

Remember-know judgments can depend on how memory is tested

JASON L. HICKS and RICHARD L. MARSH
University of Georgia, Athens, Georgia

Remember-know judgments reflect the subjective state of awareness that accompanies episodic memory retrieval. We tested an old-new recognition condition, an old-new recognition followed by remember-know judgment condition, and a simultaneous remember-know-new judgment condition. These three conditions were tested for both a short (1-sec) and a long (4.5-sec) study duration. Reassuringly, results from the first two conditions did not differ from each other. Results from the third condition, however, differed from those in the first two conditions at both long and short study durations. Simultaneous consideration of all three alternatives resulted in a markedly liberal response bias, both in recognition detection and in the ascriptions of remember and know judgments. Discussion of the results is framed in terms of the single-process signal detection models that have been proposed to account for these subjective states of awareness.

There has been considerable interest lately in the subjective states of awareness that arise when people engage in different recollective tasks. Most notably, Tulving (1985) introduced the distinction of remembering versus knowing in episodic memory retrieval. Remembering is defined as a true state of conscious recollection, whereas knowing is defined as feelings of familiarity that arise in the absence of recollection. Originally, Tulving asked for these *remember-know* (R-K) judgments following free recall, cued recall, and recognition. The proportion of remember judgments dropped across these three tasks, and the proportion of know responses increased. Subsequently, theorists have requested R-K judgments primarily following recognition memory decisions (e.g., Gardiner, 1988; Gardiner, Gawlik, & Richardson-Klavehn, 1994; Gardiner & Java, 1990). Their use following recognition judgments has basically led researchers to equate the "remember-know paradigm" with tests of item recognition (but see Conway & Dewhurst, 1995, and Mather, Henkel, & Johnson, 1997, for R-K judgments in a source-monitoring task). Donaldson (1996) and Hirshman and Master (1997) independently modeled R-K judgments using an underlying signal detection theory (SDT) analysis (for a critique of this analysis, see Gardiner & Gregg, 1997).

Our point in this brief article was neither to question the validity of the R-K procedure (e.g., Strack & Förster, 1995) nor to debate whether remembering and knowing

as assessed in this paradigm reflects conscious versus unconscious processing as might be depicted from, say, the process dissociation technique (e.g., Jacoby, 1991). These debates concerning one process versus two are being played out elsewhere (e.g., Jacoby, Yonelinas, & Jennings, 1997, vs. Gardiner, Java, & Richardson-Klavehn, 1996). Rather, our point in this article was to explore one methodological issue in the R-K paradigm and examine its consequences for the SDT analyses offered both by Hirshman and Master (1997) and by Donaldson (1996). To date, the empirical evidence suggests that various depth of processing judgments will have dissociable effects on judgments (recent summaries can be found in Gardiner et al., 1996, and Rajaram & Roediger, 1997). For example, deeper encoding generally results in more Rs than shallow processing (Gardiner, 1988) and can lead to fewer Ks as well (Rajaram, 1993). Some of these dissociations may be attributable to how participants assign their guesses or whether guessing is explicitly discouraged (see Gardiner et al., 1996).

Although there have been several studies that have examined specific aspects of the test sequence, to our knowledge, the character of the recognition test itself has received very little attention in this literature. For example, Rajaram (1993) found that masked repetition priming at test increased "know" responses. Dewhurst and Conway (1994) found longer latencies to make "know" judgments than to make "remember" judgments. Gardiner et al. (1996) tested two-alternative forced-choice recognition. Finally, Hirshman and Henzler (1998) manipulated what participants believed to be true about the proportion of old test items. Other than these four reports, there have been no direct investigations concerning how the characteristics of the test sequence might affect performance in the R-K paradigm.

For example, we could not find one instance of an empirical comparison between standard old-new recognition

The authors thank Travis Nelson, Shari Nevins, Beth Stewart, Mark Weisner, Dianne James, Kristen Slette, Jennifer Cyran, and Rebecca Green for their dedicated help in collecting the data. The authors also thank Roddy Roediger, John Gardiner, Elliot Hirshman, and an anonymous reviewer for their helpful comments. J. L. Hicks is now at Louisiana State University. Correspondence should be addressed to R. L. Marsh, Department of Psychology, University of Georgia, Athens, GA 30602-3013 (e-mail: jhicks@lsu.edu or marsh@meme.psy.uga.edu).

memory and the use of the R-K procedure in conjunction with recognition. Perhaps this is because most researchers believe that the R-K judgment is a postrecognition judgment that is unaffected by, and also does not affect, the recognition judgment that is made prior to rendering an R-K judgment (e.g., Donaldson, 1996). However, there is no uniform procedure for collecting these sorts of data. Gardiner and his colleagues traditionally have used the two successive binary judgments (first old-new [O-N], then R-K), whereas others essentially have had participants choose one of three alternatives simultaneously: remember-know-new (R-K-N; e.g., Lindsay & Kelley, 1996; Parkin & Rousseau, 1993; Perfect, Mayes, Downes, & Van Eijk, 1996). The three testing procedures theoretically should yield equivalent rates of O-N recognition performance. Likewise, the successive O-N followed by R-K should lead to equivalent levels of Rs and Ks relative to the R-K-N testing procedure. We believe that there are principled reasons why one or both of these equivalences may not hold empirically.

First, an R-K judgment need not be a postrecognition judgment. When tested with the three-alternative response format of R-K-N, the test question is very similar to a source-monitoring judgment with two old alternatives and one new (e.g., Johnson, Hashtroudi, & Lindsay, 1993). Just as there is no agreement on whether a person must obligatorily make an O-N judgment prior to a source judgment, so too it need not be the case that a recognition judgment "precedes" an R-K judgment when testing occurs with the R-K-N format (see the Discussion section of Marsh & Landau, 1995, as it relates to this issue in source monitoring). Second, psychologically speaking, weighing the evidence for each of the two successive judgments seems somewhat different, and perhaps easier, than simultaneously weighing the evidence for the three alternatives when subjective states are measured with the alternative R-K-N judgment. Thus, the purpose of this article was to compare performance on O-N recognition with the two methods of collecting R-K judgments (i.e., successive judgments or simultaneous R-K-N). Of course, we cannot know whether participants in the R-K-N condition make two judgments successively or make only one judgment, but we can assess whether performance differs in these two conditions that measure subjective states of awareness.

To understand why these conditions might differ, we have reproduced as Figure 1A the general SDT model used by both Hirshman and Master (1997) and Donaldson (1996). This figure depicts the distributions of old and new items residing in memory with differing mean familiarity values. For convenience, they are depicted as Gaussian with equal variances (both questionable assumptions, but they will suffice for the present analysis). For expositional considerations, we have adopted the notation of Hirshman and Master. According to these SDT models, participants set a criterion *C* above which they will label an item "old." This criterion leads to the standard hit and false-alarm rates of O-N recognition. In an R-K exper-

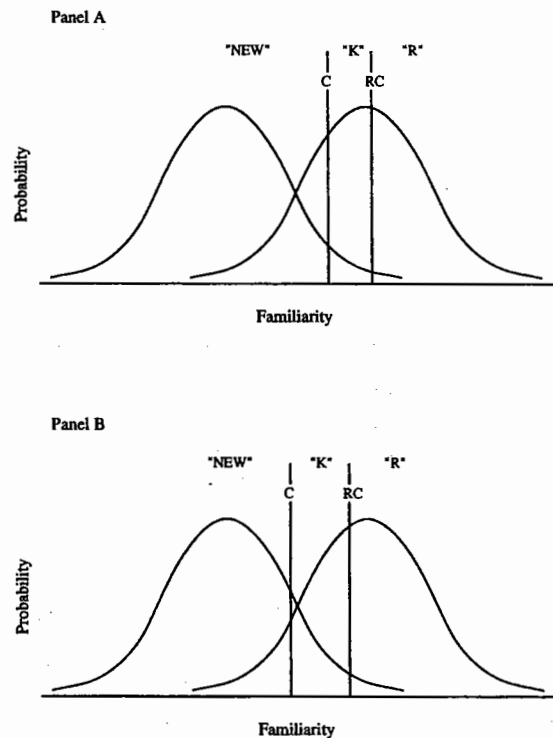


Figure 1. Donaldson's (1996) and Hirshman and Master's (1997) dual-criterion model of the remember-know paradigm. (A) Two decision criteria: the standard signal detection criterion *C* above which items are labeled "old" and the *RC* criterion above which items are labeled "remembered." (B) More liberally placed decision criteria as a consequence of the R-K-N test methodology being investigated.

iment according to the SDT models, participants set an additional criterion *RC*. If items have assessed familiarity above the *RC* criterion, they are labeled "remembered." Items with assessed familiarity falling between the two decision criteria are labeled "know." Donaldson (1996) has shown that changes in the criterion *C* are consistent with changes in the proportions of Rs and Ks. He has also shown that both criteria can be empirically calculated.

In the best of circumstances for those interested in subjective states of awareness, the placement of *C* and *RC* would be independent of whether the O-N and the R-K judgments were made successively or combined into a single R-K-N judgment. Our intuition, however, is that the R-K-N procedure is slightly more difficult than two sequential decisions. If we are correct, then more difficult recognition conditions often lead to more liberally placed decision criteria. This criterion shift has been demonstrated in the older signal detection literature by making discriminations more difficult under conditions of smaller signal-to-noise ratios (e.g., Clarke, 1960; Healy & Jones, 1973; Pollack & Decker 1958). Because the signal is more difficult to perceive, the detector must relax the criterion in order to achieve a similar hit rate. The consequence of a more liberal criterion

does increase the hit rate but also increases the false-alarm rate. We have recently argued that the same analysis may hold true with more difficult recognition judgments resulting in more liberal criterion shifts (see Hicks & Marsh, 1998).

We have depicted these shifts in Figure 1B as shifts in both criteria because we have assumed that this would be the consequence of R-K-N testing. That result need not come to pass. The different forms of testing R-K judgments theoretically could affect either *C* or *RC* independently or could affect both as we have assumed. If criterion placement is different with the different types of test formats, then shifts in states of awareness should be associated with the two types of testing. Moreover, these criterion shifts can occur only if the Hirshman and Master (1997) and Donaldson (1996) models accurately depict one way to conceptualize the underlying psychological mechanism of how R-K judgments are rendered. Thus, although our experiment was motivated by an empirical question, the results could have important theoretical ramifications for these SDT models of the R-K paradigm. Stated slightly differently, the manipulations of levels of processing that have resulted in changes in the d' measure of discriminability in previous research are consistent with the model, but only Hirshman and Henzler (1998) have manipulated the decision criteria in these models with a test manipulation. We turn now to the procedural details.

METHOD

Participants

Two-hundred ten students volunteered in exchange for partial course credit toward fulfilling a research requirement. The participants were randomly assigned to one of three between-subjects conditions that differed only in how the recognition memory test was conducted. Equal numbers of participants (i.e., 70) were randomly assigned on the basis of their arrival at the laboratory to the following conditions: (1) the standard old-new recognition condition, (2) the old-new recognition judgment followed by remember-know judgments on those items first identified as old, and (3) the single remember-know-new testing condition. Henceforth, we will refer to these three conditions with the short-hand notations of O-N, O-N-then-R-K, and R-K-N, respectively. Half of the participants in each of these three conditions (i.e., 35) were tested with a short study duration, and half were tested with a long study duration. The participants were tested individually in sessions that lasted approximately 25 min.

Materials and Procedure

The stimuli were 112 common nouns selected from Thorndike and Lorge's (1944) corpus. Fifty-six were used for study, and the remaining 56 served as lures on the 112-item recognition test. Because R-K judgments require several minutes to "stabilize" (Rajaram, 1993) and because we did not want ceiling performance, we introduced a 12-min distractor activity between encoding and test. This activity consisted of rating Gary Larson's *The Far Side* cartoons for humor on a 5-point Likert scale.

The participants were seated in front of an IBM-compatible computer monitor and were asked to pay attention to a list of words for an unspecified memory test. The 56 study words were chosen randomly by the software for each participant, with the remaining 56

items serving as lures on the recognition memory test. During study in the long condition, the words were presented at a 5.25-sec rate, the duration of which was composed of a 250-msec fixation point and warning tone, a 4.5-sec study period, and a 500-msec intertrial interval (ITI). In the short condition, the study period was 1 sec. After the encoding phase, the 12-min cartoon rating task was administered. The cartoons were photocopied one to a page and placed in a binder with sheet protectors. The experimenter paced the participants with a stopwatch at a rate of 20 sec per cartoon. The participants circled the numbers 1 (*not funny*) through 5 (*very funny*) on a rating form.

Following distraction, the participants then read instructions on the computer for the recognition test. These instructions were tailored to each of the respective experimental conditions. The participants in the O-N condition were asked to press a key labeled *OLD* for items they remembered being on the study list and to press a key labeled *NEW* for items they decided were not on the study list. The participants in the O-N-then-R-K condition were given additional instructions to indicate for items chosen as old whether they were remembered (i.e., associated with conscious recollection) or simply known (i.e., seemed familiar in the absence of recollective information) by pressing a second set of keys labeled *R* or *K*, respectively. The R-K instructions closely followed those used by Gardiner (1988) and Rajaram (1993), and these instructions were followed up in a brief conversation with the experimenter about the distinction prior to testing. Additional examples of what would constitute an *R* and a *K* were provided by the experimenter. The R-K-N condition required that the participants respond by pressing only one of three labeled keys (*R*, *K*, or *NEW*) to indicate whether an item was remembered, known, or new, respectively. The participants in the R-K-N condition received the identical instructions used in the O-N-then-R-K condition with the exception of how to respond. Thus, the only difference between these two conditions was that two binary decisions were made in the O-N-then-R-K condition, whereas a single judgment was called for in the R-K-N condition. Each of the three types of recognition tests was self-paced, with each trial being terminated by the participant's response, which was followed by a 500-msec ITI.

RESULTS AND DISCUSSION

Unless specified otherwise, the probability of an alpha error did not exceed 5%. We begin first with the traditional signal detection analysis of performance that is presented in Table 1. Next, we discuss the pattern of R-K judgments for the two conditions in which those data were collected as summarized in Table 2. The data in Table 1 were analyzed with a series of 2×2 analysis of variance (ANOVA) models with the factors of study duration (short and long) and type of test (O-N, O-N-then-R-K, and R-K-N). In no cases was the interaction term significant; therefore, we do not report them. As seen in Table 1, there was a main effect in the hit rate for study duration [$F(1,204) = 17.0$], but there were also differences due to type of test administered [$F(2,204) = 5.7$]. The same was true of the false-alarm rate for study duration [$F(1,204) = 32.6$] and type of test [$F(2,204) = 9.2$]. The differences due to study duration simply reflect a higher hit rate in the long condition than in the short condition and a greater false-alarm rate in the short condition than in the long condition. These differences were expected. The interesting difference is reflected in the fact

Table 1
Standard Recognition Performance and Signal Detection
Measures Across the Three Types of Recognition Tests

Condition	Hit	FA	d'	C	A'	B_D''
Short Study Duration						
O-N	0.73 (.02)	0.23 (.02)	1.51 (.12)	0.08 (.04)	0.63 (.01)	0.18 (.08)
O-N-then-R-K	0.71 (.02)	0.21 (.02)	1.49 (.09)	0.14 (.06)	0.61 (.01)	0.22 (.08)
R-K-N	0.77 (.02)	0.31 (.03)	1.37 (.10)	-0.12 (.06)	0.60 (.01)	-0.14 (.08)
Long Study Duration						
O-N	0.77 (.02)	0.12 (.02)	2.16 (.12)	0.23 (.06)	0.70 (.02)	0.35 (.07)
O-N-then-R-K	0.79 (.02)	0.12 (.02)	2.29 (.14)	0.25 (.07)	0.71 (.02)	0.37 (.08)
R-K-N	0.84 (.02)	0.20 (.03)	2.15 (.15)	-0.05 (.06)	0.70 (.02)	0.00 (.08)

Note—Standard errors are in parentheses. O-N refers to old-new recognition; O-N-then-R-K refers to remember-know judgments taken subsequent to recognition; R-K-N refers to a simultaneous remember-know-new response format. FA, false alarms. d' and C are measures of discriminability and criterion. A' and B_D'' are the nonparametric alternatives of discriminability and criterion, respectively.

that the O-N condition and the O-N-then-R-K condition did not differ from each other, but the R-K-N condition differed significantly from those two conditions. These results were confirmed with Student-Newman-Keuls post hoc tests.

The data suggest that the R-K-N testing method increased both the hit rate and the false-alarm rate above the other two methods of testing. We combined these measures into the standard measures of discriminability (d') and bias (C being preferred over β). We have also reported the analogous nonparametric detection measures A' and B_D'' , respectively, that have sometimes been considered better measures (e.g., Donaldson, 1992). Because the three recognition test conditions differed only in what the participants were asked to do at test, we did not expect differences in the d' measure of discriminability, and none were found [$F(2,204) < 1.0$, n.s.]. This null result was corroborated by the A' measure [$F(2,204) < 1.0$,

n.s.]. Of course, discriminability was higher for the long study condition than for the short study condition [for d' , $F(1,204) = 55.4$; A' , $F(1,204) = 48.4$]. Consistent with our intuitions, conditions that made for more difficult recognition decisions produced more liberal shifts in bias. More difficult decisions due to shorter study durations produced more liberal criteria than did decisions due to longer study durations [for C and B_D'' , the smaller $F(2,204) = 5.73$]. The critical result was that differences in bias were reflected by a marked liberal tendency in the R-K-N condition relative to the other two conditions [for C , $F(2,204) = 14.2$; for B_D'' , $F(2,204) = 13.2$].

One reassuring aspect of these data is that taking R-K judgments subsequent to O-N recognition (i.e., the O-N-then-R-K condition) does not influence the criterion C (or B_D'') and therefore does not influence performance relative to standard recognition (i.e., the O-N condition). This is the first direct test of which we are aware that has confirmed what many already have assumed about assessing subjective states of awareness. However, combining the two sequential binary judgments into a single R-K-N judgment resulted in a substantial liberal criterion shift that led to greater hit and false-alarm rates. The question of interest now becomes whether this method of testing affected people's reported states of awareness as remember versus know. Having found that C (and B_D'') is placed more liberally in the R-K-N condition, we now address whether the RC criterion shifts as well (see Figure 1).

Table 2 sets forth the relevant data. RHit and KHit refer to the hit rates labeled "remember" and "know," respectively, and the same convention is used for the false-alarm rates (i.e., RFA and KFA). We chose to analyze the hit rates and false-alarm rates separately focusing on differences between "remember" versus "know" responses for each type of item. For hits, we conducted a $2 \times 2 \times 2$ mixed-model ANOVA with the factors of study duration (long vs. short), type of test (O-N-then-R-K vs. R-K-N) and subjective report (remember vs. know). These correspond to the data in the RHit and KHit columns of Table 2. Although the effect of long versus short study duration

Table 2
Remember-Know Performance Measures
for Hits and False Alarms and Calculated Criterion Values

Condition	RHit	RFA	KHit	KFA	Rd'	RC	RA'	RB_D''
Short Study Duration								
O-N-then-R-K	0.41 (.03)	0.05 (.01)	0.30 (.02)	0.16 (.02)	1.51 (.12)	1.02 (.07)	0.56 (.01)	0.90 (.03)
R-K-N	0.53 (.03)	0.10 (.01)	0.24 (.03)	0.21 (.02)	1.53 (.12)	0.68 (.07)	0.60 (.01)	0.73 (.04)
Long Study Duration								
O-N-then-R-K	0.49 (.03)	0.02 (.01)	0.30 (.02)	0.10 (.02)	2.08 (.11)	1.05 (.05)	0.60 (.01)	0.96 (.01)
R-K-N	0.54 (.04)	0.06 (.01)	0.30 (.03)	0.14 (.02)	1.85 (.14)	0.80 (.07)	0.62 (.01)	0.81 (.05)

Note—Standard errors are in parentheses. RHit and RFA refer to hits and false alarms called "remembered." KHit and KFA refer to the same but called "known." Measures Rd' and RC are discriminability and criterion values; RA' and RB_D'' are nonparametric alternatives.

was significant, demonstrating a higher hit rate with more encoding time [$F(1,136) = 15.6$], the important finding was that R-K claims depended on how memory was tested [$F(1,136) = 9.4$]. More hits were labeled "remembered" under R-K-N testing than under O-N-then-R-K testing. Although the greater propensity to label items as remembered in the R-K-N test appears to have been offset by a smaller propensity to label the hits known in the short study condition, the associated interaction term was not significant. These data suggest that the testing condition caused a more liberal criterion shift, and more items were labeled as remembered. In other words, there was a criterion shift affecting the reported state of awareness for items identified as old.

In an identical analysis of the false alarms (see the RFA and KFA columns of Table 2), there was a main effect of study duration, reflecting a higher false-alarm rate in the short condition than in the long condition [$F(1,136) = 18.7$]. This result was expected, as was the finding that the participants labeled more of their false alarms as known than as remembered [$F(1,136) = 85.7$]. However, the interesting result was that more false alarms were labeled as remembered and more false alarms were labeled as known under R-K-N testing than under O-N-then-R-K testing [$F(1,136) = 13.0$]. This result suggests that the *RC* criterion became more liberal in the R-K-N condition, as was shown for the *C* criterion (see Figure 1B).

This difference in *RC* between the two conditions can be empirically calculated (see Donaldson, 1996). In Table 2, we also presented Rd' , which represents the discriminability of items labeled as remembered, and *RC*, which represents a measure of bias for just these items labeled as remembered (i.e., the upper criterion in Figure 1). RA' represents the nonparametric alternative for discriminability, and RB_D'' is the bias measure for these remembered items. For Rd' , the encoding manipulation affected discriminability [$F(1,136) = 13.2$], but the test manipulation did not [$F(1,136) < 1$]. The identical result held for the nonparametric RA' , but there was a statistically marginal effect due to the type of testing procedure [$F(1,136) = 3.27, p = .06$]. From our perspective, the important measures are those relating to bias. For *RC* and RB_D'' , the criterion measures were more liberal under R-K-N testing [the smaller $F(1,136) = 17.3$]. Thus, it would appear that the liberal shift in responding in the R-K-N condition relative to the O-N-then-R-K condition did affect the upper decision criterion and the lower decision criterion. Together, the standard signal detection analyses along with these measures for the "remember" versus "know" responses suggest that how one tests will lead to different results. We turn now to the implications of this experiment.

GENERAL DISCUSSION

There are two important methodological points made by these data. First, administering R-K judgments does not appear to change performance relative to standard O-N recognition. Second, this finding holds true only

when the recognition judgment and the R-K judgment are separated into two sequential judgments. When they are combined into a single R-K-N judgment, a liberal bias is introduced. According to the signal detection accounts offered by Donaldson (1996) and Hirshman and Master (1997), that bias could have affected either the *C* criterion or the *RC* criterion in Figure 1. As it turned out, the test manipulation affected both decision criteria. We can only speculate on the underlying psychological mechanism for this effect because we do not know positively whether or not the participants in the R-K-N condition made two sequential judgments. Our intuition is that they did not, because, otherwise, there should have been no differences among the three conditions tested, contrary to the data.

One simple mechanism that might account for these effects is that people are considering two old alternatives versus one new alternative in the R-K-N condition. When two response options (R and K) favor responding "old" and one favors responding "new," perhaps people have a tendency to equate or otherwise even out their responses among the three alternatives (see Greene, 1996). However, large differences between the use of remember versus know response options in previous research articles, and in this experiment, would argue that people do not have any problem using one option much more frequently than another. Therefore, although this possibility might account for our results, it is not favored from these data. As described momentarily, we believe that the difficulty of simultaneously entertaining the three alternatives of the R-K-N format is somewhat greater than entertaining the evidence for the two successive binary judgments. Differences in performance have been found with source-monitoring judgments, where two binary questions about source led to a different pattern of responding than entertaining all alternatives simultaneously (e.g., Dodson & Johnson, 1993). A similar outcome seems to have occurred in our experiment.

Judgments can be made more difficult in a variety of ways. In terms of the present experiment, weighing all of the evidence to render an R-K-N judgment involves simultaneous consideration of three response alternatives and the evidence in the memory trace that might be consistent with each alternative. Compared with a relatively quick and heuristic O-N decision, followed by a more reflective, but still binary R-K decision, we believe that sheer consideration of the options is more difficult with an R-K-N judgment. However, it would be unsatisfying if we did not then operationalize how difficulty affected the placement of the hypothetical criteria posited in Hirshman and Master's (1997) and Donaldson's (1996) models. For that, we appeal to the signal-to-noise ratio, as did Hicks and Marsh (1998). On any given R-K-N trial, the signal value is interpreted by weighing that evidence against three possible alternatives. These alternatives constitute the background noise against which the signal is weighed. Therefore, the signal-to-noise ratio might be smaller on R-K-N trials (larger background noise) than weighing the same evidence against only two

alternatives (smaller background noise). Thus, the mechanism being posited to account for this liberal criterion shift is identical to that demonstrated in the older signal detection literature using perceptual discriminations of a signal against background noise. The more difficult the trial, the more leniently placed is the criterion in order to avoid missing a weaker signal. This tendency also has the unfortunate consequence of creating more false alarms, as we found in this experiment. Although this account fits with the existing signal detection literature, it does not contravene other hypotheses. For example, the R-K-N method of testing may simply introduce a presupposition that the item is old.

This experiment demonstrated at least one important point: subjective judgments of states of awareness can be biased by the test methodology. To the extent that previous articles have reported that encoding manipulations affect R-K judgments, we have now demonstrated that these judgments can be biased at test as well (see also Hirshman & Henzler, 1998, and Rajaram, 1993). Does this limit the generalizability or usefulness of the R-K paradigm? Probably not. But, it does suggest that test conditions that change performance in a manner consistent with the detection models provide some empirical support for those models. The notion that a criterion can be placed along an underlying continuum of, say, familiarity is indeed at odds with the all-or-none depiction of recollection (see also Yonelinas & Jacoby, 1995). Had the RC criterion been found to be invariant, the interpretation of the results would be more straightforward for those who hold the all-or-none assumption strongly. In whatever manner these issues sort themselves out, the results of our experiment suggest that, in the interim, researchers interested in remember-know judgments should give some careful consideration to how those judgments are empirically collected and the different testing parameters that could affect the outcome of their experiments.

REFERENCES

- CLARKE, F. R. (1960). Confidence ratings, second-choice responses and confusion matrices in intelligibility tests. *Journal of the Acoustical Society of America*, *32*, 35-46.
- CONWAY, M. A., & DEWHURST, S. A. (1995). Remembering, familiarity, and source monitoring. *Quarterly Journal of Experimental Psychology*, *48A*, 125-140.
- DEWHURST, S. A., & CONWAY, M. A. (1994). Pictures, images, and recollective experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 1088-1098.
- DODSON, C. S., & JOHNSON, M. K. (1993). Rate of false attributions depends on how questions are asked. *American Journal of Psychology*, *106*, 541-557.
- DONALDSON, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*, 275-277.
- DONALDSON, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*, 523-533.
- GARDINER, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309-313.
- GARDINER, J. M., GAWLIK, B., & RICHARDSON-KLAVEHN, A. (1994). Maintenance rehearsal affects knowing, but not remembering; elaborative rehearsal affects remembering, not knowing. *Psychonomic Bulletin & Review*, *1*, 107-110.
- GARDINER, J. M., & GREGG, V. H. (1997). Recognition memory with little or no remembering: Implications for a detection model. *Psychonomic Bulletin & Review*, *4*, 474-479.
- GARDINER, J. M., & JAVA, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, *18*, 23-30.
- 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*, 114-122.
- GREENE, R. L. (1996). Mirror effect in order and associative information: Role of response strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 687-695.
- HEALY, A. F., & JONES, C. (1973). Criterion shifts in recall. *Psychological Bulletin*, *79*, 335-340.
- HICKS, J. L., & MARSH, R. L. (1998). A decrement to familiarity interpretation of the revelation effect from forced-choice tests of recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 1105-1120.
- HIRSHMAN, E., & HENZLER, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, *9*, 61-65.
- HIRSHMAN, E., & MASTER, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition*, *25*, 345-351.
- JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, *30*, 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 J. Cohen & J. W. Schooler (Eds.), *Scientific approaches to the study of consciousness* (pp. 13-47). Hillsdale, NJ: Erlbaum.
- JOHNSON, M. K., HASHTRUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28.
- LINDSAY, D. S., & KELLEY, C. M. (1996). Creating illusions of familiarity in a cued-recall remember/know paradigm. *Journal of Memory & Language*, *35*, 197-211.
- MARSH, R. L., & LANDAU, J. D. (1995). Availability in cryptomnesia: Assessing its role in two paradigms of unconscious plagiarism. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 1568-1582.
- MATHER, M., HENKEL, L. A., & JOHNSON, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, *25*, 826-837.
- PARKIN, A. J., & RUSSO, R. (1993). On the origin of functional differences in recollective experience. *Memory*, *1*, 231-237.
- PERFECT, T. J., MAYES, A. R., DOWNES, J. J., & VAN EUIK, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *Quarterly Journal of Experimental Psychology*, *49A*, 797-813.
- POLLACK, I., & DECKER, L. (1958). Confidence ratings, message reception, and the receiver operating characteristics. *Journal of the Acoustical Society of America*, *30*, 286-292.
- RAJARAM, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*, 89-102.
- RAJARAM, S., & ROEDIGER, H. L., III (1997). Remembering and knowing as states of consciousness during recollection. In J. Cohen & J. W. Schooler (Eds.), *Scientific approaches to the study of consciousness* (pp. 213-240). Hillsdale, NJ: Erlbaum.
- STRACK, F., & FÖRSTER, J. (1995). Reporting recollective experiences: Direct access to memory systems? *Psychological Science*, *6*, 352-358.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College Press.
- TULVING, E. (1985). Memory and consciousness. *Canadian Psychologist*, *40*, 1-12.
- YONELINAS, A. P., & JACOBY, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory & Language*, *34*, 622-643.