeness of source judgments, and their relationship with subjective experience.

Further relationships between source, subjective experience, and confidence

The relationship between source and confidence and subjective experience judgments was also able to be analysed in the opposite direction in the current experiments. Patterns of subjective experience and confidence responses split by the accuracy of the accompanying source judgment demonstrated further differences between how subjective experience and confidence judgments were used. For subjective experience responses, a clear relationship between source and subjective experience was demonstrated for both correct and incorrect source judgments. Correct retrieval of source was associated with recollection whilst incorrect retrieval of source was associated with familiarity. In contrast, patterns of confidence judgments were only reliable for correct source items. When source was correct, a greater proportion of items were assigned to High, compared to Medium, compared to Low confidence; but when source was incorrect, confidence judgments were evenly split across High, Medium, and Low confidence levels. While this is not incorrect of participants as confidence judgments were made for recognition of the item, not confidence in their source judgment, it shows that there was no relationship between recognition confidence and source when source was incorrect. Critically, it is suggested that judgments of subjective experience are more sensitive and more reflective of memory processes than judgments of confidence. In addition to judgments of confidence being more lenient than judgments of subjective experience, confidence did not vary systematically when items were recognised without retrieval of source.

The current findings may best be interpreted in terms of recent conceptualisations of recollection as a continuous process (e.g., Mickes, Wais, & Wixted, 2009; Onyper, Zhang, & Howard, 2010; Parks & Yonelinas, 2007; Rotello, Macmillan, & Reeder, 2004; Slotnick, 2010; Slotnick & Dodson, 2005; Wixted & Mickes, 2010). With the separation of Know and Familiar

responses, the finding that Know and Remember responses differ only in terms of source accuracy while Remember, Know, and Familiar responses were all accompanied by retrieval of varying degrees of source accuracy is taken to support the idea that there are varying degrees of recollection. Crucially, it is not suggested that retrieval of source for Know and Familiar items was accompanied by recollective experience; instead the suggestion is that retrieval of source was experienced either as semantic knowledge (Know) or with feelings of familiarity (Familiar). Though recollection as a process might be continuous, it is suggested that the *feeling* of Remembering is threshold. The phenomenological experience associated with retrieval that warrants a Remember response is what separates this Remember recognition from Know and Familiar recognition, even if that Know and Familiar recognition can include retrieval of enough contextual details to garner an accurate judgment of source. Evidence to support this viewpoint comes from examination of reaction time.

Reaction time for recognition

To our knowledge, the current experiments were the first to examine RTs using the four categories of Remember, Know, Familiar and Guess. Here, with the separation of Know and Familiar subjective experiences, recognition decisions for Know items were as fast as for Remember items, and both Remember and Know were faster than Familiar. These results are therefore interpreted as consistent with earlier studies that found Remember to be faster than Know, as in those studies Know and Familiar had not been separated (Dewhurst & Conway, 1994; Dewhurst et al., 1998, 2006; Henson et al., 1999; Stretch & Wixted, 1998; Vilberg & Rugg, 2007; Wixted & Stretch, 2004). Though many studies have demonstrated RT differences for Remember and Know recognition decisions, many dual-process models predict that responses based on familiarity should be quicker than those based on recollection as familiarity

is considered to be rapid and automatic while recollection is thought to be slower and more effortful (e.g., Jacoby, 1991; Mandler, 1980; Yonelinas & Jacoby, 1994, 1996). This inconsistency between the models' predictions and experimental findings led Yonelinas (2002) to suggest that Remember responses are faster than Know responses because instructions require participants to only respond Know if an item is 'familiar and not recollected'. If Remember and Familiar subjective experiences are considered to directly map onto the processes of recollection and familiarity then Yonelinas's (2002) suggestion explains the current patterns – Familiar recognition decisions are slower because instructions require the recollection process to be completed first. In addition, with the separation of Know and Familiar categories, Yonelinas's suggestion is also consistent with the present findings that Know and Remember decisions were equally fast. If Know is considered to be high confidence without recollection then whether the recollection process results in anything from encoding being recollected (Remember) or nothing being recollected (Know) this should take the same amount of time and recognition RTs for Remembering or Knowing should be equal, as demonstrated in the current experiments.

However, Dewhurst et al. (2006) provided evidence against the suggestion that the speed advantage for Remember responses is an artefact of experimental instructions. In their experiment, faster recognition decisions for Remember items were demonstrated even when the judgment of subjective experience was decoupled from the recognition decision. Under these conditions, RK instructions could not influence the timing of recognition decisions and thus Dewhurst et al. concluded against the demand characteristics account proposed by Yonelinas (2002). Dewhurst et al. suggested that instead of directly reflecting the time-course of the *processes* of recollection and familiarity, Remember and Know responses reflect the time taken to make recognition decisions based on *the information which results from* recollection and

familiarity processes. If a test item cues retrieval of contextual information from encoding such as thoughts, images, or associations, then recognition decisions can be made more quickly than if it requires additional processing to evaluate its familiarity relative to other items on the test (Conway & Dewhurst, 1994; Dewhurst et al., 2006; Henson et al., 1999). For the present experiments, this explains the difference in speed of Remember and Familiar decisions but not the similarity in speed of Remember and Know decisions. Moreover, in Experiment 2 both Know and Familiar responses were accompanied by above-chance source accuracy demonstrating that some contextual information from encoding was retrieved for these items, even if this retrieval was not accompanied by a subjective feeling of Remembering.

As Know and Remember were equally fast in terms of RT, and Remember, Know, and Familiar were all accompanied by above-chance source accuracy, it is suggested it is the *subjective feeling* concerning recognition that is critically important in influencing the speed with which decisions are made. For Remember and Know responses, whether an item is retrieved with recollection of contextual information from encoding or high confidence without recollection (accompanied by 'just known' contextual information or not) results in two different subjective experiences which are experienced equally rapidly and which result in fast recognition decisions. In contrast, when the subjective experience is one of Familiarity, additional processing regarding the source of this feeling of familiarity (and any accompanying contextual information retrieved) results in recognition decisions taking longer. The subjective feeling is related to the information in memory and in what manner context is retrieved, and assessment of all these elements of retrieval influences recognition decision RTs.

For confidence judgments, patterns of recognition RTs categorised by later confidence judgment were similar to those for subjective experience judgments. Recognition decisions for

High and Medium confidence items were faster than for Low and None. The general finding that the higher the confidence the faster the response replicates previous research (Henson et al., 2000; Mandler & Boeck, 1974; Ratcliff & Murdock, 1976). However, the similarity between the patterns for confidence and subjective experience RTs suggests that the categories within the two types of judgment are similar in terms of processing time while they were not the same in terms of distribution of responses or recognition accuracy.

For source judgments, Experiments 2 and 3 did not find a consistent speed advantage for accurate source items but re-analysis of data combined across both experiments demonstrated that recognition was faster for items when the subsequent source judgment was correct. The differing findings across experiments mirrors previous research. Some studies found a difference between RT for incorrect and correct source items (e.g., Kahn et al., 2004; Lundstrom et al., 2005) and other studies did not (e.g., Wilding & Rugg, 1996). From the combined analysis in the present experiments, it is tentatively concluded that RTs are faster for correct recognition decisions where source judgments are ultimately correct. The retrieval of source information at recognition speeds the recognition decision.

Regarding interpretations of the present RT findings in terms of single- and dual-process accounts, Rotello and Zeng (2008) argued against such RT differences being support for dual-process theories by demonstrating that when Remember and Know responses were equated for confidence the RT advantage for Remember responses was reduced. In contrast to their findings, the current experiments showed that when Remember and Know responses are equated for accuracy (Experiment 2) or equated for both accuracy and confidence (Experiment 1), both Remember and Know responses are made more quickly than Familiar responses, which had lower levels of accuracy and confidence. These patterns do not support the conclusion that

subjective experience judgments are based on a single underlying memory processes and instead demonstrates that splitting Know into separate Know and Familiar categories can reveal important patterns of data otherwise obscured.

Reaction time to make judgments

These experiments were the first time that reaction times to make post-recognition judgments of confidence, source, and subjective experience have been compared. For confidence and subjective experience, mean judgment RTs in Experiments 2 and 3 paralleled those obtained in Experiment 1 and between-subjects comparisons demonstrated no significant differences in RTs across experiments, all $t < 1$. The time it took participants to make judgments of subjective experience and confidence was equivalent across experiments.

Comparisons across judgment types demonstrated that source judgments were quicker than confidence, which in turn were quicker than subjective experience. This result was not predicted. As both source and subjective experience judgments require assessment of what information can be retrieved regarding context at encoding it was predicted that both source and subjective experience judgments would be slower than judgments of confidence. One suggestion for the observed pattern is that judgments of subjective experience and confidence are both subjective types of assessment whereas source is more objective. Thus, confidence and subjective experience judgments may require greater metacognitive processing because of their non-objective nature. An alternative suggestion is that source judgments may have been so quick because only a one-dimensional source judgment was used and encoding proficiency had been improved due to the instructions provided (after Yonelinas, 1999). These manipulations led to source judgments being accurate over 80% of the time and may have also influenced RTs. Future experiments could examine whether source judgments are slower when the source

judgment is more difficult, or when the delay between encoding and retrieval is longer, as this might increase forgetting of source information. Previous research has shown that coarse source discriminations are faster than finer ones (Lindsay, 2008). For example, Johnson, Kounios, and Reeder (1994) analysed the time-course of item recognition and source memory and found that item recognition became accurate more quickly than source memory. However, multi-dimensional source experiments have not yet investigated the time course of source judgments. If this was explored in future research it might be expected that, in line with the findings of Meiser and Sattler (2007), RTs might differ for source judgments which require retrieval of specific (e.g., individual voice) and global (e.g., gender of voice) information.

In the present experiments, in addition to recognition judgments being faster when a subsequent judgment of source was correct, source judgments themselves were also faster when source was correct. This finding is novel, though unsurprising. It also adds to the findings of Kahn et al. (2004) and Lundstrom et al. (2005) who used one-step methods where recognition and source were judged at the same time. The current findings show that when separate source and recognition judgments are made, both these types of judgment are faster when source is accurate. Critically, using the two-step method in the current experiments, the speed advantage for correct source judgments demonstrated a larger and more reliable difference than the speed advantage observed for recognition decisions later shown to have correct source. Moreover, further analysis revealed that RTs for source judgments were faster when accompanied by Remember or High confidence. Thus, the subjective experience or confidence associated with recognition of an item was related to how quickly a source judgment was made to that item.

This novel finding is particularly interesting in terms of the relationship between source and subjective experience as for subjective experience judgments themselves no reliable

relationship was demonstrated between category of subjective experience and speed of judgment. The subjective experience associated with recognition of an item did not influence RTs to make subjective experience judgments yet it did influence RTs for source judgments. This is surprising as source judgments were made significantly faster than subjective experience judgments. Though source judgments were made more quickly, within that quick response there was significant variation and this was associated with the subjective experience with which the item was recognised. It is suggested that this occurred because a judgment of source is a more objective one-dimensional type of judgment compared to a subjective experience judgment – at least in these experiments. Thus, if the critical source information was recollected from encoding and this was experienced subjectively as Remembering then this led to speeding of the source judgment. This interplay between source and subjective experience fits with the suggestion made earlier that recognition decision RTs themselves are based on the relationship between the *subjective feeling* of recognition and the manner in which contextual information accompanying them is retrieved.

In contrast to the finding that source judgments were speeded by accompanying recollection, Remember responses themselves were not made significantly more quickly than other subjective experience responses. One explanation for this could be that all processing of subjective experience has already taken place at recognition and judgment RT simply reflects time taken to make a response. However, as subjective experience judgments took longer to make than recognition decisions (1794ms compared to 1560ms respectively[9]), this suggests that additional processing was undertaken at the time of making the judgment. Participants could have been checking that they did not in fact retrieve any contextual information; or checking the content of the contextual information they had retrieved – was it accurate? Metacognitive

processing such as this would result in the patterns observed; as soon as processing ends a response is made, and RT for this response is not dependent on the outcome of the processing. This suggestion fits with the pattern of results obtained for confidence judgments where judgments were made more quickly as confidence increased. For these more inferential, less subjective judgments, it is suggested that the type of processing required is easier than that for subjective experience judgments. Although the subjective experience with which an item is recognised is suggested to be what gives rise to feelings of confidence (Gardiner, 2002; Tulving, 1985), once subjective experience has been triggered by presentation of an item at test, the lack of further processing required post-recognition results in quicker judgments when confidence is higher.

Conclusions

The experiments presented here used novel experimental designs and measures to explore the differences, similarities, and reaction times of confidence, source, and subjective experience judgments. Results demonstrate that making multiple judgments does not influence those judgments; however participants use the three types of judgment in different ways – supporting the argument that there are differences between confidence and recollection.

The current experiments have shown confidence and subjective experience judgments to dissociate in many different ways. Based on these dissociations, confidence is interpreted as an inferential assessment of subjective experiences whereas subjective experience judgments themselves are considered to be more reflective of underlying memory processes and experiential states. The current findings are therefore interpreted as supporting dual-process as opposed to single-process accounts of recognition. In agreement with Dewhurst et al. (2006), subjective experience responses in the current experiments are not considered to be direct

reflections of the underlying processes of recollection and familiarity, rather they are considered to be assessments of the information provided by these processes *plus* the experiential state which results. Although the present findings may be able to be interpreted using unidimensional SDT models, that Remember and Know responses were differentiated only by source accuracy while Know and Familiar responses were dissociated by many factors adds to evidence from other behavioural and brain imaging studies that recognition decisions associated with different types of subjective experience judgments differ in ways other than just confidence or memory strength (e.g., Dewhurst et al., 2006; Diana & Ranganath, 2011).

Endnotes

[1] Additionally, in Dewhurst et al. (2009) 78% of correct responses at Time 1 were Remembered so there were only 28% of items which could be assigned to any other category of subjective experience.

[2] While RTs for incorrect recognition decisions are worthy of note (cf. Wixted & Stretch, 2004), for the purposes of this paper discussion will focus on RTs for correctly recognised items.

[3] Words obtained from the MRC Psycholinguistic Database. Familiarity values refer to the printed frequency in the language and were derived from merging three sets of familiarity norms: Pavio (unpublished), Toglia and Battig (1978) and Gilhooly and Logie (1980).

[4] After data sorting it was discovered that the word 'wiggle' had been used as both a target and a lure item. All data for this item were therefore deleted for all three experiments.

[5] The categories utilised by Martin (2007) were actually 'Recollect' and 'Familiar' instead of Remember and Know but for consistency of explanation the standard labels are used here.

[6] ANOVA using source judgment RT regardless of source accuracy demonstrated the same patterns: source judgments were made faster than subjective experience judgments, $F(1,31) = 13.46$, $p = .001$, no main effect judgment order, $F<1$, and no interaction, $F<1$. Similarly, source judgments were made faster than confidence judgments, $F(1,33) = 6.32$, $p = .017$, no main effect judgment order, $F(1,33) = 2.47$, $p = .13$, and no interaction, $F<1$.

[7] Based on the literature it would be predicted that Remember judgments are related to judgments of source and therefore a one-tailed t-test could be used here (and the *p* value

obtained would be .033). A one-tailed test was not used as the analysis of source judgment RT split by response category was a post-hoc analysis.

[8] As mentioned previously, medium and Low confidence responses could not be analysed in terms of source accuracy as 70% of items had been assigned to High confidence resulting in missing data.

[9] Average RTs calculated across experiments. RTs for Guess excluded from calculations.