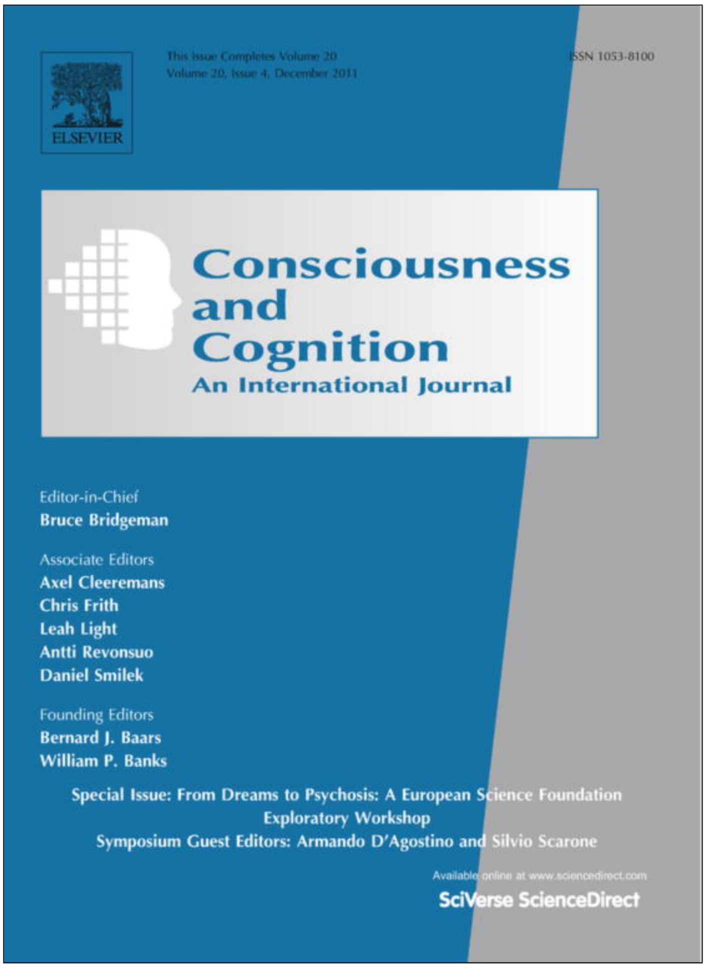
This Issue Completes Volume 20
Volume 20, Issue 4, December 2011
ISSN 1053-8100
ELSEVIER
Consciousness and Cognition
An International Journal
Editor-in-Chief
Bruce Bridgeman
Associate Editors
Axel Cleeremans
Chris Frith
Leah Light
Antti Revonsuo
Daniel Smilek
Founding Editors
Bernard J. Baars
William P. Banks
Special Issue: From Dreams to Psychosis: A European Science Foundation Exploratory Workshop
Symposium Guest Editors: Armando D'Agostino and Silvio Scarone
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Consciousness and Cognition 20 (2011) 1363–1375

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Recognition memory performance as a function of reported subjective awareness

Heather Sheridan, Eyal M. Reingold *

Department of Psychology, University of Toronto at Mississauga, Ontario, Canada L5L 1C6

ARTICLE INFO

Article history:
Received 21 December 2010
Available online 24 May 2011

Keywords:
Recognition memory
Consciousness
Unconscious memory
Remember know paradigm
Subjective measures of awareness

ABSTRACT

Three experiments introduced a recognition memory paradigm designed to investigate reported subjective awareness during retrieval. At study, in Experiments 1A and 2, words were either generated or read (generation), while modality of presentation (auditory versus visual) was manipulated in Experiment 1B. Word pairs (old/new or new/new) were presented during test trials, and participants indicated if they contained an old word by responding "remember", "know" or "new" in Experiments 1A and 1B, and by responding "strong no", "weak no", "weak yes", or "strong yes" in Experiment 2. Participants were then required to decide which of the 2 words was old. We demonstrated that the proportion measures used in the Remember Know paradigm substantially underestimated the influence of generation on familiarity resulting in an artificial dissociation between indices of knowing (familiarity) and remembering (recollection). We also found a qualitatively different pattern of forced-choice recognition performance as a function of claimed awareness.

1. Introduction

The subjective, or 'claimed awareness', method for measuring conscious awareness has been the source of ample findings and controversies (for reviews see Reingold, 2004; Reingold and Merikle, 1988, 1990; Reingold and Sheridan, 2009). This approach requires participants to report on their subjective phenomenal awareness while performing the experimental task. Although subjective reports have often been met with skepticism, the subjective approach is the oldest and arguably the most intuitive method for measuring awareness. As pointed out by Reingold and Toth (1996), there is a curious discrepancy between the use of the subjective measurement approach in the unconscious perception literature and the unconscious memory literature. Specifically, while the unconscious perception literature has a long history of employing subjective measures of awareness (for a review of early studies, see Adams (1957)), the field of memory has only more recently incorporated such measures, primarily within the framework of the Remember Know (RK) paradigm that was first introduced by Tulving (1985) and later refined by Gardiner (1988) to study subjective states of awareness at the time of retrieval. In advocating the use of the subjective measurement approach, Gardiner and Richardson-Klavehn (2000) argue that "one cannot tell what subjects experience mentally from purely objective measures of their performance. If one wants to be able to take into account subjective awareness of memory, there is no alternative to the use of subjective reports" (p. 230).

In the most commonly used "two-step" version of the RK paradigm, participants are asked to first judge if a given test item was from the study phase (i.e., old) or was not previously presented (i.e., new), and then to further classify each old

* Corresponding author. Address: University of Toronto at Mississauga, Department of Psychology, 3359 Mississauga Road N. RM 2037B, Mississauga, Ontario, Canada L5L 1C6. Fax: +1 905 569 4326.

E-mail address: reingold@psych.utoronto.ca (E.M. Reingold).

1053-8100/$ - see front matter Crown Copyright © 2011 Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2011.05.001

response as either *remember* if they experienced the recollection of episodic details for that item (i.e., specific details concerning their previous encounter with the item), or *know* if they experienced familiarity in the absence of recollection. The overall proportion of recognized old items (hit rate) in each experimental condition is then partitioned based on the subjective classifications provided by participants to form the separate remember and know proportions that constitute the primary dependent measures in this paradigm. Similarly, in another version of the paradigm (i.e., the "one-step" version), participants make a simultaneous "Remember", "Know" or "New" judgment for each item, and the primary dependent measure is the proportion of items that received each type of judgment. The vast majority of RK studies have explored the differential effects of experimental manipulations on these proportion measures. In summarizing this literature, Gardiner and Richardson-Klavehn (2000) concluded that remembering and knowing are dissociable states of awareness because it has been demonstrated that "some variables affect one or the other of the two states of awareness, that some variables have opposing effects on them, and that some variables have parallel effects on them" (p. 231).

However, the contemporary RK literature is very much divided on the question of whether remember and know judgments reflect two qualitatively different processes or systems. In this debate, single-process accounts contend that remembering and knowing reflect different response criterions along a continuum of confidence or memory strength (e.g., Donaldson, 1996; Dunn, 2004, 2008; Heathcote, Bora, & Freeman, 2010; Hicks & Marsh, 1999; Hirshman & Master, 1997; Inoue & Bellezza, 1998; Malmberg, Zeelenberg, & Shiffrin, 2004; Rotello, Macmillan, Hicks, & Hautus, 2006; Wais, Mickes, & Wixted, 2008; Wixted & Stretch, 2004), while a competing class of dual-process models interprets remembering and knowing as reflecting the operation of two qualitatively different systems or processes, such as the semantic versus episodic memory systems (Tulving, 1985), distinctiveness versus fluency processing (Rajaram, 1996) and familiarity versus recollection (Jacoby, Yonelinas, & Jennings, 1997; for a review, see Yonelinas, 2002).

Despite this marked lack of agreement over what exactly remember and know responses represent, the RK paradigm is currently being used extensively to make conclusions about the effects of a variety of variables on recollection and familiarity. In reviewing the current state of the RK literature, Wixted and Mickes (2010) noted that the RK paradigm was used in over 30 publications in the year 2009 alone. Importantly, these publications spanned a diverse range of fields and included high profile journals, and "In nearly all of these studies, remember judgments are assumed to reflect recollection, and know judgments are assumed to reflect familiarity (p. 3)."

Given the RK paradigm's prominence in the literature, it is necessary to work towards a better understanding of the processes underlying remembering and knowing, and of the measures that are used to index them. The main purpose of the present paper is to further examine the RK paradigm's know response proportion measure, in light of past criticisms of this measure that have arisen out of both the dual-process and the single-process perspectives. To accomplish this goal, we introduce a modified RK paradigm that combines the standard RK paradigm with an approach developed by Merikle and Reingold (1990) in the field of unconscious perception. Accordingly, we begin by briefly describing some of the controversies surrounding the standard RK paradigm, with a focus on past criticisms of the know response proportion measure. We then explain the rationale behind the present approach and report on the findings from experiments employing the modified RK paradigm to examine the effects of generation, (Experiment 1A and Experiment 2) and modality (Experiment 1B) manipulations on the RK response proportion measures as well as on new measures derived from our modified paradigm.

Most relevant for the present context is the controversy surrounding the attempts to use the standard RK paradigm's know response proportion measure to make inferences about familiarity. In particular, Jacoby et al. (1997) argued that the know response proportion measure tends to provide an underestimation of the contributions of familiarity to memory performance. According to the approach advocated by Jacoby et al. (1997), it is possible for some items to be both recollected and familiar. However, in the RK paradigm participants are forced to choose between either a remember or a know response, and are not allowed to provide both responses to the same item (i.e., response exclusivity). Note that in the RK paradigm, when recollection occurs, regardless of whether or not familiarity is present, participants are instructed to respond remember, and consequently if some items are both recollected and familiar then by definition the know response proportion measure will underestimate the contributions of familiarity to performance.

This underestimation critique has important implications for dissociations that are based on the standard RK paradigm's response proportion measures of remembering and knowing. Critically, conditions with high levels of remembering may constrain the number of leftover trials that can receive know responses. Thus, the RK paradigm could produce artificial dissociations such that experimental manipulations with a large effect on remembering appear to have a null effect, or even a reverse effect on knowing, even if the manipulation actually has a more parallel effect on recollection and familiarity. Jacoby et al. (1997) used this logic to explain the puzzling finding that know responses to new words (i.e., lures) sometimes exceed know responses to old words (i.e., targets). If knowing reflects familiarity such a finding might suggest a negative familiarity effect such that new words were more familiar than old words. It is hard to envision a conceptual framework that would predict such an effect. Similarly, this criticism might call into question dramatic dissociations such as the one reported by Parkin and Walter (1992) showing that there were higher levels of remembering for younger than older adults, but there were higher levels of knowing for older than for younger adults. In commenting on this dissociation as a function of aging, Jacoby et al. (1997) stated that "Although it may be comforting to think that deficits in recollection are offset by improvements in familiarity, this pattern of results appears to be an artifact of the RK procedure" (p. 35). Specifically, according to Jacoby et al. (1997) knowing is underestimated for young adults because of their higher levels of remembering compared to older adults.

Arguing against the criticism that knowing is underestimated when there are high levels of remembering, Gardiner, Java, and Richardson-Klavehn (1996) questioned the relationship between the strength of remember effects and the incidence of reversal effects on knowing. Gardiner et al. (1996) demonstrated a pattern of a large effect on remembering but no effect on knowing, and argued that such a dissociation is "at variance with the idea that as a general rule response exclusivity produces opposite effects on know responses when there are large effects on remember responses"(p. 119). However, even if it can be reliably shown that variables with large effects on remembering lead to null effects on knowing, this pattern could still be the result of underestimation if the true effect of the variable was to increase both recollection and familiarity. Thus, it is still an open issue whether or not the know response proportion measure is subject to underestimation and more importantly whether dissociations demonstrated with such a measure, as argued by Jacoby et al. (1997), might represent an artifact of the RK paradigm.

An additional critique of the know response proportion measure that was based on considerations from signal detection theory (Green & Swets, 1974; Macmillan & Creelman, 2005) was independently advanced by both Donaldson (1996) and Hirshman and Master (1997). Specifically, Donaldson (1996) pointed out that despite the fact that hit rates confound response criteria with memory sensitivity (Snodgrass & Corwin, 1988), the vast majority of the RK literature employed hit rates as the primary measure of performance under remembering and knowing. Moreover, based on a meta-analysis and an experimental demonstration, Donaldson (1996) documented that even converting the proportion of know responses to targets and lures into a know sensitivity measure (e.g., d' and A') does not save the measure. Specifically, Donaldson (1996) showed that there was a substantial correlation between criterion placement and know sensitivity measures. This finding was later replicated by Inoue and Bellezza (1998), but not by Gardiner, Ramponi, and Richardson-Klavehn (2002). Consequently, Donaldson (1996) argued that the absence of an unbiased sensitivity measure for knowing might lead to a contradictory pattern of results concerning the impact of experimental manipulations on knowing. For example, Donaldson (1996) suggested that a conservative response criterion produced an effect of a generation manipulation on both remember and know responses in Wippich (1992), while a more lenient response criterion resulted in a generation effect on remember but not on know responses in Gardiner (1988). Similar to Donaldson (1996), Hirshman and Master (1997) contended that criterion changes can lead to an inconsistent pattern of know results across studies. To support this viewpoint, Hirshman and Master (1997) reviewed the RK literature and identified seven variables, including generation, that produce remarkably consistent results for remembering, but inconsistent results for knowing, such that some studies report a reverse effect of the variable on knowing and other studies report a null effect on knowing (see Donaldson, MacKenzie, & Underhill, 1996 for a related demonstration).

Thus, an important criticism of the know response proportion measure is that it appears to be susceptible to criterion differences across participants and studies. As a possible solution to this problem, Gardiner et al. (1996) introduced a two-alternative forced-choice (2AFC) version of the RK paradigm which differs from the standard yes/no recognition version of the paradigm. In the 2AFC version of the paradigm, participants first decided which of two words was old, and then categorized this decision as being a remember or a know response. Critically, Gardiner et al. (1996) argued that the dissociations demonstrated using this 2AFC paradigm would provide strong convergent evidence for the dissociations previously shown with the yes/no version of the paradigm. In support of this argument, Gardiner et al. (1996) cited the observation by Green and Swets (1974) that 2AFC tasks eliminate the problem of variations in participants' response criterions and provide a direct, unbiased measure of sensitivity. Gardiner et al. (1996) reported replicating several key dissociations that were previously shown in the yes/no RK paradigm with their 2AFC version of the paradigm and concluded that obtaining such dissociations "in a two-alternative forced-choice test means that the dissociation directly reflects target-lure discriminability, since subjects are forced to adopt the same criterion for all responses" (p. 118).

Unfortunately, even though as asserted by Gardiner et al. (1996), the 2AFC proportion correct measure is generally considered to be an unbiased measure of sensitivity (Green & Swets, 1974; Macmillan & Creelman, 2005)[1], these authors did not compute a separate 2AFC proportion correct value for each subjective response category. Instead, they simply partitioned the overall proportion correct into separate components based on the subjective classifications. Specifically, for each subjective response category, Gardiner et al. (1996) divided the number of correct trials that received a particular subjective response (the numerator) by the total number of trials (the denominator). A proportion correct measure would have used the same numerator as the Gardiner et al. (1996) method, but the denominator would instead have been the number of trials that received each subjective response. For example, the Gardiner et al. (1996) know measure is the number of correct know trials divided by the total number of trials, whereas the proportion correct know measure is the number of correct know trials divided by the number of trials that received a know response (see Supplementary materials for details).

Importantly, separate 2AFC proportion correct values for each subjective response category could independently vary from chance to perfect performance. However, in marked contrast, the Gardiner et al. (1996) remember and know values do not vary independently, and consequently a high level of remembering might still produce an underestimation of knowing. Thus, although in principle, a 2AFC version of the RK paradigm has potential methodological advantages over the yes/no version of this paradigm, the implementation by Gardiner et al. (1996) did not capitalize on these potential advantages.

[1] In yes/no tasks, bias refers to a participant's willingness to respond "old". The 2AFC task eliminates variations across participants in this type of bias. However, as pointed out by Macmillan and Creelman (2005) in a 2AFC task there can still be variations in position bias (i.e., a participant's tendency to prefer a particular location).

In the present study, to further investigate the validity of the RK know measure and the dissociations obtained using that measure, we introduce a modified version of the RK paradigm which was designed to permit concurrent measurement of remembering and knowing in both yes/no recognition and 2AFC recognition decisions. Specifically, participants were shown pairs of words at test (50% of trials = one old word and one new word, 50% of trials = two new words). The pair of words remained on the screen for two questions. For the first question (the yes/no recognition decision), participants had to decide if either word was old by choosing one of three response options: "remember", "know", "new". They were instructed to respond "remember" if either word triggered the conscious recollection of episodic details, "know" if either word felt familiar in the absence of recollection, and "new" if they believed that neither word was from the study phase. For the second question (the 2AFC recognition decision), regardless of their response to the first question, participants were asked to decide which word was old.[2]

The main rationale for creating this modified paradigm was to directly compare the RK literature's response proportion measures of knowing (the yes/no response proportion measure and the Gardiner et al., 1996, 2AFC response proportion measure) with a proportion correct measure, in which the proportion of correct trials is calculated separately for the subset of trials that received each type of subjective response. The two response proportion measures are calculated to confirm that the paradigm can replicate previous RK findings, and the proportion correct measure is calculated to assess the validity of past RK dissociations. Importantly, in our paradigm, the know proportion correct measure is not subject to constraints in the presence of high levels of remembering. This is because regardless of the magnitude of a variable's effect on remembering, the know proportion correct measure is always free to independently vary between chance and perfect performance. Thus, if we find that the know proportion correct measure shows the same results as the standard RK know response proportion measures, then such a pattern of results would help to validate the RK literature's know response proportion measures. However, if the proportion correct measure fails to replicate the know response proportion measures, then such a pattern of results would lend support to the criticisms that were put forth by Donaldson (1996), Hirshman and Master (1997), and Jacoby et al. (1997).

A second important advantage of the present paradigm is that it allows us to examine recognition memory under conditions of claimed null awareness (i.e., trials in which participants claim that both words are new). This aspect of our paradigm was inspired by an unconscious perception study by Merikle and Reingold (1990). These authors demonstrated that participants could perform above chance at deciding which of two stimuli had been presented, even for trials in which they believed that they did not see a stimulus (i.e., claimed null awareness trials). Furthermore, a qualitative difference occurred between claimed awareness and claimed null awareness trials. Specifically, for claimed awareness trials, performance was above chance for both word and nonword stimuli, while for claimed null awareness trials, performance was above chance for words but not for nonwords. Using similar logic, the present paradigm is used to examine the possibility of above chance recognition memory in claimed null awareness trials, and to explore potential qualitative differences between 2AFC performance as a function of the presence or absence of claimed awareness.

2. Experiments 1A and 1B

Experiments 1A and 1B were designed to examine the validity of past findings concerning the effect of encoding variables on knowing, in light of past criticisms of the RK literature's response proportion measures of knowing (Donaldson, 1996; Hirshman & Master, 1997; Jacoby et al., 1997), and to explore recognition memory under conditions of claimed null awareness. To accomplish these goals, we examine the effects of two experimental manipulations (generation, modality) on three different measures that are built into our modified RK paradigm: the yes/no response proportion measure, the Gardiner et al. (1996) 2AFC response proportion measure, and a proportion correct measure that is calculated separately for remember and know trials.

To test Jacoby et al. (1997)'s assertion that the know response proportion measure tends to underestimate familiarity in conditions with high levels of remembering, the present study compares a generation manipulation (Experiment 1A) with a modality manipulation (Experiment 1B). Past RK studies employing generation manipulations have found that generating a word at study (e.g., from an anagram, a clue, or a rule), compared to simply reading the word, leads to a substantial increase in remembering but has no effect on knowing, and this pattern of results has been found with both the yes/no RK paradigm (Gardiner, 1988; but see, Wippich, 1992) and the 2AFC RK paradigm (Gardiner et al., 1996). In contrast, RK studies have shown that presenting words in the same modality rather than a different modality at study and at test increases knowing but does not impact levels of remembering (Gregg & Gardiner, 1994; but see, Rajaram, 1993).

We expect that the present study's generation and modality manipulations will replicate the above pattern of results when the RK literature's response proportion measures are used. However, if the Jacoby et al. (1997) underestimation critique is correct, then a different pattern of results might emerge for the know proportion correct measure in the case of Experiment 1A. This is because the generation manipulation is expected to have a large impact on remembering, which would potentially produce constraints on knowing for the response proportion measures, but not for the proportion correct

[2] Note that the inclusion of both yes/no and 2AFC decisions in the present paradigm enables a wide range of comparisons with past RK findings. Although participants are required to provide the yes/no response prior to the 2AFC decision, we make no assumptions about the order in which participants internally process the information relevant to their decisions. In particular, we do not assume that yes/no decisions are independent from, and/or necessarily precede 2AFC decisions. For example, it is possible that on some trials participants determined their 2AFC response prior to finalizing their yes/no decision.

measure. Thus, if underestimation occurs, Experiment 1A is expected to produce a null effect of generation for the response proportion measures of knowing, coupled with a significant generation effect for the proportion correct measure of knowing. In contrast, because the Experiment 1B modality manipulation is expected to produce little or no effect on remembering, the underestimation problem should not be a factor in this case. Accordingly, the modality manipulation is expected to yield a congruent pattern of results across the response proportion measures and the proportion correct measure. Taken together, the results from Experiment 1A and 1B would provide an empirical test of Jacoby et al. (1997)'s assertion that the know response proportion measure produces an underestimation of familiarity, and that such underestimation is particularly likely in conditions with high levels of remembering.

In addition, to investigate the argument by Donaldson (1996) and Hirshman and Master (1997) that the know response proportion measure is not independent from response bias, we attempted to replicate Donaldson (1996)'s finding of a correlation between know sensitivity measures and response bias measures. Finally, the present study was also designed to explore recognition memory under conditions of claimed null awareness. Accordingly, Experiment 1A and 1B examined whether participants were still significantly above chance on the proportion correct measure for the subset of trials in which they gave a new response. Furthermore, we explored whether the generation (Experiment 1A) and the modality (Experiment 1B) manipulations differentially influenced 2AFC proportion correct performance for claimed null awareness trials (i.e., new response trials) as compared to claimed awareness trials (i.e., remember and know response trials).

2.1. Method

2.1.1. Participants

All 64 participants (32 in each of the two experiments) were undergraduate students at the University of Toronto. They were given either course credit or $10 (Canadian) per hour. All had normal or corrected-to normal vision. A total of 5 additional participants (2 in Experiment 1A, and 3 in Experiment 1B) were replaced due to difficulty in understanding and following the RK instructions.

2.1.2. Materials

The stimuli were 784 words (50% = four letters, 50% = five letters) with a mean Francis and Kučera (1982) word frequency of 18.1 occurrences per million (range = 8–47). Participants were shown 192 single words at study. In Experiment 1A, for the generate condition (50% of study trials), words were displayed with the letters in a backwards order (e.g., "elbat" for "table"), while for the read condition (50% of study trials), the letters were displayed in the correct order (e.g., "table" for "table"). In Experiment 1B, the presentation of study words was either visual (50% of study trials) or auditory, through headphones (50% of study trials). At test, there were 384 pairs of words, with half of the pairs containing one old word (Experiment 1A: 50% = generated, 50% = read; Experiment 1B: 50% = visual, 50% = auditory) and one new word, and the remaining half containing two new words. In addition, there were three practice trials that were presented at the start of the test phase, and a total of 12 filler study words that were shown at the beginning and end of the study phase to minimize primacy and recency effects. The order of the study and test phase trials was determined randomly for each participant, and the words were counterbalanced across participants such that each word appeared equally often in each condition.

2.1.3. Procedure

The participants were not informed about the recognition test until after the completion of the study phase. Instead, prior to the study phase they were told that the purpose of the experiment was to investigate reading. In the Experiment 1A study phase, generate and read trials were mixed together in a random order. To indicate the study trial type, each study trial began with the presentation of a central arrow cue for 1200 ms. For generate trials, the arrow pointed left and was followed by a backwards spelled word. For read trials, the arrow pointed right and was followed by a correctly spelled word. Participants were instructed to read each word out loud and then to press a button to proceed to the next trial. In the Experiment 1B study phase, the auditory and visual trials were presented in separate blocks, and the order of the blocks was counterbalanced across participants. Both the auditory and visual study trials began with the presentation of a central fixation cross. The participant then pressed a button to start the trial. For the auditory study trials, participants listened to a 1500 ms sound clip through headphones, and during this time the screen remained blank. For the visual study trials, a word was presented for 1300 ms, followed by a blank screen for 200 ms. Participants were instructed to read the visual words out loud, while for auditory words they repeated out loud the word that they had just heard.

The test phase procedure was identical for Experiment 1A and Experiment 1B. On each test phase trial, participants were shown a pair of words in the center of the screen, with one word above the other. For 50% of trials, both words were new, and for the remaining 50% of trials there was one new word and one old word (half of the time the old word was the upper word, and half the time it was the lower word). Participants were informed that 50% of the trials would contain an old word. During the test phase, participants answered two questions about each word pair. For the first question, participants had to decide if either word was old by giving one of three response options: remember, know, or new. They were told to respond "remember" if they could remember specific details about their past encounter with one of the words, and to respond "know" if one of the words seemed familiar in the absence of recollection of episodic details. They were told to respond, "new", if they believed that neither one of the two words was old. On every trial, regardless of which of the three responses (remember, know or new) was given to question one, participants had to decide which of the two words was old, by pressing the up or the down arrow on the keyboard. The word pair remained on the screen for both questions, and a prompt at the top of the screen reminded the participants of the current question and of the available response options (see Supplementary materials for further details about the study and test phase trial sequences). Prior to starting the test phase, participants were given written instructions, followed by a verbal explanation by the experimenter. To ensure understanding, participants were also asked to explain the test phase procedure in their own words, and to generate their own examples and explanations of remembering and knowing.

2.2. Results and discussion

The main purpose of Experiment 1 was to compare the subjective response proportion measures that are commonly used in the RK literature (i.e., the yes/no response proportions and the 2AFC response proportions), with the present paradigm's proportion correct measure. Accordingly, for each encoding condition (Exp. 1A: generate, read; Exp. 1B: visual, auditory) and subjective response category (remember, know, new), the following measures were derived from our modified RK paradigm: (1) yes/no response proportions (i.e., the mean proportion of trials containing an old word that received each type of subjective response. For example, the yes/no know response proportion is the number of trials that received a know response divided by the total number of trials containing an old word), (2) 2AFC response proportions (i.e., the number of correct trials that received each type of subjective response, divided by the total number of trials containing an old word. For example, the 2AFC know response proportion is the number of correct know trials divided by the total number of trials containing an old word), (3) proportion correct (i.e., the number of correct trials that received a particular subjective response, divided by the total number of trials that received this subjective response. For example, the know proportion correct value is the number of correct know trials divided by the number of trials that received a know response).

In the following analyses, we first confirm that our modified RK paradigm can replicate past RK findings using the yes/no and 2AFC response proportion measures. We then report the proportion correct findings, which revealed an interesting discrepancy in the pattern of results for the know proportion correct measure as compared to the know response proportion measures. Specifically, in Experiment 1A, the response proportions for knowing showed a null generation effect, while know proportion correct performance revealed a significant generation effect. To follow-up on this discrepancy across measures, we verify that a similar discrepancy still occurs after the yes/no response proportion measure is converted into an A' sensitivity measure, to place it on the same scale as the proportion correct measure. Finally, we report analyses in support of two additional hypotheses. Specifically, we verify that our paradigm replicates the Donaldson (1996) correlation between know sensitivity and bias, and we show that 2AFC recognition memory performance can be above chance under conditions of claimed null awareness (i.e., following a "new" response).

2.2.1. Response proportions

The first step in the analysis was to confirm that despite the modifications that were made to the RK paradigm, past findings were still replicated when the RK literature's response proportion measures were derived. Fig. 1 shows the means and standard errors for the yes/no and 2AFC response proportion measures,[3] by experiment and by encoding condition. As can be seen from the Fig. 1 means, Experiment 1A replicated the dissociation reported by past RK studies (Gardiner, 1988; Gardiner et al., 1996), by showing a significant generation effect for overall recognition performance (i.e., the overall hit rate) and for remembering (all t's > 8, all p's < .001), coupled with a null generation effect for knowing (all t's < 2, all p's > .1), and this pattern was consistent across both of the response proportion measures (yes/no and 2AFC).

Similarly, as can be seen from the means in Fig. 1, Experiment 1B produced results that were consistent with prior findings. Most importantly, both the yes–no and 2AFC response proportion measures demonstrated no effect of modality on remembering (all t's < 2, all p's > .1). The results for knowing were somewhat mixed, with the yes/no response proportion measure revealing a non-significant numerical trend of a modality effect on knowing, $t(31) = 1.46$, $p = .15$, and the 2AFC response proportion measure showing a significant effect such that knowing was higher for visual trials than for auditory trials, $t(31) = 3.44$, $p < .01$. This mixed pattern concerning the effect of modality on knowing is similar to the pattern that emerged in prior investigations with Gregg and Gardiner (1994) showing an influence of modality on knowing, while Rajaram (1993) failed to do so. Finally, in Experiment 1B, the overall proportion of old responses showed a significant modality effect for both the yes/no and 2AFC response proportion measures (all t's > 3, all p's < .01). In sum, these analyses confirm that even though we used a modified version of the RK paradigm, our paradigm yielded the same results as past RK studies when we used the RK literature's response proportion measures of remembering and knowing.

2.2.2. Proportion correct

The next step in the analysis was to compare the response proportion measures with a bias-free proportion correct measure that has not yet been used in the RK literature. To obtain this measure, we calculated the proportion of correct trials separately for the subset of trials that received each type of subjective response (remember, know, new). Fig. 2 (right side)

[3] In addition, some RK studies report a measure of the proportion of responses to lures. We obtained a yes/no proportion of responses to lures measure by calculating the proportions of remember and know responses given to "new new" trials (Experiment 1A: $R = .07$, $K = .28$; Experiment 1B: $R = .10$, $K = .34$) and we obtained a 2AFC responses to lures measure by calculating the proportions of remember and know responses for incorrect "old new" trials (Experiment 1A: $R = .03$, $K = .09$, Experiment 1B: $R = .04$, $K = .12$).

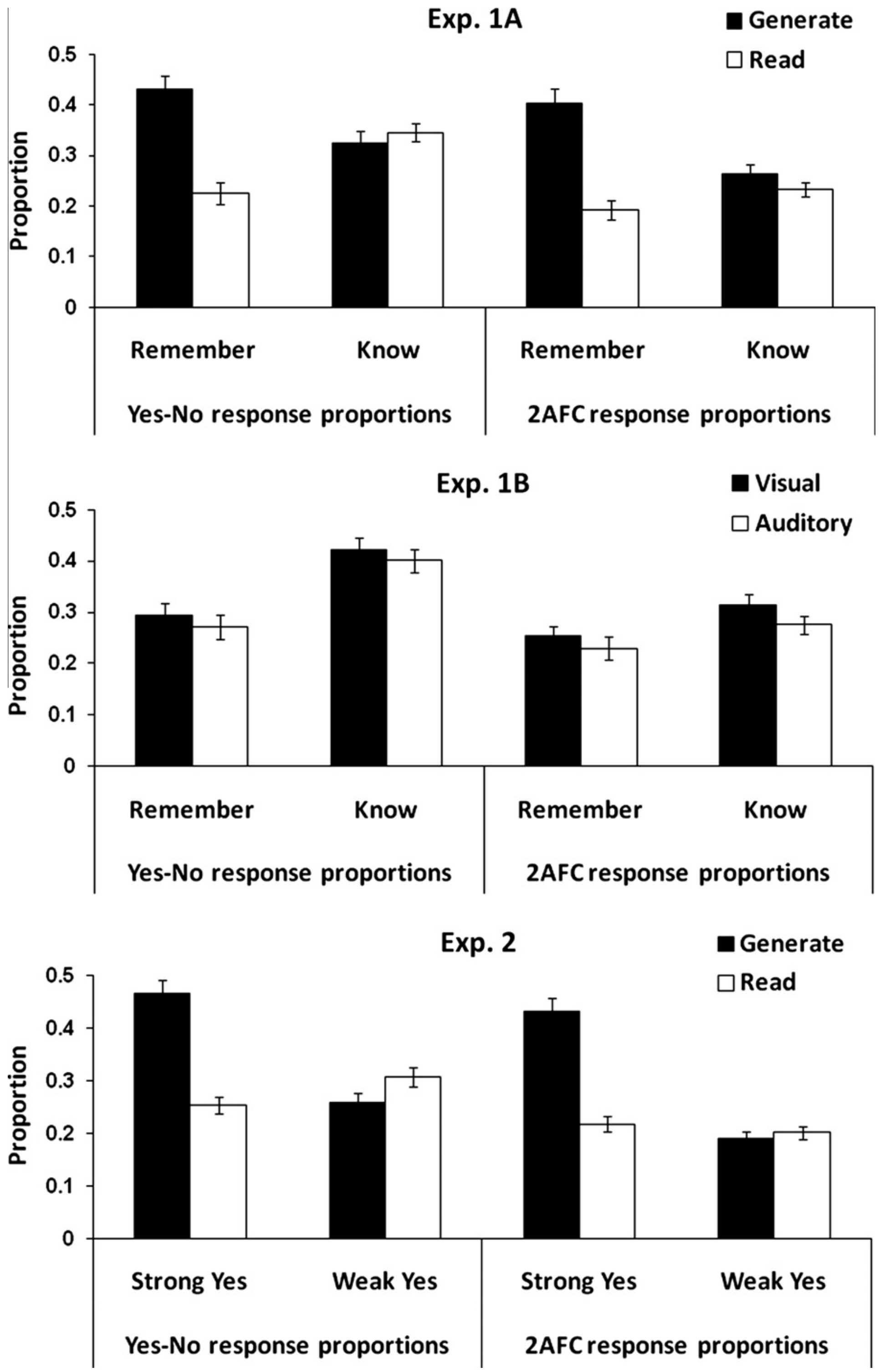

Fig. 1. Experiment 1A, Experiment 1B, and Experiment 2 mean proportions of subjective responses to targets as a function of encoding condition and calculation method: for the Yes–No method a target is defined as a trial containing an old word, and for the 2AFC method a target is defined as a correct 2AFC response.

displays the means and standard errors for the proportion correct measure. For Experiment 1A, the proportion correct t-tests revealed a different pattern as compared to the response proportion findings. In contrast to the dissociation shown by both of the response proportion measures, the proportion correct measure revealed a parallel effect of generation on remembering and knowing, such that performance was better for the generate than the read condition, for both remember trials, $t(31) = 3.43$, $p < .01$, and know trials, $t(31) = 6.70$, $p < .001$, as well as for overall performance, $t(31) = 8.70$, $p < .001$.

In contrast to Experiment 1A, the proportion correct results from Experiment 1B showed a similar pattern of results to the two response proportion measures. Similar to both the yes/no and 2AFC response proportion measures, across both

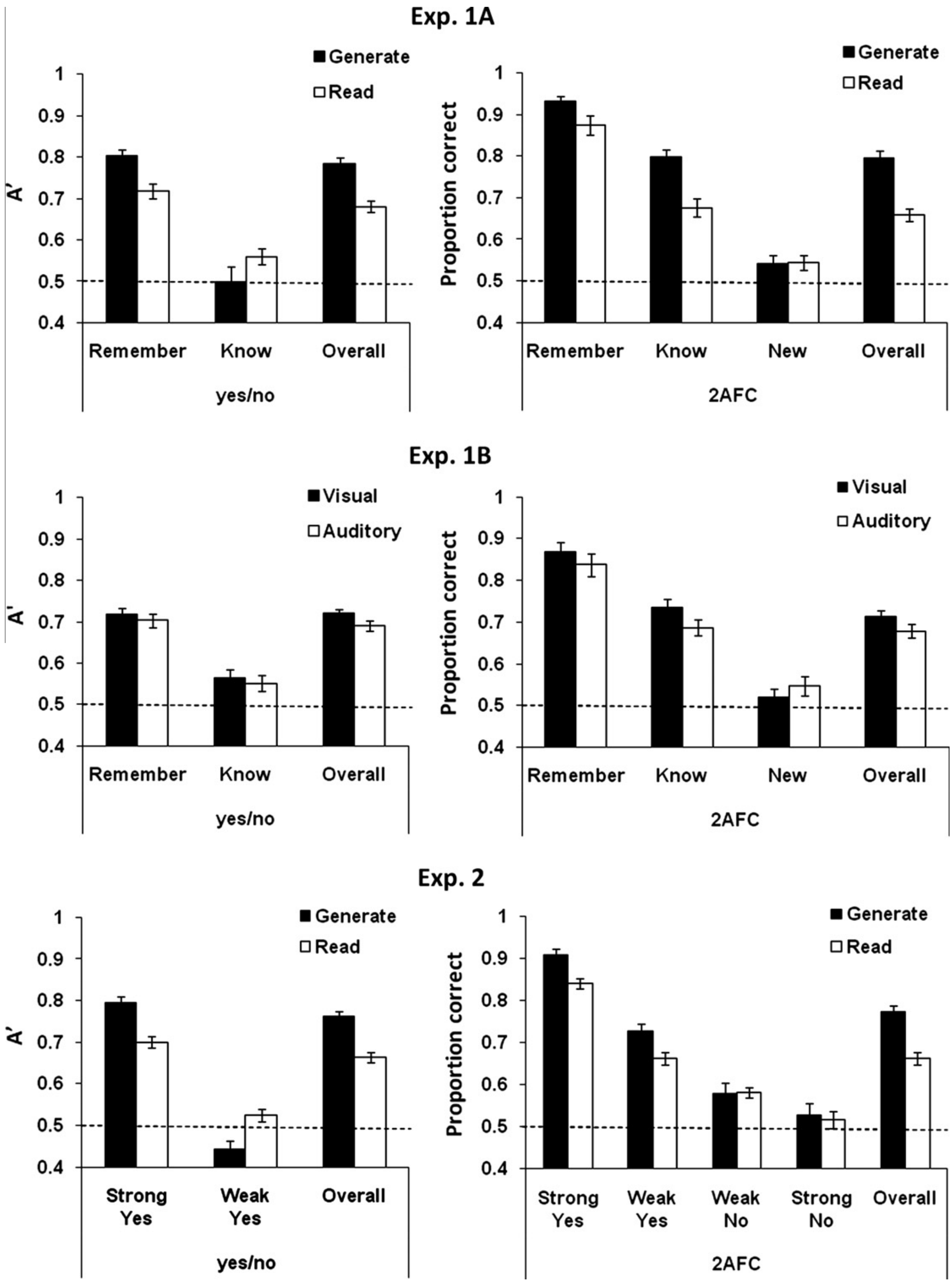

Fig. 2. Experiment 1A, Experiment 1B, and Experiment 2 mean yes/no A' sensitivity and 2AFC proportion correct performance by encoding condition and subjective response category.

remember and know trials, proportion correct performance was always numerically higher for the same modality trials than for the different modality trials. Furthermore, like with the 2AFC response proportion measure, this difference was significant for know trials, $t(31) = 3.84$, $p < .001$, and for overall performance, $t(31) = 2.94$, $p < .01$, but not for remember trials, $t(31) = 1.46$, $p = .15$.

Thus, when a manipulation with a strong effect on remembering was used (i.e., generation), there was no effect on knowing for the two response proportion measures. However, a know generation effect emerged for the proportion correct measure, suggesting that the effect of generation on knowing was underestimated by the RK literature's know response proportion measures. In contrast, when a manipulation with a smaller effect on remembering was used (i.e., modality), the proportion correct findings were consistent with the RK literature's response proportion measures. Taken together, the above findings are consistent with the Jacoby et al. (1997) view that knowing is underestimated in conditions with high levels of remembering.

2.2.3. A' versus proportion correct

To further investigate the discrepancy that emerged between the know proportion correct measure and the know response proportion measures, we converted the yes/no response proportion measure and the proportion correct measure to the same scale. This can be done either by converting both of these measures to a d' sensitivity measure, or by converting the yes/no response proportion measure into the A' sensitivity measure, which provides an approximation of 2AFC performance and is therefore on the same scale as the proportion correct measure.[4] As both of these analyses yielded a qualitatively similar pattern of results, only A' is reported here. Fig. 2 displays the A' means and standard errors (left side) and the proportion correct means and standard errors (right side). As can be clearly seen by an inspection of this figure, when overall and remember sensitivity are considered, the yes/no A' and 2AFC proportion correct measures produced an extremely similar pattern of results. In contrast, when know sensitivity is examined, a striking difference between yes/no A' and 2AFC proportion correct performance was observed in Experiment 1A, but not in Experiment 1B. In support of this pattern in the know sensitivity means in Experiment 1A, a 2 × 2 ANOVA revealed a significant interaction between measure type (yes/no A' versus 2AFC proportion correct) and encoding condition (generate versus read), $F(1, 31) = 30.15$, $p < .001$. Planned comparisons revealed that this interaction occurred because, for 2AFC proportion correct, performance was substantially better for generate trials as compared to read trials, $t(31) = 6.70$, $p < .001$, while for A' yes/no sensitivity there was a non-significant numerical trend in the opposite direction, $t(31) = 1.84$, $p = .08$. Thus, Experiment 1A showed a robust generation effect for 2AFC proportion correct performance coupled with a numerical trend of a reversed generation effect for the A' yes/no sensitivity measure. Such a pattern of results is consistent with the Jacoby et al. (1997) critique that the standard RK paradigm's response proportion measures tend to underestimate the effect of a variable on knowing, whenever the variable has a large effect on remembering. Furthermore, it appears that this underestimation issue might apply to the standard RK paradigm's know response proportion measure, as well as to know sensitivity measures that are derived from the yes/no hit and false alarm rates.

2.2.4. Correlations between sensitivity and response bias

An additional goal of the present study was to examine whether the present paradigm could replicate the Donaldson (1996) finding of a correlation between know sensitivity and response bias. Consequently, the overall recognition hit and false alarm rates from the yes/no task were used to calculate the appropriate measure of bias associated with A' (i.e., B''_D). For Experiment 1A, the mean value of B''_D was −.005 (range = −.90 to .68), while for Experiment 1B, the mean value of B''_D was −.246 (range = −.98 to .93). Importantly, in replication of Donaldson (1996), know sensitivity was significantly correlated with bias (Experiment 1A: $r(30) = .61$, $p < .001$; Experiment 1B: $r(30) = .60$, $p < .001$). In contrast, remember sensitivity was not correlated with bias (Experiment 1A: $r(30) = .20$, $p = .28$; Experiment 1B: $r(30) = .19$, $p = .31$). Thus, the present paradigm provides convergent evidence for the Donaldson (1996) critique concerning the RK know response proportion measure's susceptibility to bias even if converted into a signal detection sensitivity measure.

2.2.5. Memory in the absence of claimed awareness

Finally, the present study examined 2AFC recognition memory performance under conditions of claimed null awareness (i.e., following a new response). Fig. 2 contains the means and standard errors for proportion correct performance following a new response. Importantly, performance was still significantly above chance (i.e., >0.5) for new response trials, both in Experiment1A, $t(31) = 3.09$, $p < .01$, and in Experiment 1B, $t(31) = 2.14$, $p < .05$. Furthermore, in Experiment 1A, a qualitative difference was observed between memory in the presence versus the absence of claimed awareness. Specifically, following an old response (i.e., a remember or a know response), there was a generation effect such that proportion correct performance was higher for generate than for read trials, both for remember trials, $t(31) = 3.43$, $p < .01$, and for know trials, $t(31) = 6.70$, $p < .001$. In contrast, following a new response, proportion correct performance did not differ between the generate and read trials, $t < 1$. However, caution should be exercised in interpreting this pattern of results, because the absence of a generation effect following a new response might be partly due to floor effects in that condition. Consequently, as out-

[4] The yes/no hit and false alarm rates were converted into the A' sensitivity measure using the A' formula provided by Snodgrass and Corwin (1988):

$$A' = 0.5 + [(H - FA) * (1 + H - FA)]/[(4H * (1 - FA))]$$

FA = false alarms. Separate H and FA values were calculated for overall recognition, remember and know responses. Specifically, for the present paradigm, the remember hit rate is the proportion of "old new" trials that were given a remember response, the know hit rate is the proportion of "old new" trials given a know response, and the overall recognition hit rate is the sum of the remember and know hit rates. In a similar manner, the FA values are calculated based on the proportions of "new new" trials that received each type of response. Following Donaldson (1992), the following formula was used to calculate B''_D: $B''_D = [(1 - H) * (l - FA) - H * FA]/[(1 - H) * (l - FA) + H * FA]$.

lined below, Experiment 2 attempted to replicate the qualitative difference pattern shown in Experiment 1A under conditions in which floor effects are unlikely to occur.

3. Experiment 2

Experiment 2 was designed to test the generality of the main findings from Experiment 1, by applying the generate read manipulation from Experiment 1A to a version of the paradigm that employed a four-point confidence rating scale instead of remember, know and new responses. Specifically, in response to each pair of words at test, participants indicated whether either word was old (i.e., yes/no recognition) by selecting one of four options: "strong no", "weak no", "weak yes", or "strong yes". As in Experiment 1, participants then decided which word was old (i.e., 2AFC recognition). Using this confidence ratings paradigm, Experiment 2 attempted to replicate the main findings from Experiment 1.

For the "yes" side of the confidence scale, we were interested in examining whether the overall pattern of results from Experiment 1 can be replicated when participants are asked to subdivide their "yes" responses into "strong yes" and "weak yes" responses instead of remember and know responses. In particular, Experiment 2 will test whether the Jacoby et al. (1997) and Donaldson (1996) criticisms of the know response proportion measure also apply to a similar "weak yes" response proportion measure. In general, if the "weak yes" and the know response proportion measures are susceptible to similar problems, then such a pattern of results would imply that the problems stem from fundamental properties of the response proportion measure and the method of subdividing hit rates based on subjective responses, rather than from unique properties of the RK paradigm.

For the "no" side of the confidence scale, of particular interest was whether Experiment 2 could replicate and strengthen Experiment 1's finding of above chance 2AFC recognition memory performance under conditions of claimed null awareness (i.e., following a new response). Critically, in the field of unconscious perception, Merikle and Reingold (1990) noted that findings of above chance performance in the absence of subjective claimed awareness are quite prevalent, but it is not possible to interpret such findings as evidence of unconscious processing on an a priori basis. Instead, subjective measures of awareness must first be validated by demonstrating qualitative differences between processing with and without claimed awareness. Consistent with the qualitative differences logic, Experiment 1A did provide preliminary evidence that the present paradigm's claimed awareness and null awareness states were qualitatively different, by demonstrating a generation effect under conditions of claimed awareness (i.e., following a remember or a know response), but not in the absence of claimed awareness (i.e., following a new response). However, this dissociation is difficult to interpret because floor effects may have prevented the new response trials from showing a generation effect. Consequently, in Experiment 2, we attempt to rule out the possibility of floor effects, and provide finer grained information about performance under conditions of claimed null awareness, by including two types of null awareness responses (i.e., "strong no" and "weak no"), instead of a single "new" response as in Experiment 1. Of particular interest is whether a dissociation will occur between "weak yes" and "weak no" trials, as it is expected that 2AFC performance following these responses will be less susceptible to potential ceiling and floor effects, compared to responses at more extreme parts of the confidence scale.

3.1. Method

3.1.1. Participants

All 64 participants were undergraduate students at the University of Toronto. They were given either course credit or $10 (Canadian) per hour. All had normal or corrected-to normal vision.

3.1.2. Materials and procedure

Experiment 2 employed a generate read manipulation at study. The materials and procedure were identical to Experiment 1A, except that the test phase employed a confidence rating scale instead of the remember know instructions. Specifically, for each pair of words at test, participants responded to the first question ("Is either word old?") using a confidence scale that ranged from one to four (1 = "strong no", 2 = "weak no", 3 = "weak yes", 4 = "strong yes"). Next, and regardless of which of the four responses were given to question one, participants had to decide which of the two words was old, by pressing the up or the down arrow on the keyboard. The word pair remained on the screen for both questions, and a prompt at the top of the screen reminded the participants of the current question and of the available response options.

3.2. Results and discussion

The purpose of Experiment 2 was to examine if Experiment 1's main findings could be replicated using a confidence ratings version of the present paradigm. Accordingly, for each encoding condition (generate, read) and subjective response category (strong yes, weak yes, weak no, strong no), we derived the same measures as in Experiment 1. Thus, the analyses below are based on the following measures: (1) yes/no response proportions (i.e., the mean proportion of trials containing an old word that received each type of subjective response), (2) 2AFC response proportions (i.e., the number of correct trials that received each type of subjective response, divided by the total number of trials containing an old word), (3) proportion correct (i.e., the number of correct trials that received a particular subjective response, divided by the total number of trials that received this subjective response).

3.2.1. Response proportions and proportion correct

For the "yes" side of the confidence scale, Experiment 2 replicated the overall pattern of results from Experiment 1A. As can be seen by an inspection of Figs. 1 and 2, the "strong yes" category mirrored Experiment 1A's remember results by showing a generation effect across all measures (all *t*s > 3, all *p*s < .001) and the "weak yes" response category mirrored Experiment 1A's know results, by showing a discrepancy across measures such that there was a generation effect for the proportion correct measure, $t(62) = 3.88$, $p < .001$, but the response proportion measures showed an effect in the opposite direction, such that read was greater than generate (yes/no response proportion: $t(63) = 4.40$, $p < .001$; 2AFC response proportion: $t(63) = 1.27$, $p = .21$).

3.2.2. A' versus proportion correct

Furthermore, as can be seen from Fig. 2, the "weak yes" category's discrepancy across measures was also obtained when the yes/no response proportion measure was converted to A' to place it on the same scale as proportion correct performance. In support of this pattern in the "weak yes" sensitivity means, a 2×2 ANOVA revealed a significant interaction between measure type (yes/no A' versus 2AFC proportion correct) and encoding condition (generate versus read), $F(1, 62) = 37.49$, $p < .001$. Planned comparisons revealed that this interaction occurred because, for 2AFC proportion correct, performance was substantially better for generate trials as compared to read trials, $t(62) = 3.88$, $p < .001$, while for A' yes/no sensitivity there was a significant effect in the opposite direction (read > generate), $t(62) = 4.84$, $p < .001$. Thus, Experiment 2's response proportion measures may have failed to show a generation effect for "weak yes" because the large number of "strong yes" trials in the generate condition constrained the number of leftover trials that could receive "weak yes" responses.

3.2.3. Correlations between sensitivity and response bias

Furthermore, in replication of the know results from Experiment 1, there was a significant correlation between "weak yes" sensitivity and bias, $r(62) = .57$, $p < .001$, and in replication of Experiment 1's remember results, "strong yes" sensitivity was not correlated with bias, $r(61) = -.13$, $p = .33$. Overall, it appears that the Jacoby et al. (1997) and Donaldson (1996) critiques of the know response proportion measure are also applicable when a similar response proportion measure is calculated for the "weak yes" response category.

3.2.4. Memory in the absence of claimed awareness

For the "no" side of the confidence scale, we replicated Experiment 1's finding that 2AFC recognition memory performance can be above chance under conditions of claimed null awareness, by showing that proportion correct performance was above chance following both a "weak no" response, $t(62) = 5.52$, $p < .001$ and a "strong no" response, $t(57) = 2.16$, $p < .05$.[5] Furthermore, as can be seen from Fig. 2, we also replicated Experiment 1's pattern of a qualitative difference between proportion correct performance under conditions of claimed awareness versus null awareness, by showing a generation effect for "strong yes" and "weak yes" trials (all *t*s > 3, all *p*s < .001) but not for "strong no" and "weak no" trials (all *t*s < 1). Importantly, since the average performance level for "weak no" trials was 58%, it is unlikely that this pattern of results is simply due to floor effects.

4. General discussion

In line with past criticisms by Jacoby et al. (1997) and Donaldson (1996), the findings from the present study strongly suggest that caution is necessary when interpreting findings that are based on the RK paradigm's know response proportion measure. Most importantly, the present study raises serious concerns regarding the extensive and ongoing use of this measure as an index of familiarity (see Wixted and Mickes (2010) for a review). As a case in point, the present study challenges the widespread conclusion that the null effect of generation on the know response proportion measure is due to an absence of an influence of generation on familiarity. In direct opposition to this conclusion, the present paradigm's know proportion correct measure documented robust generation effects. We argue that a key advantage of the proportion correct measure is that, unlike the standard RK response proportion measure, it is free to vary independently across the different subjective response categories. Thus, we strongly advocate a critical reexamination of the interpretations of past findings concerning the impact of experimental manipulations on the know response proportion measure.

The present demonstration of an influence of generation on familiarity is also consistent with the results obtained in studies employing the Process Dissociation Procedure (Jacoby, 1991) and the Independent Remember Know (IRK) paradigm (Jacoby et al., 1997; see also Yonelinas & Jacoby, 1995). The latter approach utilizes the standard RK paradigm to derive estimates of recollection and familiarity. Specifically, while the remember response proportion measure is used directly as an estimate of recollection, the familiarity estimate is obtained by dividing the proportion of know (K) responses by one minus the proportion of remember (R) responses (i.e., $F = K/(1 - R)$). Interestingly, when we calculated the IRK

[5] Six participants were excluded from this analysis because they gave fewer than 5 "strong no" responses.

paradigm's familiarity estimate using Experiment 1A's data, we obtained a strong generation effect, which is in agreement with the present paradigm's know proportion correct findings.[6] Importantly, the IRK paradigm and the process dissociation procedure both incorporate the assumption that recollection and familiarity vary independently. This independence assumption is adopted by many dual-process theories (for a review, see Yonelinas, 2002), although alternative relational assumptions have also been proposed (e.g., Joordens & Merikle, 1993; Richardson-Klavehn, Gardiner, & Java, 1996). A unique advantage of the present paradigm is that, unlike the IRK paradigm and the process dissociation procedure, it provides a method for studying recognition memory performance under different reported subjective states of awareness without requiring an a priori relational assumption.

Another advantage of the present paradigm is that it permits an exploration of 2AFC recognition memory performance when participants claim to be unaware of the presence of an old stimulus. The present finding that participants can show above chance memory performance in the absence of claimed awareness may not be surprising in and of itself. However, the differential effects of generation on performance as a function of claimed awareness provided preliminary support for the notion that subjective reports could be successfully employed to distinguish between qualitatively different patterns of objective memory performance. Building on this finding, future research could explore additional qualitative differences between recognition memory with and without claimed awareness, and could employ variations on the present paradigm, such as the addition of a "guess" response option (Gardiner, Ramponi, & Richardson-Klavehn, 1998; Gardiner et al., 1996, 2002). Furthermore, our finding of above chance 2AFC performance in the absence of claimed awareness fits well with recent demonstrations of an unconscious component in 2AFC recognition memory performance (e.g., Voss, Baym, & Paller, 2008; Voss & Paller, 2009), and with well-established findings that densely amnesic patients show above chance 2AFC recognition memory performance in the absence of subjective phenomenal awareness of the study episode (for a review, see Aggleton and Shaw (1996)). Taken together, these findings suggest that the traditional focus on investigating the explicit or conscious memory component of recognition memory should be supplemented by greater emphasis on examining the implicit or unconscious component of performance in this task (see Reingold (2002) for related findings and discussion).

The present paradigm might provide a number of additional opportunities for future memory research. As illustrated in the present study, the paradigm can be used with both RK and confidence judgments and it permits the simultaneous assessment of yes/no and 2AFC recognition memory performance. Consequently, the present paradigm might prove useful in the context of the recent controversy concerning the extent to which yes/no and 2AFC tasks differ in terms of recognition accuracy and in terms of the relative contributions of recollection and familiarity (Bastin & Van der Linden, 2003; Cook, Marsh, & Hicks, 2005; Khoe, Kroll, Yonelinas, Dobbins, & Knight, 2000; Kroll, Yonelinas, Dobbins, & Frederick, 2002). Importantly, in contrast to these past studies, which assessed 2AFC and yes/no performance on separate trials, the present paradigm produces both measures of performance within a single trial. Thus, a methodological advantage of the present paradigm is that it matches retrieval conditions across the yes/no and 2AFC tasks, but the inevitable cost of this modification is that our yes/no task differs from the traditional yes/no task in that an extra distractor word is now present on every trial. In the present experiments, the presence of the distractor word likely did not diminish yes/no performance, as we consistently found no differences in overall recognition accuracy across the yes/no and 2AFC tasks.

In conclusion, the RK paradigm has played a pivotal role in integrating the measurement of subjective reports of phenomenal awareness as part of the assessment of recognition memory performance. As discussed earlier, we believe that our modified RK procedure has the potential of addressing the important theoretical questions which motivated the introduction of the original paradigm, without being subject to the methodological limitations that are inherent to the original paradigm's know response proportion measure. Moreover, our modified procedure provides a novel methodology for the exploration of recognition memory with and without claimed awareness.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.concog.2011.05.001.

References

Adams, J. K. (1957). Laboratory studies of behavior without awareness. *Psychological Bulletin, 54*(5), 383–405.

Aggleton, J. P., & Shaw, C. (1996). Amnesia and recognition memory: A re-analysis of psychometric data. *Neuropsychologia, 34*(1), 51–62.

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology, 17*(1), 14–24. doi:10.1037/0894-4105.17.1.14.

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2005). Revisiting the role of recollection in item versus forced-choice recognition memory. *Psychonomic Bulletin & Review, 12*(4), 720–725.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121*(3), 275–277.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*(4), 523–533.

[6] When we applied the Jacoby et al. (1997) Independence Remember Know (IRK) procedure to our yes/no data, by dividing the proportion of know (K) responses by one minus the proportion of remember (R) responses (i.e., $F = K/(1 - R)$), this IRK familiarity estimate showed the same pattern of results as the present paradigm's proportion correct measure. Specifically, in Experiment 1A, the IRK familiarity estimate was significantly higher for generate trials ($M = 0.56$, $SE = .03$) than for read trials ($M = 0.45$, $SE = .02$), $t(31) = 4.70$, $p < .001$ (for a related finding, see Turriziani, Serra, Fadda, Caltagirone, & Carlesimo, 2008) and in Experiment 1B, the IRK familiarity estimate was significantly higher for visual trials ($M = 0.59$, $SE = .02$) than for auditory trials ($M = 0.56$, $SE = .03$), $t(31) = 2.52$, $p < .05$.

Donaldson, W., MacKenzie, T., & Underhill, C. (1996). A comparison of recollective memory and source monitoring. *Psychonomic Bulletin & Review, 3*(4), 486–490.

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*(2), 524–542. doi:10.1037/0033-295X.111.2.524.

Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review, 115*(2), 426–446. doi:10.1037/0033-295X.115.2.426.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*(4), 309–313.

Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology, 50*(1), 114–122.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness & Cognition, 26*(7), 1–26.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory*, *10*(2), 83–98. doi:10.1080/09658210143000281.

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

Green, D. M., & Swets, J. A. (1974). Signal detection theory and psychophysics (vol. xiii). Oxford, England: RE Krieger Pub., Co.

Gregg, V. H., & Gardiner, J. M. (1994). Recognition memory and awareness: A large effect of study-test modalities on "know" responses following a highly perceptual orienting task. *European Journal of Cognitive Psychology, 6*(2), 131–147.

Heathcote, A., Bora, B., & Freeman, E. (2010). Recollection and confidence in two-alternative forced choice episodic recognition. *Journal of Memory and Language, 62*(2), 183–203. doi:10.1016/j.jml.2009.11.003.

Hicks, J. L., & Marsh, R. L. (1999). Remember-know judgments can depend on how memory is tested. *Psychonomic Bulletin & Review, 6*(1), 117–122.

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition, 25*(3), 345–351.

Inoue, C., & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition, 26*(2), 299–308.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513–541.

Jacoby, L. L., Yonelinas, A.P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In E. Jonathan, D. Cohen, & W. Schooler (Eds.), Scientific approaches to consciousness (p. 13–47). Hillsdale, NJ: Erlbaum.

Joordens, S., & Merikle, P. M. (1993). Independence or redundancy? Two models of conscious and unconscious influences. *Journal of Experimental Psychology: General, 122*(4), 462–467. doi:10.1037/0096-3445.122.4.462.

Khoe, W., Kroll, N. E., Yonelinas, A. P., Dobbins, I. G., & Knight, R. T. (2000). The contribution of recollection and familiarity to yes–no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia, 38*(10), 1333–1341.

Kroll, N. E. A., Yonelinas, A. P., Dobbins, I. G., & Frederick, C. M. (2002). Separating sensitivity from response bias: Implications of comparisons of yes–no and forced-choice tests for models and measures of recognition memory. *Journal of Experimental Psychology: General, 131*(2), 241–254. doi:10.1037//0096-3445.131.2.241.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed., Vol. xix). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Malmberg, K. J., Zeelenberg, R., & Shiffrin, R. M. (2004). Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by midazolam. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *30*(2), 540–549. doi:10.1037/0278-7393.30.2.540.

Merikle, P. M., & Reingold, E. M. (1990). Recognition and lexical decision without detection: unconscious perception? *Journal of Experimental Psychology: Human Perception and Performance, 16*(3), 574–583.

Parkin, A. J., & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging, 7*(2), 290–298.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*(1), 89–102.

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*(2), 365–377.

Reingold, E. M. (2002). On the perceptual specificity of memory representations. *Memory (Hove England), 10*(5–6), 365–379. doi:10.1080/09658210244000199.

Reingold, E. M. (2004). Unconscious perception: Assumptions and interpretive difficulties. *Consciousness & Cognition, 13*(1), 117–122. doi:10.1016/j.concog.2003.11.002.

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics, 44*(6), 563–575.

Reingold, E. M., & Merikle, P. M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind and Language, 5*(1), 9–28.

Reingold, E. M., & Sheridan, H. (2009). Dissociation methods. In P. Wilken, T. Bayne, & A. Cleeremans (Eds.), *Oxford companion to consciousness* (pp. 230–233). Oxford, UK: Oxford University Press.

Reingold, E. M., & Toth, J. P. (1996). Process dissociations versus task dissociations: A controversy in progress. In G. Underwood (Ed.), *Implicit cognition* (pp. 159–202). Oxford: Oxford University Press.

Richardson-Klavehn, A., Gardiner, J. M., & Java, R. I. (1996). Memory: Task dissociations, process dissociations, and dissociations of consciousness. In G. F. Underwood (Ed.), *Implicit cognition* (pp. 85–158). Oxford: Oxford University Press.

Rotello, C. M., Macmillan, N. A., Hicks, J. L., & Hautus, M. J. (2006). Interpreting the effects of response bias on remember-know judgments using signal detection and threshold models. *Memory & Cognition, 34*(8), 1598–1614.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34–50.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne, 26*(1), 1–12.

Turriziani, P., Serra, L., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2008). Recollection and familiarity in hippocampal amnesia. *Hippocampus, 18*(5), 469–480. doi:10.1002/hipo.20412.

Voss, J. L., Baym, C. L., & Paller, K. A. (2008). Accurate forced-choice recognition without awareness of memory retrieval. *Learning & Memory, 15*(6), 454–459.

Voss, J. L., & Paller, K. A. (2009). An electrophysiological signature of unconscious recognition memory. *Nature Neuroscience, 12*(3), 349–355. doi:10.1038/nn.2260.

Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience, 20*(3), 400–405. doi:10.1162/jocn.2008.20041.

Wippich, W. (1992). Implicit and explicit memory without awareness. *Psychological Research, 54*(3), 212–224.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*(4), 1025–1054. doi:10.1037/a0020874.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*(4), 616–641.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441–517. doi:10.1006/jmla.2002.2864.->.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: effects of size congruency. *Journal of Memory and Language, 34*, 622–643.