Memory & Cognition
2010, 38 (4), 389-406
doi:10.3758/MC.38.4.389

Can corrective feedback improve recognition memory?

Justin Kantner and D. Stephen Lindsay
University of Victoria, Victoria, British Columbia, Canada

An understanding of the effects of corrective feedback on recognition memory can inform both recognition theory and memory training programs, but few published studies have investigated the issue. Although the evidence to date suggests that feedback does not improve recognition accuracy, few studies have directly examined its effect on sensitivity, and fewer have created conditions that facilitate a feedback advantage by encouraging controlled processing at test. In Experiment 1, null effects of feedback were observed following both deep and shallow encoding of categorized study lists. In Experiment 2, feedback robustly influenced response bias by allowing participants to discern highly uneven base rates of old and new items, but sensitivity remained unaffected. In Experiment 3, a false-memory procedure, feedback failed to attenuate false recognition of critical lures. In Experiment 4, participants were unable to use feedback to learn a simple category rule separating old items from new items, despite the fact that feedback was of substantial benefit in a nearly identical categorization task. The recognition system, despite a documented ability to utilize controlled strategic or inferential decision-making processes, appears largely impenetrable to a benefit of corrective feedback.

Previous studies examining the effects of corrective feedback on recognition memory are minimal in number and widely scattered across the last 30 years. The near absence of a literature on this topic is conspicuous, especially given the longstanding attention both to uncovering the mechanisms of recognition memory (see S. E. Clark, 1999; Yonelinas, 2002) and to investigating the effects of feedback on various cognitive performance measures (see Kluger & DeNisi, 1996, for a review), including recall (e.g., Bjork, 1994; Pashler, Cepeda, Wixted, & Rohrer, 2005). The potential informativeness of feedback effects for recognition theory and for the development of memory training/rehabilitation programs suggests that a detailed investigation into the issue is overdue. This article describes several experiments collectively meant to provide groundwork for such an investigation.

In examining the effects of feedback on recognition, the present work is concerned with changes in old/new discrimination, rather than the learning of responses to specific exemplars. Gardner, Sandoval, and Reyes (1986) observed a large feedback-based improvement in an old/new recognition task, but their design included three presentations of the same test list; thus, feedback most likely increased accuracy by teaching participants the correct responses to the repeated items, not by enhancing old/new discrimination per se (see also W. C. Clark & Greenberg, 1971). Furthermore, the present experiments were designed to address the question of whether feedback can influence old/new discrimination at *test*. In some previous studies (reviewed below), continuous recognition or multiple study–test cycles have been employed to assess feedback effects, leaving open the question of whether such effects arose strictly at test or whether they arose because feedback at test influenced subsequent study processing (Estes & Maddox, 1995; Han & Dobbins, 2008; Jennings & Jacoby, 2003; Rhodes & Jacoby, 2007). In the experiments reported here (and in a few previously published studies described below; Titus, 1973; Verde & Rotello, 2007), a single test list followed a single study list, and each test item was presented only once.

From the perspective of signal detection theory (Green & Swets, 1966; Parks, 1966), feedback might be expected to affect recognition memory test performance by either or both of two means: guiding participants to establish a more appropriate response criterion and increasing old/new sensitivity. Feedback would seem most likely to facilitate criterion placement when base rates of old and new items are unequal, a condition not typically included in recognition experiments but more common in studies of category learning (e.g., Kruschke, 1996). By conveying the correct responses on a trial-by-trial basis, feedback should enable participants to tune in to underlying probabilities of old and new items and to adjust criterion accordingly (Estes & Maddox, 1995; Titus, 1973). The impact of feedback on response bias when base rates of old and new items are equal has been the subject of more recent work (Rhodes & Jacoby, 2007; Verde & Rotello, 2007), which we review below. Existing models of recognition include parameters that index response criterion, but as Estes and Maddox noted, little progress has been

J. Kantner, jkantner@uvic.ca

made toward a formal account of how criterion is adjusted (see also Whittlesea, 2002a). These models can be refined, therefore, by information regarding the effect of feedback on response criterion.

Although the present study examined the effects of feedback on response criterion, our primary aim was to determine whether feedback can enhance sensitivity to the differences between old and new items. Unlike a criterion shift, which can occur without effecting an improvement in accuracy, an increase in sensitivity necessarily entails better performance and is thus of central interest in terms of the application of feedback to memory training programs. Theories proposing that consciously controlled recall-like retrieval processes can contribute to recognition judgments (dual-process models; e.g., Jacoby, 1991; Mandler, 1980) or theories that emphasize attribution-making processes in recognition (e.g., Jacoby, Kelley, & Dywan, 1989; Johnson, Hashtroudi, & Lindsay, 1993; Lindsay, 2008; Whittlesea, 2002b) allow multiple potential avenues for improvement in recognition sensitivity via feedback. As Han and Dobbins (2008) noted, feedback might selectively reinforce responses that are based on the use of effective retrieval strategies or might sharpen participants' interpretations of their internal responses to test items such that they glean from these responses information that is diagnostic of oldness or newness (Dodson & Johnson, 1993; Gruppuso, Lindsay, & Kelley, 1997; Jacoby, Yonelinas, & Jennings, 1997; Jennings & Jacoby, 1993). Evidence of just such effects of feedback has been obtained in tasks manipulating participants' interpretations of cognitive fluency (Unkelbach, 2006, 2007) and on source-monitoring performance in an eyewitness suggestibility paradigm (Lane, Roussel, Villa, & Morita, 2007). Benefits of feedback for sensitivity could also conceivably arise in single-process models of recognition memory (e.g., by selectively reinforcing more effective ways of encoding test probes).

Previous Research on Feedback and Recognition

In early work on the factors affecting response bias in recognition, Titus (1973) gave trial-by-trial feedback during recognition of CVC trigrams. Only 20% of the test items were old. Feedback significantly increased response criterion, but no compelling evidence of an impact on sensitivity was obtained. These results demonstrate the potential utility of feedback in cases of highly uneven base rates, although the use of only 15 old items (each of which was viewed six times during study) likely limited the variance of the signal distribution relative to that of the noise distribution, complicating the interpretation of the reported computations of sensitivity and bias (Titus, 1973).[1]

Estes and Maddox (1995) used a continuous recognition test and manipulated three between-subjects variables: proportion old (.33 or .67), stimulus type (digits, letters, or words), and trial-by-trial accuracy feedback (provided or withheld). For digit and letter stimuli, bias was more liberal in the .67-old condition and more conservative in the .33-old condition for participants receiving feedback than for controls. Feedback did not, however, affect response bias for words. Estes and Maddox proposed that base rates are learned more slowly for words than for digits and letters but offered no rationale for this account.

Estes and Maddox (1995) obtained only small and marginally significant positive effects of feedback on sensitivity for digits and letters and no effect on sensitivity for words. Note that even if marginal sensitivity effects in the Estes and Maddox study represent a true benefit of feedback, its locus is ambiguous: Because Estes and Maddox used a continuous recognition test, feedback may have influenced encoding instead of or in addition to recognition processes.

Jennings and Jacoby (2003) utilized feedback in a novel training procedure for improving recognition test performance in elderly participants. The training procedure was designed to use trial-by-trial feedback to develop participants' ability to reject test-item repetitions at increasing lags. Jennings and Jacoby (2003) found dramatic reductions in false alarms to repeated items by the end of training. Although a feedback-absent control condition was not included in the study, the observed reduction was presumably due at least in part to the presence of corrective feedback. Importantly, however, feedback did not result in an improvement in overall recognition accuracy; performance was enhanced only in rejection of repeated lures. In addition, Jennings and Jacoby (2003) used multiple study–test cycles, leaving open the question of whether feedback affected study and/or test processes.

Four recent studies have examined the effect of feedback on criterion placement in recognition (Han & Dobbins, 2008, 2009; Rhodes & Jacoby, 2007; Verde & Rotello, 2007). Rhodes and Jacoby conducted a recognition test in which the probability of a probe's oldness (.67 or .33) and its location on the computer screen covaried. With trial-by-trial feedback, participants were able to tune in to this correspondence and make dynamic trial-by-trial response criterion adjustments appropriate to presentation location.

Verde and Rotello (2007) divided recognition test trials into two blocks: one containing *strong* items studied at long durations (or on multiple occasions) and the other containing *weak* items studied at short durations (or only once). Strong items were assumed to produce a more conservative response criterion than weak items; therefore, participants sensitive to the change in item strength at the onset of the second block were expected to adjust criterion accordingly. Verde and Rotello did not manipulate the presence of feedback within experiments, but across-experiment comparisons indicated that the participants only made such adjustments with the aid of feedback (Experiment 5).

Han and Dobbins (2008) demonstrated that feedback can produce dynamic criterion shifting without manipulations of target base rates (Estes & Maddox, 1995; Rhodes & Jacoby, 2007) or target strength (Verde & Rotello, 2007). These researchers elicited liberal or conservative bias shifts across test blocks by giving false positive feedback either to false alarms (the lax criterion condition) or misses (the conservative criterion condition). Participants generally appeared unaware of the inaccuracies in the feedback, leading Han and Dobbins (2008) to conclude that feedback, at least under some circumstances, can guide implicit learning of adaptive criterion placement. Accurate feedback, however, did not affect response bias (Han & Dobbins, 2008, Experiment 1), presumably because the base rates of old and new items did not differ. In a follow-up study, Han and Dobbins (2009) demonstrated that implicitly learned criterion shifts can occur when false feedback is given probabilistically instead of deterministically (Experiment 1) and when feedback to false alarms or misses is omitted (Experiment 2). The latter finding indicates that feedback need not be false to elicit criterion shifts as long as the probability of positive feedback is sufficiently weighted in favor of a particular response.

Both Verde and Rotello (2007) and Han and Dobbins (2008) reported null effects of feedback on recognition sensitivity (because no feedback-absent control condition was reported in the Han & Dobbins, 2009, or Rhodes & Jacoby, 2007, articles, sensitivity effects could not be assessed). In summary, then, these recent results harmonize with those of earlier published reports in suggesting that recognition sensitivity is not influenced by feedback. Feedback appears to enhance recognition performance only for repeated test items, for which the knowledge of correct responses to particular exemplars conveyed through feedback may be utilized when those same items reappear (Gardner et al., 1986; Jennings & Jacoby, 2003). However, with the exception of the small and marginally significant positive effects reported in the digit and letter conditions of the Estes and Maddox (1995) experiments, the small literature to date suggests that feedback fails to enhance a generalized sensitivity to the differences between old and new items and is thus of little or no benefit to recognition accuracy when items are not repeated at test.

The accumulated evidence as to the potential of feedback to improve recognition is far from conclusive. From a dual-process or attribution-making standpoint, a feedback advantage would be most likely under conditions conducive to recollective, inferential, and/or strategic processing at test. By contrast, past studies in this domain have used a traditional recognition design in which the study list is composed of an essentially miscellaneous assortment of items. In such experiments, the participants have little obvious means of organizing members of the study set along thematic dimensions or, consequently, of using such organization to inform recognition judgments at test. Lacking a strategic basis for making judgments, the participants may default to a more automatic, familiarity-based approach to recognition (see, e.g., Jacoby et al., 1997). Although past experiments may have been best suited to studying the effects of feedback on automatic recognition processes, their designs may have precluded the strategic components to recognition that are proposed to be engaged or improved by feedback.

A more complete understanding of feedback effects in recognition requires tests of feedback under conditions that promote inferential or recollective test processing. The experiments reported in this article were designed to provide such tests. In Experiment 1, we used categorized study lists with the aim of increasing confusability between studied and nonstudied category exemplars and thereby increasing the extent to which recognition memory test performance would rely on recollection as opposed to familiarity. In Experiment 2, we employed a base-rate manipulation akin to that of Estes and Maddox (1995) to determine whether participants could strategically regulate responses after learning through feedback that most items were old (or new). Observing such an effect on bias would sharpen the informativeness of a null effect of feedback on sensitivity because it would demonstrate that the participants were attending to and (in some ways) using the feedback and that the experiment had sufficient power to detect effects of feedback. In Experiment 3, we implemented a false-memory procedure in which critical test items were highly familiar lures; it was hypothesized that feedback would teach participants not to respond to these items solely on the basis of their familiarity but to engage a more controlled process, such as retrieval, to verify their study list presence. In Experiment 4, we tested whether participants could use feedback to learn a simple category rule that distinguished old from new items and to respond strategically on the basis of that rule. These manipulations were intended to provide stronger tests of feedback effects on recognition sensitivity than have been reported in the literature to date.

In the present experiments, we also addressed questions regarding the relationship between feedback and response bias with word stimuli. Han and Dobbins (2008, 2009) provided the only published demonstrations of criterion shifts without manipulations of old/new base rates, but only manipulations of the reinforcement (feedback) schedule produced these effects. Of the studies in which base rates were manipulated, Rhodes and Jacoby (2007) and Verde and Rotello (2007) found feedback-based criterion shifts in recognition of words, whereas Estes and Maddox (1995) did not. This pattern of findings suggests that a base-rate manipulation alone is not a sufficient basis for criterion shifting via feedback; rather, the base-rate distinction must be made more salient (e.g., via contextual cues; Rhodes & Jacoby, 2007), the items themselves must possess some readily apparent feature diagnostic of oldness (e.g., strength; Verde & Rotello, 2007), or the base rates of the feedback itself must be manipulated (Han & Dobbins, 2008, 2009). The present experiments provide several tests of the potential of complete and accurate feedback to influence response bias, without overt contextual or featural cues.

Importantly, the present experiments were designed to provide unambiguous information as to the locus of any sensitivity or bias effects observed. In nearly all published studies on feedback and recognition, repeated study–test cycles have been employed (Estes & Maddox, 1995; Han & Dobbins, 2008; Jennings & Jacoby, 2003; Rhodes & Jacoby, 2007), or the studies have lacked a control group receiving no feedback (Han & Dobbins, 2009; Jennings & Jacoby, 2003; Verde & Rotello, 2007). In each of the present experiments, a single test phase followed a single study phase (such that sensitivity effects could only accrue at test), and to ensure that any effects obtained resulted from feedback and not other factors (e.g., practice), each experiment included a feedback-absent comparison group.

EXPERIMENT 1

The main purpose of Experiment 1 was to provide a test of feedback effects in old/new recognition when the nature of the study set encouraged a recollective strategy at test. Participants studied words from 10 semantic categories (e.g., birds) and then took a standard recognition test in which the studied words were intermixed with new words from the same categories. The participants were given either corrective feedback after each response (for the first two thirds of the test) or no feedback. Lures were expected to elicit a sense of familiarity because of their categorical association with multiple studied items; it was hypothesized that negative feedback in response to familiarity-driven false alarms would train the participants to adopt a more deliberate approach to making recognition judgments (e.g., searching memory for evidence that an individual test item, and not others of its category, had been studied).

A second goal was to address Estes and Maddox's (1995) finding of small increases in recognition sensitivity with feedback when the stimuli were digits and letters but no effect when the stimuli were words. The relatively poor performance in the digit and letter conditions against the high level of performance in the word condition (see above) raises the question of whether feedback can benefit recognition only when overall levels of accuracy are low (a point also raised by Han & Dobbins, 2008). Perhaps the difficulty of the digit and letter conditions in the Estes and Maddox experiments led the participants in the no-feedback condition to respond haphazardly, whereas those receiving feedback put more effort into making their recognition judgments. Conversely, the high level of performance in the word condition may indicate that these participants did not need to pay attention to the feedback (and in any case, there was little room for improvement). Whereas Rhodes and Jacoby (2007) and Verde and Rotello (2007) recently reported performance levels near those of the digit and letter conditions of the Estes and Maddox experiments and did not observe a benefit of feedback to sensitivity, they did not directly manipulate discriminability or cross feedback and discriminability manipulations within a single experiment. In Experiment 1, we did so by having the participants study half of the items at a deep level of processing (LOP) and half at a shallow LOP (Craik & Lockhart, 1972).

Method

Participants. In each of the present experiments, the participants were University of Victoria undergraduates volunteering in exchange for bonus credit in an introductory psychology course. There were 46 participants in Experiment 1. The participants were evenly divided between the feedback and control conditions.

Materials. Two versions of the study list were created, each consisting of six words from each of 10 semantic categories (e.g., *trees*, *birds*). Twelve exemplars from each of 10 categories of concrete nouns were selected from the Battig and Montague (1969) norms. In general, the 12 most frequent exemplars of each category were selected, excluding words whose dominant meaning was inconsistent with the corresponding category (e.g., *cherry* was not used as an exemplar of *tree*), as well as long (more than eight letters), orthographically unusual, or very low-frequency words. The order of items on the study lists was randomized, with adjustments made thereafter to ensure that no 2 exemplars of a given category occurred in immediate succession. Finally, 6 additional concrete nouns were used as primacy and recency buffers during the study phase.

The test list consisted of the 60 studied words randomly intermixed with the 60 nonstudied words, with the constraints that no more than 2 words of a given *old–new* status and no more than a single word from a given semantic category occurred in immediate succession and that there be approximately equal numbers of old and new words and exemplars of each category in each third of the test. Blocks of test items were rotated across thirds of the test, such that across participants in each condition, each test item appeared equally often in each block.

The experiment was conducted with Micro Experimental Laboratory (Schneider, 1988).

Procedure. In each of the present experiments, participants were tested individually and were randomly assigned to the feedback or control condition. The procedure was identical for the two conditions, with the exception of the presence or absence of feedback at test. Study instructions informed the participants that each word on the study list would be preceded by an instruction as to how that word was to be studied: by judging whether the word contained the letter "a" (the shallow LOP task) or by judging whether the object named by the word would fit in a shoebox. Assignment of items to orienting tasks was counterbalanced across participants. The deep and shallow LOP study trials were randomly intermixed anew for each participant. On each trial, the study instruction appeared on the screen for 1 sec before the word appeared, and then both remained on the screen until a response was entered. The participants responded by saying "yes" or "no"; responses were keyed in by the experimenter. The study items were separated by a blank 1-sec interval. The study and test phases were separated by a 5-min delay interval during which the participants wrote down the names of as many countries as they could.

Test instructions correctly informed the participants that the test consisted of approximately equal numbers of studied and nonstudied words and that the nonstudied words were drawn from the same semantic categories as the studied words. In the feedback condition, the instructions went on to explain that, for the first two thirds of the test, a message would appear on the screen after each recognition response indicating whether that item had, in fact, been studied. The feedback participants were asked to attend to this message, because doing so might improve their performance.

Recognition judgments were collected via a 6-point, confidence-graded scale (1, *sure not studied*; 2, *probably not studied*; 3, *maybe not studied*; 4, *maybe studied*; 5, *probably studied*; 6, *sure studied*). Responses were spoken aloud and keyed in by an experimenter. Test items appeared in the center of the screen along with the response scale and remained on the screen until a response was entered. In the feedback condition, entry of the response triggered immediate feedback (the word "Studied" if the test word had been studied or the words "Not Studied" if the test word had not been studied) that remained in the center of the screen along with the word for 1 sec. Only the studied–not-studied decision (not the level of confidence) was relevant to the type of feedback received: Any response of 1–3 to a new item or 4–6 to an old item was counted as correct and given positive feedback, and any response of 1–3 to an old item or 4–6 to a new item was counted as incorrect and given negative feedback. Feedback was followed by a 1-sec blank interval, after which the next test trial began. In the control condition, responses were followed by a blank screen for 600 msec, an asterisk presented in the center of the screen for 800 msec, and an additional blank 600-msec interval, after which the next test trial began. Thus, the two conditions contained an equivalent intertrial interval (2,000 msec).

When participants had completed two thirds of the test phase, a message appeared on the screen informing them of their progress and, for those in the feedback condition, indicating that feedback

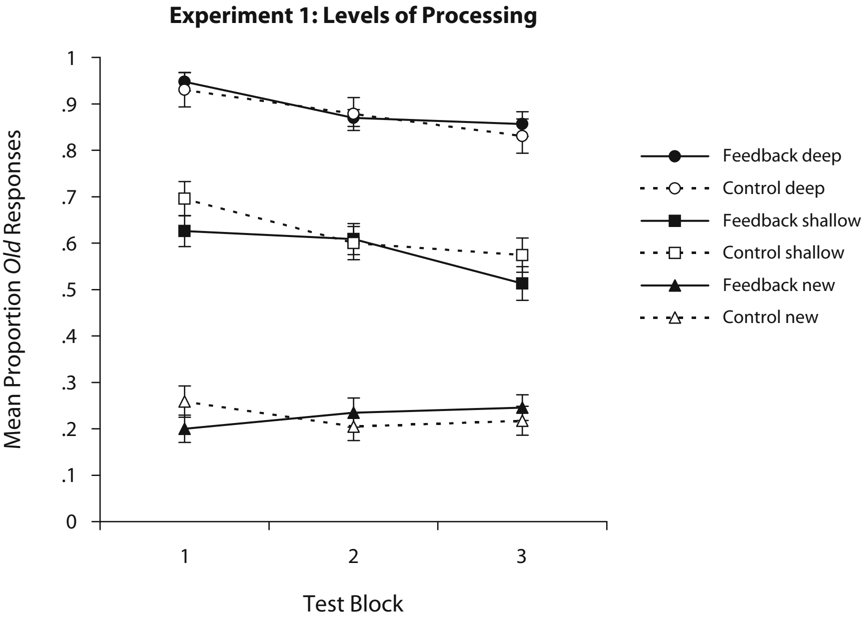

Figure 1. Mean proportion of *old* responses (4, 5, or 6 on the response scale) as a function of depth of processing and test block in Experiment 1. Error bars represent one standard error of the mean.

information would not be presented during the final third of the test. In this portion of the test phase, the procedure of the control condition was adopted for both feedback and control groups.

Results and Discussion

The recognition ratings data were converted to hits and false alarms by scoring responses of 4, 5, or 6 as hits for studied items and as false alarms for nonstudied items. To uncover any trends in the effects of feedback over the course of the test phase, hit and false alarm rates were analyzed by test block and are displayed as such in Figure 1.[2] These data were submitted to a 3 (study status: deep encoding, shallow encoding, or no encoding [new items]) × 3 (test block: 1, 2, or 3) × 2 (feedback: present or absent) mixed-factor ANOVA. In this and subsequent analyses of hit and false alarm rates, a main effect of feedback may be taken to reflect a change in response bias (e.g., an increase in both hits and false alarms), whereas a study status × feedback interaction indexes a change in performance (i.e., a greater difference between hit rates and false alarm rates in the feedback condition than in the control condition or vice versa). In Experiment 1, the expected effect of study status was significant [$F(2,88) = 662$, $MS_e = 0.023$, $p < .001$, $\eta_p^2 = .938$]; however, there was no evidence of a study status × feedback interaction ($p = .33$), indicating that the feedback and control groups performed similarly, regardless of the level at which items were encoded. The trend for performance to worsen across blocks was significant [$F(2,88) = 8.885$, $MS_e = 0.020$, $p < .001$, $\eta_p^2 = .168$], but this trend did not interact with feedback ($p = .35$). A significant study status × block interaction reflected the tendency for performance to decline more across blocks for old items than for new items [$F(4,176) = 3.781$, $MS_e = 0.013$, $p < .01$, $\eta_p^2 = .079$], but this tendency did not further interact with feedback ($p = .26$). The main effect of feedback was also nonsignificant ($p = .67$), suggesting that the feedback and control groups displayed similar response bias.

We measured old/new recognition discrimination by calculating A_z. This index is derived from the receiver operating characteristic (ROC), a plot of pairs of hit and false alarm rates as a function of the confidence with which those judgments were made. The area under the best-fitting curve connecting these points (A_z) indexes sensitivity, typically ranging between .5 (chance discrimination) and 1 (perfect discrimination). A_z is a well-supported estimate of sensitivity (Verde, Macmillan, & Rotello, 2006) and has recent precedent in work on feedback and recognition (Han & Dobbins, 2008; Verde & Rotello, 2007). A_z values were calculated using the Web-based program JROCFIT (Eng, n.d.).

Response bias was indexed using the ROC-based c_a measure (Macmillan & Creelman, 2005). Estimates of the response criterion at each level of confidence on the ROC were calculated using the Web-based RscorePlus program (Harvey, 2005). Negative values of c_a indicate a liberal response bias, whereas positive values reflect a conservative bias. For each of the present experiments, a 5 (ROC confidence level) × 2 (feedback) ANOVA assessed the overall effects of feedback on the response criterion, as well as changes in the magnitude of these effects across criterial confidence levels. In each experiment, criterion estimates rose as a function of rising confidence levels (all ps < .001), an expected property of c_a that we will not mention in discussing the results of the individual experiments.

Table 1
Mean Sensitivity (A_z) and Response Criterion (c_a) Values for the Feedback and Control Groups in Experiments 1–4

			Response Criterion									
	A_z		c_1		c_2		c_3		c_4		c_5	
Condition	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
					Experiment 1							
Feedback	0.81	0.01	−0.64	0.16	−0.19	0.07	0.05	0.05	0.32	0.05	0.63	0.08
Control	0.84	0.02	−0.89	0.12	−0.24	0.06	0.04	0.06	0.40	0.06	0.89	0.08
					Experiment 2							
.75-old												
Feedback	0.88	0.01	−1.15	0.17	−0.80	0.13	−0.34	0.04	0.00	0.08	0.21	0.12
Control	0.91	0.01	−1.18	0.17	−0.59	0.09	−0.18	0.06	0.21	0.08	0.51	0.12
.25-old												
Feedback	0.85	0.02	−0.71	0.20	−0.32	0.11	0.11	0.04	0.46	0.08	0.73	0.14
Control	0.88	0.01	−1.36	0.18	−0.69	0.13	−0.21	0.08	0.19	0.09	0.66	0.11
					Experiment 3							
Feedback	0.89	0.02	−0.84	0.14	−0.26	0.09	0.22	0.04	0.60	0.06	0.95	0.09
Control	0.91	0.01	−0.83	0.11	−0.20	0.08	0.12	0.06	0.45	0.06	0.79	0.06
					Experiment 4							
CRR												
Feedback	0.78	0.02	−1.12	0.17	−0.53	0.13	−0.03	0.05	0.36	0.07	0.87	0.14
Control	0.81	0.02	−1.21	0.15	−0.43	0.10	0.08	0.08	0.55	0.08	0.99	0.09
SR												
Feedback	0.72	0.02	−1.22	0.20	−0.60	0.19	−0.02	0.06	0.45	0.10	0.86	0.13
Control	0.74	0.02	−1.07	0.16	−0.42	0.11	0.05	0.10	0.43	0.09	0.79	0.10

Note—Values at c_1–c_5 represent response criterion at each level of confidence drawn from the ROC curve. Negative (positive) values indicate a liberal (conservative) response criterion. For Experiment 3, the results displayed are for #2–#5 exemplars only (all #1 exemplars were new). CRR, category-rule recognition; SR, simple recognition.

Greenhouse–Geisser correction was used for violations of sphericity, which were common across confidence levels in the c_a data.

The mean A_z and c_a values for the feedback and control conditions of Experiments 1–4 are presented in Table 1. In Experiment 1, A_z was directionally but nonsignificantly *lower* in the feedback condition than in the control condition, providing no indication of any benefit of feedback ($p = .33$). The analysis of c_a estimates revealed a marginal confidence level × feedback interaction that reflected the trend for participants receiving feedback to establish a more conservative criterion than the controls at the lowest confidence level, a more liberal criterion than controls at the highest confidence level, and approximately the same criterion as the controls at intermediate confidence levels [$F(1.179, 51.868) = 3.140$, $MS_e = 0.440$, $p = .08$, $\eta_p^2 = .067$]. There was no main effect of feedback ($p = .91$), however, indicating that bias was generally unaffected by feedback.

The results of Experiment 1 were straightforward: Feedback exerted no discernible effect on sensitivity or response bias, even though the categorized nature of the study list was expected to elicit improved recollective test processing through feedback. This null effect was observed regardless of whether the overall level of sensitivity was high (deep LOP condition) or lower (shallow LOP condition). These findings suggest that the failure of feedback to increase sensitivity in the word condition of the Estes and Maddox (1995) study was not the result of ceiling-level performance. Rather, these results suggest that feedback simply does not influence recognition sensitivity.

EXPERIMENT 2

Experiment 1 provided no indication of feedback effects on either sensitivity or response bias. Experiment 2 was focused on the question of whether feedback can produce strategic biases in recognition of words without any manipulation beyond that of old item base rates (see the introduction). Following Estes and Maddox (1995), the test contained either mostly old or mostly new items, and the participants were not informed of the base-rate disparity a priori. Categorized study lists were not used. Unlike the Estes and Maddox study, Experiment 2 consisted of a single study list followed by a single test list. To increase the salience of the base-rate manipulation, the proportion of old items was changed from .67 (or .33) in the Estes and Maddox study to .75 (or .25) in the present experiment.

In addition to testing the Estes and Maddox (1995) findings, this design addressed the possibility that the null effects in Experiment 1 resulted from inattention to the feedback. If feedback affected response criterion, we could be confident that the participants were, in fact, attending to and using feedback but had simply not derived a benefit from such processing in Experiment 1. Evidence that the participants could make use of feedback would allow a clearer interpretation of its effect (or lack thereof) on sensitivity.

In Experiment 2, we also introduced design modifications intended to increase the potential utility of feedback for old/new discrimination. The length of the test was extended to 240 trials (from 120 in Experiment 1), and feedback was delivered throughout test (as opposed to only the first two thirds of the test). These measures were taken to increase the likelihood that effects of feedback on performance had sufficient time to accrue. In addition, the feedback itself was augmented to become more informative and more salient. Whereas in Experiment 1 the feedback simply consisted of the words "Studied" or "Not Studied," test responses in Experiment 2 were followed by a message conveying the nature of the error or correct response (e.g., "No, wrong, that item was NOT on the study list!" following a false alarm) accompanied by a tone that differed for correct and incorrect responses. More minor modifications for Experiment 2 are described in the Method section.

Finally, in this and each subsequent experiment in the article, we report reaction time (RT) differences between the feedback and control conditions (RT data were not available in Experiment 1 because responses were entered by the experimenter). Although the focus of the present work was on feedback effects on sensitivity and response bias, RTs may also reveal an influence of feedback on recognition processing. For example, even if feedback does not result in increased accuracy, it might enable the participants to become more efficient at weighing decision evidence, an alternative form of improvement that would be revealed in decreased RTs relative to controls (a possibility not examined in prior publications on effects of feedback on recognition). By contrast, slower RTs in the feedback condition might result if the prospect of feedback causes participants to more carefully consider their judgments before responding.

Method

Participants. There were 72 participants in Experiment 2. The data of 1 participant performing near chance on the recognition test were removed from subsequent analyses, leaving 18 participants in the .75-old/feedback, .75-old/control, and .25-old/feedback conditions and 17 participants in the .25-old/control condition.

Materials. The stimuli were 360 four- to eight-letter English nouns of medium frequency selected from the MRC psycholinguistic database (Coltheart, 1981). Study lists were composed of 180 words selected randomly and without replacement from the pool of 360. A unique study list was constructed for each participant. Three primacy and three recency buffers were included in the study phase. The buffers consisted of words not included in the pool of 360 but adhering to the same specifications. Test lists were composed of the 180 studied items plus 60 randomly selected nonstudied items (.75-old condition) or the 180 nonstudied items plus 60 randomly selected items from the study list (.25-old condition), producing a 240-item test. The stimuli appeared in the center of a computer screen in black lettering against a white background. The experiment was conducted with DMDX experimental software (Forster & Forster, 2003).

Procedure. Participants were informed that they would study a series of words for a later memory test. Each study word was presented for 1 sec, followed by a blank 1-sec interval and a prompt to indicate whether the preceding word had been concrete or abstract. The participants registered a concrete or abstract judgment by hitting the "1" or "2" key, respectively. This selection triggered another 1-sec pause, after which the next trial began. A total of 180 trials plus three primacy and three recency buffers composed the study list.

The test phase was very similar to that of Experiment 1, with the following exceptions. First, participants in the feedback condition received one of four messages, depending on whether the preceding response had been a hit, false alarm, correct rejection, or miss (e.g., "Yes, correct, that item WAS on the study list!" for a hit or "No, wrong, that item was NOT on the study list!" for a false alarm). Second, the message was accompanied by a high tone for correct responses and a low tone for incorrect responses. Third, the feedback remained on the screen until participants initiated the next trial. Test instructions in the feedback condition encouraged participants to "observe and consider the feedback for as long as may be useful before continuing." The next trial then began after a 1-sec interval during which only the scale remained on the screen. The control condition was identical to the feedback condition, except that responses were followed only by the instruction to press a key when ready to continue. Fourth, as was noted above, there were twice as many test trials (240) as in Experiment 1.

Finally, to help motivate consistent effort throughout the test phase, participants in both the feedback condition and the control condition were issued a performance-based score after every 60 trials. Scores began at 0; each correct response added 1 point, and each incorrect response subtracted 1 point. Participants were informed of the scoring system via the test instructions. Quarterly score displays provided benchmarks to be surpassed in the succeeding test block and served as rest breaks between blocks.

Results and Discussion

Mean hit and false alarm rates are displayed in Figure 2; results from the .75-old and .25-old conditions appear in the top and bottom panels, respectively. A 2 (study status: old or new) × 4 (test block: 1, 2, 3, or 4) × 2 (feedback: present or absent) × 2 (proportion old: .75 or .25) mixed-factor ANOVA revealed that old items were recognized more often than new items [$F(1,67) = 1{,}243$, $MS_e = 0.042$, $p < .001$, $\eta_p^2 = .949$]. Performance tended to decline across the test, as indexed by a significant study status × block interaction [$F(3,201) = 8.469$, $MS_e = 0.007$, $p < .001$, $\eta_p^2 = .112$]; a marginal study status × block × proportion old interaction reflected the tendency for this decline to be greater in the .25-old condition [$F(3,201) = 2.431$, $MS_e = 0.007$, $p < .07$, $\eta_p^2 = .035$]. A nonsignificant study status × feedback interaction indicated that the difference between hit and false alarm rates did not vary between the feedback and control conditions ($p = .69$). Feedback did not interact with block ($p = .30$), although the study status × block × feedback interaction was significant [$F(3,201) = 2.736$, $MS_e = 0.007$, $p < .05$, $\eta_p^2 = .039$], reflecting the trend for the performance of controls to decline across blocks more than that of the feedback group. Follow-up analyses revealed a tendency for the performance of the control group to worsen to a greater extent than the performance of the feedback group on new items [$F(3,207) = 2.427$, $MS_e = 0.011$, $p < .07$, $\eta_p^2 = .035$], but not on old items ($p = .43$). The fact that false alarm rates rose across blocks in the control condition but not the feedback condition suggests that feedback might have helped the participants avoid false alarms as the test progressed. We are wary of putting stock in such an interpretation, however, because the difference between conditions was small in magnitude and was nonexistent when collapsed across blocks.

As is evident from inspection of Figure 2, feedback did have a marked impact on response bias: The proportion of *old* responses was higher for the participants receiving feedback than for the controls in the .75-old condition and lower for the participants receiving feedback than for the controls in the .25-old condition. There was a near-significant tendency toward a main effect of feedback

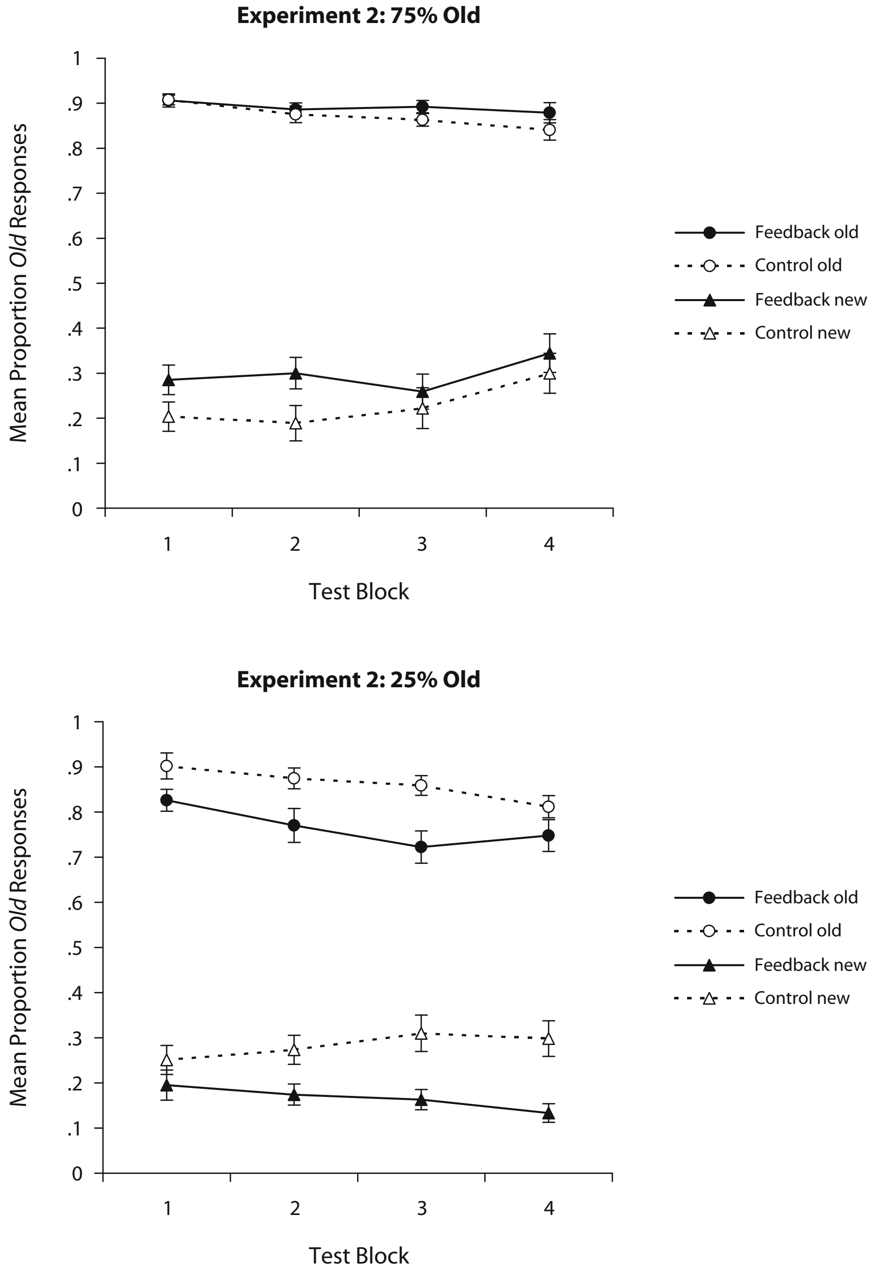

Figure 2. Mean proportion of *old* responses in the .75-old condition (top panel) and the .25-old condition (bottom panel) in Experiment 2. Error bars represent one standard error of the mean.

[$F(1,67) = 3.658$, $MS_e = 0.037$, $p = .06$, $\eta_p^2 = .052$], most likely because the biasing effect of feedback was stronger in the .25-old condition, leaving the feedback participants with lower hit and false alarm rates on balance. The main effect of proportion old was reliable [$F(1,67) = 10.6$, $MS_e = 0.037$, $p < .01$, $\eta_p^2 = .137$], driven largely by greatly reduced *old* responses in the .25-old feedback condition relative to the .75-old feedback condition. A significant feedback × proportion old interaction [$F(1,67) = 21.4$, $MS_e = 0.037$, $p < .001$, $\eta_p^2 = .242$] corroborates this interpretation and captures the effect of feedback on response bias. A separate ANOVA on the .75-old condition indicated that the proportion of *old* responses was significantly higher in the feedback condition [$F(1,34) = 4.250$, $MS_e = 0.033$, $p < .05$, $\eta_p^2 = .111$], despite the fact that effect was weak for old items. A similar ANOVA on the .25-old condition revealed a significant reduction in *old* responses with feedback [$F(1,33) = 18.8$, $MS_e = 0.042$, $p < .001$, $\eta_p^2 = .363$]. The only other significant interaction was block × proportion old [$F(3,201) = 2.795$, $MS_e = 0.010$, $p < .05$, $\eta_p^2 = .040$], indexing the trend for bias to grow more conservative across blocks in the .25-old condition but to increase slightly in the .75-old condition.

Values of A_z were again lower in the feedback condition than in the control condition, but in this experiment the difference was statistically significant [$t(69) = 2.395$, $p < .05$]. Values of c_a were significantly lower for the feedback participants than for the controls in the .75-old condition [$F(1,34) = 4.920$, $MS_{\mathrm{e}} = 0.262$, $p < .05$, $\eta_{\mathrm{p}}^2 = .126$], indicating that feedback increased bias toward giving an *old* response in this condition. In the .25-old condition, values of c_a were higher for the feedback group than for the controls, a significant difference [$F(1,33) = 10.4$, $MS_{\mathrm{e}} = 0.481$, $p < .01$, $\eta_{\mathrm{p}}^2 = .240$]. There were no interactions with confidence level (both *p*s > .20). These measures are consonant with those of the hit and false alarm data in suggesting that feedback exerted a strong influence on response bias but a negative influence, if any, on sensitivity.

The mean RTs in the feedback and control conditions of Experiments 2–4 are displayed in Table 2. All RT data in the article were drawn from nonspeeded responses and were trimmed to reduce the influence of outliers. The longest 0.5% of RTs and those under 200 msec were removed prior to analysis. RTs from error trials were also removed. RTs in each experiment were submitted to an ANOVA containing the same factors used to analyze the hit and false alarm data.

In Experiment 2, participants responded more quickly to old words than to new words [$F(1,67) = 38.1$, $MS_{\mathrm{e}} = 450{,}537$, $p < .001$, $\eta_{\mathrm{p}}^2 = .363$], but this difference was significantly greater in the .75-old condition than in the .25-old condition [$F(1,67) = 21.3$, $MS_{\mathrm{e}} = 450{,}537$, $p < .001$, $\eta_{\mathrm{p}}^2 = .241$]. Response speed increased across test blocks [$F(3,201) = 55.2$, $MS_{\mathrm{e}} = 128{,}569$, $p < .001$, $\eta_{\mathrm{p}}^2 = .452$]; this increase was greater for new items than for old items [$F(3,201) = 9.180$, $MS_{\mathrm{e}} = 78{,}250$, $p < .001$, $\eta_{\mathrm{p}}^2 = .121$]. Participants receiving feedback responded nearly 300 msec faster than controls, a significant difference [$F(1,67) = 4.436$, $MS_{\mathrm{e}} = 2{,}445{,}307$, $p < .05$, $\eta_{\mathrm{p}}^2 = .061$]. This trend showed a nonsignificant tendency to decrease across blocks [$F(3,201) = 2.191$, $MS_{\mathrm{e}} = 128{,}659$, $p = .09$, $\eta_{\mathrm{p}}^2 = .032$] and was significantly greater for new items than for old items [$F(1,67) = 4.344$, $MS_{\mathrm{e}} = 450{,}537$, $p < .05$, $\eta_{\mathrm{p}}^2 = .061$]. Apparently, feedback elicited faster responding from participants, particularly on new items, but reduced recognition sensitivity. The reasons for these trends are unclear, but they discourage characterization of the RT advantage in the feedback condition as an improvement to processing.

Table 2
Mean Reaction Times (in Milliseconds) for the Feedback and Control Groups in Each of the Conditions of Experiments 2, 3, and 4

Condition	Feedback		Control	
	M	*SE*	*M*	*SE*
	Experiment 2			
.75-old	1,621	98	1,859	140
.25-old	1,916	175	2,246	84
	Experiment 3			
Critical #1 lures	1,302	139	1,388	110
Noncritical #1 lures	1,189	162	1,236	137
#2–#5 exemplars	969	75	1,004	63
	Experiment 4			
Category-rule recognition	1,754	108	2,323	134
Simple categorization	2,231	227	2,201	162
Simple recognition	1,932	123	1,879	151

In summary, the results of Experiment 2 demonstrate that feedback can affect response criterion in recognition under uneven base-rate conditions alone. As was noted above, the effect of feedback was weak for old items in the .75-old condition (see Figure 2). One possible reason is that bias was generally liberal, as was evidenced especially by generally liberal values of c_a for controls in both proportion-old conditions; perhaps feedback could not push hit rates much higher than the level exhibited by controls. Nonetheless, the bias shift with feedback was significant in both proportion-old conditions and in each case was in the direction expected if participants used feedback to gain knowledge of underlying base rates and respond accordingly.

These results are consistent with those of Titus (1973) and the digit and letter conditions of Estes and Maddox (1995) but are in contrast to the findings in the word condition of the latter study. Although the reason for this disparity is not obvious, the strengthening of the base-rate manipulation from .67-/.33-old in the Estes and Maddox experiments to .75-/.25-old in Experiment 2 is likely a factor. In addition, as was noted above, d' values were generally very high in the word condition of the Estes and Maddox experiments, raising the possibility that the participants in this condition did not need to attend to or make use of the feedback (i.e., if feedback almost always tells the participants that they are correct, they have no reason to change their response criterion). By demonstrating that feedback can affect bias in recognition of words without manipulations of the stimuli (Rhodes & Jacoby, 2007; Verde & Rotello, 2007) or uneven administration of feedback (Han & Dobbins, 2008, 2009), the present findings join those of recently published studies in delineating the circumstances under which feedback is used by participants to modulate response bias. Crucially, despite this evidence that the participants attended to and used feedback, there was no evidence that doing so enhanced recognition sensitivity.

EXPERIMENT 3

In Experiment 3, we tested for an effect of feedback on recognition sensitivity with stimuli designed to elicit responses on the basis of false familiarity. Experiment 3 was designed such that certain items triggering extremely high levels of familiarity were uniformly new, with the hypothesis that this relationship between subjective experience and objective response would be learned by feedback.

In Experiment 3, we adapted a variant of the Deese/Roediger–McDermott (DRM) false-memory paradigm (Deese, 1959; Read, 1996; Roediger & McDermott, 1995) previously used to elicit illusory recollection of foils categorically associated with studied words (Seamon, Luo, Schlegel, Greene, & Goldenberg, 2000; Smith, Ward, Tindell, Sifonis, & Wilkenfeld, 2000). Participants studied the second through fifth most common (#2–#5) exemplars of various semantic categories. The test list included these studied items plus the #2–#5 exemplars from nonstudied categories. Critically, the test also included the most common (#1) category exemplars from each studied and nonstudied category. This design is well known to produce high false alarm rates to #1 exemplars from studied categories (e.g., Seamon et al., 2000). The motivating hypothesis of Experiment 3 was that corrective feedback at test would reduce such false alarm rates by teaching the participants that the compelling sense of familiarity evoked by #1 exemplars was diagnostic of newness rather than oldness (cf. Unkelbach, 2006).

Past researchers attempting to attenuate false recognition of critical lures in the DRM paradigm have often employed warnings as to the nature of DRM study lists (see Starns, Lane, Alonzo, & Roussel, 2007, and Westerberg & Marsolek, 2006, for reviews). These warnings state explicitly what participants might be expected to learn through feedback in the present design: that the memory test contains nonstudied items that are highly associated with a corresponding group of studied items and that special care should be taken to avoid calling such items "old." Although substantial false alarm rates generally persist following warnings, they are usually significantly lower than those of unwarned controls (e.g., Gallo, Roberts, & Seamon, 1997; McDermott & Roediger, 1998). Although some evidence indicates that the warning must be delivered before encoding of the study lists for a reduction in false alarms to occur (e.g., Anastasi, Rhodes, & Burns, 2000; Neuschatz, Payne, Lampinen, & Toglia, 2001), other evidence suggests that a warning delivered between study and test can have a beneficial effect as well (e.g., Lane, Roussel, Starns, Villa, & Alonzo, 2008; McCabe & Smith, 2002; Starns et al., 2007). If, as the latter finding suggests, participants benefit from insight into the test design without having such knowledge at study, the use of this information, and the consequent decrease in critical false alarms, must be achievable via processes operating at test. Although feedback does not reveal the nature of the test design as explicitly as a warning does, it should yield a similar insight by repeatedly discouraging *old* responses to #1 lures and should thereby reduce such false alarms.

To our knowledge, published investigations into the effects of feedback in a DRM paradigm are limited to two recent studies. McConnell and Hunt (2007) tested free recall of DRM lists and gave feedback by rereading the lists as participants corrected their test forms; a second test on the same lists was given 2 days later. The feedback group repeated critical errors from the first test at a considerably lower rate than controls (surprisingly, however, this advantage was partially offset by a tendency for feedback participants to commit more *new* critical errors than controls during the second test). These results highlight the potential of feedback to draw attention to false recollection of nonpresented list prototypes (at least in a free recall task), but because McConnell and Hunt tested the same study set twice, the feedback advantage was most likely driven by the learning of correct responses to particular stimuli, rather than an enhanced ability to avoid false recollection in general.

Jou and Foreman (2007) obtained evidence of just such enhancement in both recall and recognition of DRM lists. In the recognition version of their design (Jou & Foreman, 2007, Experiment 1), participants received a standard warning as to the nature of the lists and the need to avoid false recognition of critical nonpresented items, and, critically, some were given negative feedback whenever they called such an item "old." Although even participants in a warning-only condition showed marked improvement in rejecting critical lures over the course of the experiment (relative to an unwarned control group), feedback yielded a significant additional benefit. Because no lists were repeated across study–test cycles, Jou and Foreman were able to infer that feedback-based training had increased a general sensitivity to critical lures.

Jou and Foreman's (2007) findings do not necessarily indicate that feedback will successfully lower false alarm rates to #1 lures in the present experiment. First, these researchers used a repeated study–test procedure. Given that the feedback conditions of their experiments included a warning, the benefit of feedback that they observed may have been driven in part by its contribution to processing occurring at study (e.g., generation and tagging of likely critical items during study to facilitate rejection at test; see Gallo, Roediger, & McDermott, 2001). Whether participants can use feedback to learn to reject critical lures when the appearance of such lures is not explicitly foretold and learning from feedback can only accrue at test is an open question.

We predicted that in the absence of an explicit warning and without repeated study–test cycles, feedback would enable participants to learn that the strong feelings of familiarity evoked by #1 lures should not be taken as evidence of oldness. Although feedback was given on every trial, we expected negative feedback following false alarms to critical lures to be particularly salient: In the present design, every single time a participant endorsed a #1 exemplar (e.g., *bird*: *robin*; *metal*: *gold*), he or she was told, "No, that word was not on the study list." We expected that this feedback would lead the participants to become more cautious about endorsing highly category-prototypical exemplars, leading to the use of recollection to verify their study list presence and reducing the false-memory effect. Such a finding would demonstrate one respect in which feedback can be used to improve recognition accuracy at test.

Method

Participants. There were 43 participants in Experiment 3. Twenty-one were in the feedback condition, and 22 were in the control condition.

Materials. The stimuli were taken from 51 Battig and Montague (1969) categories, with some adjustments made after norming in our lab to ensure the appropriateness of their use with a Canadian undergraduate population.[3] The study set consisted of the #2–5 (noncritical) exemplars from 40 of these categories. The study categories were randomly selected for each participant. The test list consisted of 97 randomly selected members of the study set, the 44 #2–#5 exemplars from the 11 nonstudied categories, the (noncritical) #1 exemplar from each of the nonstudied categories, and the (critical) #1 exemplar from each of the 40 studied categories, randomly intermixed. In total, the test list contained 97 old items and 95 new items. The experiment was conducted using E-Prime software (Psychology Software Tools, www.pstnet.com).

Procedure. Participants were tested individually and were randomly assigned to the feedback or control conditions. The study items and their corresponding category labels (e.g., "[unit of distance] inch") appeared one at a time in the center of the computer screen. Participants were instructed to read each word aloud and to think about its meaning with respect to the category label. The presentation of items occurred in 40 blocks according to category membership: Each block contained the four members of a given category, presented in descending order of exemplar frequency, followed by a new block containing the four members of the next category. Block presentation order was randomized for each participant. Each item and category label was presented for 1.5 sec, and a 1-sec blank interval separated the items.

During the test phase, each word was presented without its category label. A mandatory interval was imposed whereby no response could be given for 2 sec following stimulus onset. This delay was intended to encourage the participants to take the time to engage in controlled processing of each probe rather than summarily endorsing each item (including critical lures) that passed a criterial level of familiarity. Participants were advised that they need not respond immediately after the 2 sec had passed but that they were required to wait at least that amount of time before their response could be entered. The instructions urged the participants to use this delay to "give careful thought to whether or not the word in front of you was part of the study list." A tone sounded when 2 sec had elapsed. The test probes remained on the screen until a response was given.

Participants responded using the same scale as in Experiments 1 and 2. The feedback was identical to that of Experiment 2, with the exceptions that the message was changed to include the item from the current test trial (e.g., "Yes, [item] WAS on the study list!") and that the sounds that accompanied the message were changed to a voice saying "excellent" for correct responses and a buzzer for incorrect responses. Quarterly scores were no longer provided in either the feedback or the control condition. Otherwise, the control condition procedure did not differ from that of Experiment 2. A total of 192 trials composed the test list.

Results and Discussion

Although the comparison of primary interest in Experiment 3 is that of feedback versus control group false alarm rates to critical (#1) lures, we first report the results from the noncritical #2–#5 exemplars. Both sets of data are displayed by test block in Figure 3. Consistent with Experiments 1 and 2, a 2 (study status) × 4 (block) × 2 (feedback) ANOVA revealed a significant main effect of study status, with old words recognized more often than new words [$F(1,41) = 889$, $MS_e = 0.045$, $p < .001$, $\eta_p^2 = .956$] but no study status × feedback interaction ($p = .60$), indicating that feedback failed to increase old/new discrimination. The main effect of feedback was also nonsignificant ($p = .22$), suggesting that feedback had no effect on response bias. Consistent with these outcomes, neither A_z nor c_a differed significantly between the feedback and control groups (ps = .27 and .36, respectively). Analysis of c_a estimates revealed no interaction of feedback and confidence level ($p = .53$). The main effect of test block was significant [$F(3,123) = 2.827$, $MS_e = 0.009$, $p < .05$, $\eta_p^2 = .065$], reflecting a tendency for *old* responses to rise slightly from Block 1 to Block 3. Block did not interact with feedback ($p = .32$).

As in Experiment 2, an ANOVA on RTs to the #2–#5 exemplars showed that old words elicited faster responses than new words [$F(1,41) = 17.9$, $MS_e = 80,754$, $p < .001$, $\eta_p^2 = .304$] and that response speed increased across blocks [$F(3,123) = 17.2$, $MS_e = 106,537$, $p < .001$, $\eta_p^2 = .295$]. Unlike in Experiment 2, however, there was no main effect of feedback on RTs ($p = .72$). Feedback did interact with block [$F(3,123) = 4.435$, $MS_e = 106,537$, $p < .01$, $\eta_p^2 = .098$]; although both the feedback group and the control group increased response speed across blocks, the controls did so to a greater extent, reversing an initial speed disadvantage during the second half of the test. This interaction is consistent with the possibility that feedback encouraged the participants to maintain a greater amount of care in their recognition judgments as the test progressed (if this was the case, however, such care did not result in increased sensitivity). Note that the mandatory response delay imposed on the participants may have obscured potential differences between the feedback and control groups on this measure; thus, the RT results from the present experiment should be interpreted with caution.

An ANOVA on the #1 lures indicated that those from studied categories received a far higher proportion of false alarms than those from nonstudied categories, replicating the classic DRM false-memory effect [$F(1,41) = 43.2$, $MS_e = 0.047$, $p < .001$, $\eta_p^2 = .513$]. Contrary to our prediction, however, the magnitude of the DRM effect did not interact with feedback ($p = .58$). The rates of false alarms to critical lures from studied categories did not differ between the feedback and control groups (Ms = .30 and .31, respectively; $p = .86$). Although mean false alarm rates to #1 lures from nonstudied categories appear from Figure 3 to be substantially reduced in the third and fourth blocks for the participants receiving feedback, this difference is largely an artifact of high variability due to the small number of such items in each block. There were no significant differences between the feedback and control conditions on noncritical #1 lures, and even the directional differences were greatly reduced when recognition ratings—not false alarms—were used as the dependent measure. Since all of these critical items were new on the test list, there was no basis for calculation of sensitivity or bias.

Participants responded more slowly to critical than to noncritical #1 lures [$F(1,41) = 4.127$, $MS_e = 481,015$, $p < .05$, $\eta_p^2 = .091$]. Response speed to these items generally increased significantly across blocks [$F(2.174, 89.151) = 5.585$, $MS_e = 540,421$, $p < .01$, $\eta_p^2 = .120$, Greenhouse–Geisser corrected], with critical #1 lures eliciting a marginally greater increase in speed than noncritical #1 lures [$F(2.323, 95.235) = 2.621$, $MS_e = 279,282$, $p = .07$, $\eta_p^2 = .060$, Greenhouse–Geisser corrected]. Feedback did not cause the participants to take more time with judgments than the controls ($p = .75$). This null effect held for both critical and noncritical lures when analyzed separately

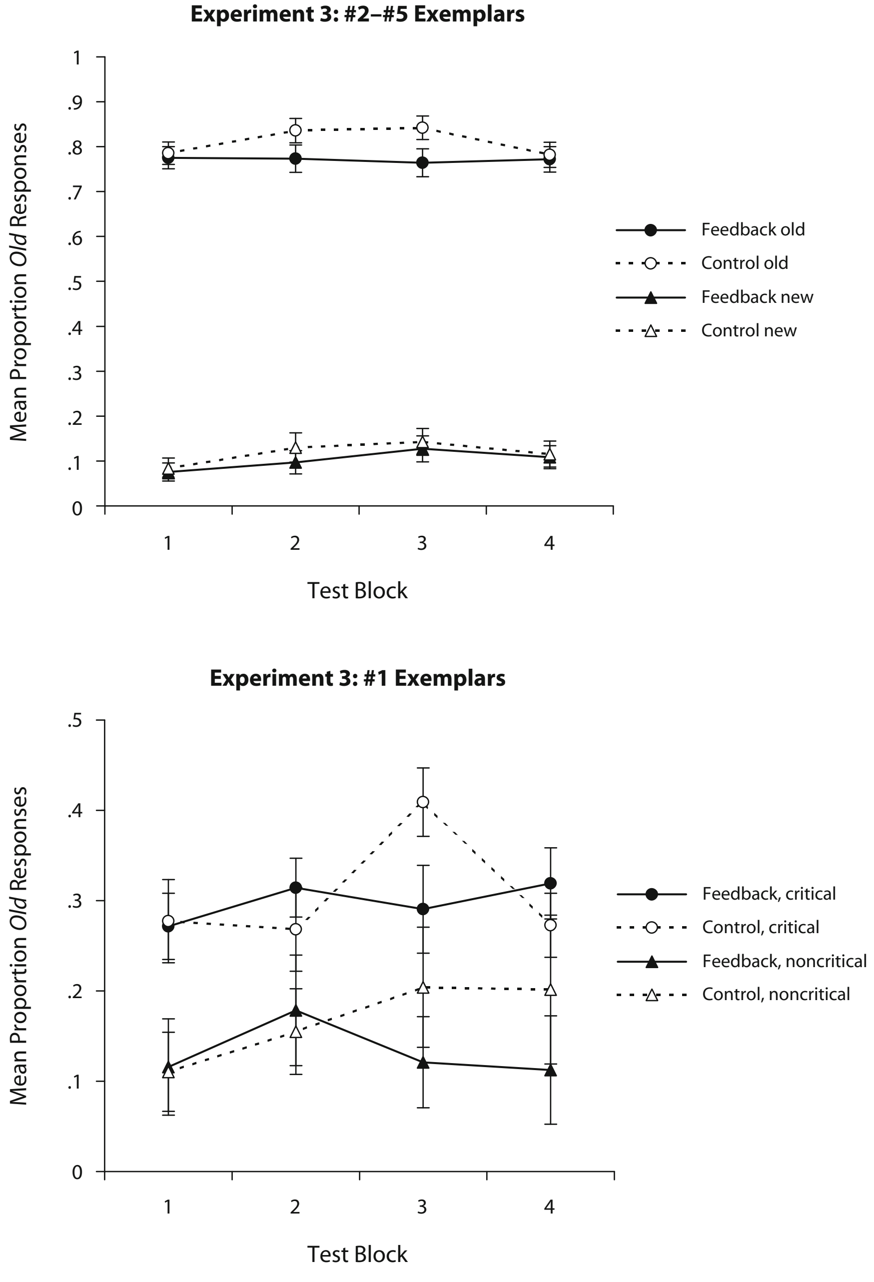

Figure 3. Mean proportion of *old* responses for #2–#5 exemplars (top panel) and #1 lures (bottom panel) from studied and nonstudied categories in Experiment 3. Error bars represent one standard error of the mean.

(both *p*s $>$ .62), but see the above caution regarding the interpretation of the RT results in the present experiment.

These results provide no indication that feedback helped participants to realize the misleading sense of familiarity engendered by critical lures. This outcome is particularly striking in light of the fact that participants were made to take at least 2 sec to consider each test probe carefully before committing to a judgment. Apparently, time and instructional encouragement to engage in controlled processing was not sufficient for feedback to improve accuracy, even on items containing such a salient and diagnostic subjective feature as the familiarity of the critical lures.[4]

EXPERIMENT 4

Experiment 4 was designed to provide the most favorable conditions yet for a feedback advantage in recognition. Although subjective responses to stimuli were to serve as the basis for improvement through feedback in Experiment 3, in Experiment 4 we created a scenario in which an objective, readily observable stimulus feature was available to drive learning. In the category-rule recognition (CRR) condition, participants could attain near-perfect performance by learning a simple categorical rule: Old items were names of large objects (e.g., *tree*, *barn*), and new items were names of small objects (e.g., *thimble*, *cup*) or, for other participants, vice versa.

In categorization tasks, trial-by-trial feedback is an essential component of learning experimenter-defined category boundaries, and research shows that such learning occurs relatively quickly with simple rules (e.g., Bruner, Goodnow, & Austin, 1956). In the present recognition design, feedback provides the same information as in category learning procedures and should result in the same benefit to performance: Because the size of the referent object is 100% diagnostic of *old–new* status, the participants may essentially treat the CRR condition as a categorization task in which large words are classified as old and small words are classified as new or vice versa.

In the simple categorization (SC) condition, the CRR task was transformed into a literal categorization task. The same test words as in the CRR condition were presented without a preceding study phase, and participants were instructed to assign items to either Category A or Category B. As in the CRR condition, category membership was determined by the size of the referent object and was to be learned through feedback.

Finally, in the simple recognition (SR) condition, recognition of the same word stimuli as in the CRR and SC conditions was tested but old/new status was not confounded with the size of the referent object (i.e., no simple category rule separated old from new items). Thus, the SR condition represents a replication of previous tests of feedback effects on recognition, as well as a comparison for the CRR condition.

Method

Participants. There were 46 participants in the CRR condition ($n = 22$ in the feedback condition, 24 in the control condition), 44 participants in the SC condition (21 in the feedback condition, 22 in the control condition), and 31 participants in the SR condition (16 in the feedback condition, 15 in the control condition). The data of 4 participants performing at or near chance were removed from subsequent analyses.

Materials. The stimuli were 160 concrete nouns selected from the MRC database; half were names of objects larger than a standard bread box, and half were names of objects smaller than a bread box (according to the intuition of the second author). In the CRR condition, the study list consisted of five primacy buffers, 75 randomly ordered target items, and five recency buffers. For a randomly selected half of the participants in the CRR condition, all of the study items were names of large objects; for the remaining participants, all of the study items were names of small objects. The test list consisted of the 75 studied target items randomly intermixed with the 75 nonstudied items. In the SC condition, all 150 items were presented in a random order with no preceding study phase. In the SR condition, the 150 items were randomly divided into targets and lures for each participant. The experiment was conducted with E-Prime software.

Procedure. Participants were randomly assigned to the task and feedback conditions. In the CRR and SR conditions, each study item appeared in the center of the screen for 500 msec. Participants were instructed to consider each word as a concrete noun in preparation for a memory test. The study items were separated by a blank 500-msec interval. The study and test phases were separated by a 10-min delay interval during which participants wrote down the names of as many popular songs as they could. The test procedure in the CRR and SR conditions was identical to that of Experiment 2.

The SC task consisted of 150 test words with no preceding study phase. The correct category (A or B for large or small object names, respectively) was selected randomly. Participants were instructed to sort the items into Categories A and B as they appeared on the screen and were told that although they would essentially be guessing at the outset of the experiment, they might eventually develop a sense of what distinguishes a member of Category A from a member of Category B. The response scale ranged from 1 (*definitely A*) to 6 (*definitely B*). Participants in the feedback condition were encouraged to use the feedback to inform their decisions. Feedback appeared as either "A" or "B." In all other respects, the test procedure was identical to that of the CRR task.

The CRR and SC conditions included posttest questions as to whether participants had noticed an attribute that separated old (or Category A) items from new (or Category B) items (i.e., the size of the referent object). After participants entered their responses, they were told that such an attribute did exist and were asked to try to name it.

Results and Discussion

The results of Experiment 4 are displayed in Figure 4. Since this experiment contained a smaller number of trials than Experiments 2 and 3, the data were broken down into three test blocks instead of four.

CRR condition. A 2 (study status) × 4 (block) × 2 (feedback) ANOVA confirmed that old items were recognized more often than new items [$F(1,43) = 411$, $MS_e = 0.033$, $p < .001$, $\eta_p^2 = .905$]. The main effect of block was also significant [$F(2,86) = 3.239$, $MS_e = 0.010$, $p < .05$, $\eta_p^2 = .070$], reflecting a general tendency for both hits and false alarms to decrease across blocks. The critical study status × feedback interaction indexing the effect of feedback on recognition sensitivity was again nonsignificant [$F(1,43) = 2.553$, $MS_e = 0.033$, $p = .12$, $\eta_p^2 = .056$]; however, a directional benefit of feedback was obtained. Follow-up analyses indicated that hit rates were significantly higher for participants receiving feedback ($M = .76$) than for controls ($M = .68$) [$F(1,43) = 4.221$, $MS_e = 0.046$, $p < .05$, $\eta_p^2 = .089$]. However, false alarm rates were nearly identical in the feedback ($M = .28$) and control ($M = .27$) conditions ($p = .90$). A_z did not differ across the two conditions ($p = .30$). The main effect of feedback was nonsignificant ($p = .22$), as was the effect of feedback on c_a ($p = .36$), indicating that response bias was unaffected by feedback. Feedback did not interact with confidence level in the estimates of c_a ($p = .55$).

The significant hit rate advantage in the feedback condition represents the first benefit of feedback to recognition observed in this line of studies. This result is consistent with the possibility that when a categorical rule separates old and new items, participants can use feedback to become sensitive to the rule and improve recognition performance. Unfortunately, multiple lines of evidence call such a conclusion into question. First, the overall improvement in sensitivity was nonsignificant, whether measured by the study status × feedback interaction or the A_z parameter. Second, the difference in hit rates between the feedback and control conditions remained virtually unchanged across test blocks, arguing against the idea that feedback allowed for (presumably gradual) learning of the category rule. Third, we conducted a replication of the CRR condition in a follow-up experiment using 37 par-

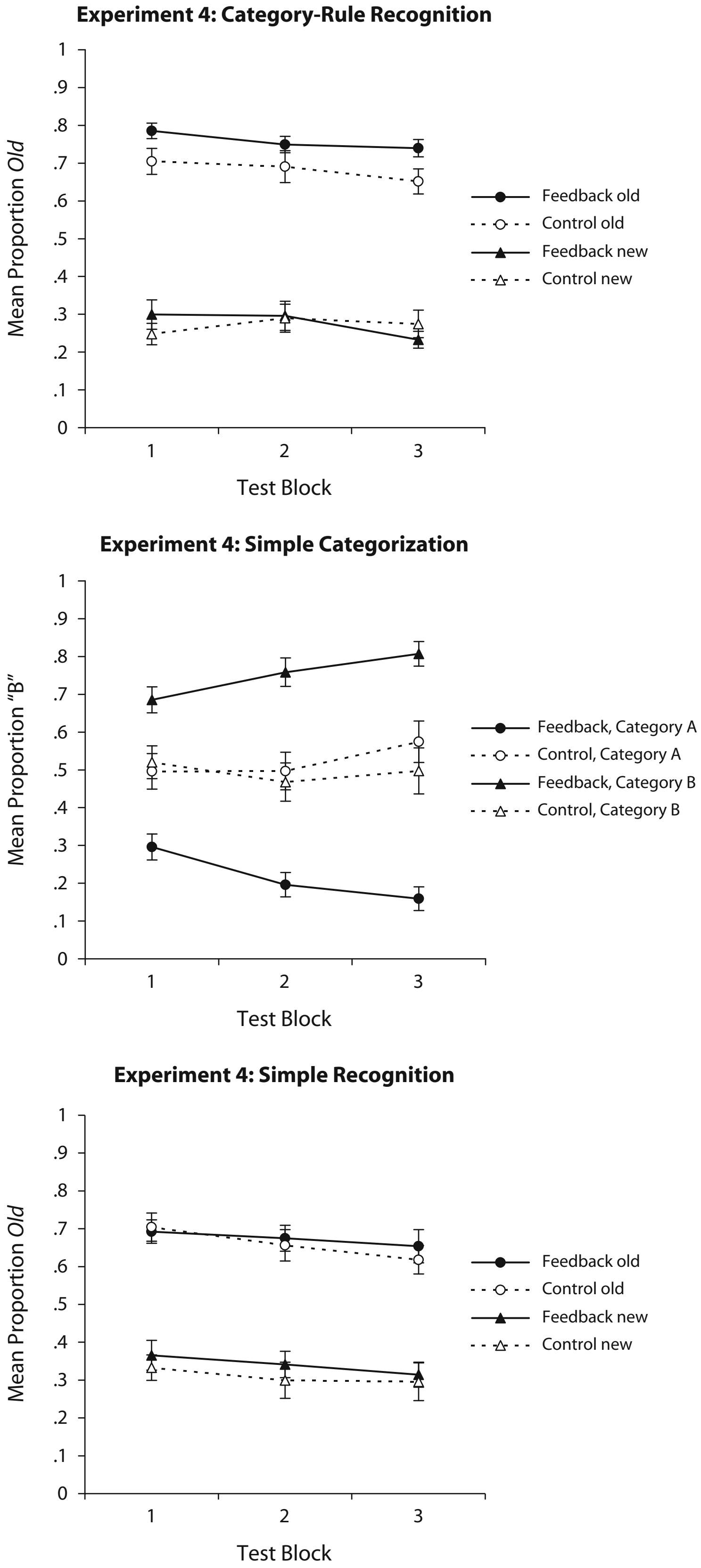

Figure 4. Mean proportion of *old* responses in the category-rule recognition condition (top panel); mean proportion of Category B responses (4, 5, or 6 on the response scale) in the simple categorization condition (middle panel), and mean proportion of *old* responses in the simple recognition condition (bottom panel) in Experiment 4. Error bars represent one standard error of the mean.

ticipants (17 in the feedback condition, 20 in the control condition) and failed to obtain any advantage of feedback: Hit rates were .746 and .749 in the feedback and control conditions, respectively, and false alarm rates were .256 and .279. Thus, the present evidence regarding increased category-rule recognition performance with feedback is equivocal at best.

The trends in the RT data generally followed those observed in Experiments 2 and 3: Responses were faster to old words than to new words [$F(1,43) = 21.9$, $MS_e = 375{,}695$, $p < .001$, $\eta_p^2 = .338$], response speed increased across test blocks [$F(2,86) = 20.1$, $MS_e = 140{,}109$, $p < .001$, $\eta_p^2 = .319$], and the increase was greater for new words than for old words [$F(2,86) = 3.883$, $MS_e = 79{,}136$, $p < .05$, $\eta_p^2 = .083$]. Strikingly, however, the participants receiving feedback responded nearly 600 msec faster than the controls, a highly significant difference [$F(1,43) = 15.0$, $MS_e = 1{,}702{,}578$, $p < .001$, $\eta_p^2 = .258$]. Despite the fact that feedback produced a nonsignificant benefit (if any) to recognition sensitivity, it apparently enabled a sizable increase in decision speed. We return to this result below.

SC condition. When the task was to categorize the items, participants receiving feedback greatly outperformed controls. A 2 (category: A or B) × 4 (block) × 2 (feedback) ANOVA revealed a significant category × feedback interaction [$F(1,40) = 37.0$, $MS_e = 0.134$, $p < .001$, $\eta_p^2 = .481$], indicating that the accurate discrimination of Category A and Category B items occurred with much higher frequency in the feedback condition. A marginal category × block interaction [$F(2,80) = 2.715$, $MS_e = 0.013$, $p < .10$, $\eta_p^2 = .064$] capturing increased accuracy across blocks was driven by learning in the feedback condition, as was evidenced by a significant category × block × feedback interaction [$F(2,80) = 13.226$, $MS_e = 0.013$, $p < .001$, $\eta_p^2 = .248$].

RT analyses[5] revealed only a main effect of block [$F(2,78) = 76.3$, $MS_e = 383{,}031$, $p < .001$, $\eta_p^2 = .662$], signifying decreasing RTs as the test progressed, and a category × feedback interaction [$F(1,39) = 4.139$, $MS_e = 222{,}711$, $p < .05$, $\eta_p^2 = .096$]. The latter effect corresponds to the tendency for the control participants to respond faster to Category B items, whereas the feedback participants were mildly faster to respond to Category A items. An interpretation for this pattern is not immediately obvious, but it does not appear to bear meaningfully on the issues of central interest here. The feedback and control conditions did not differ in RTs ($p = .91$).

SR condition. When recognition decisions could not be guided by a categorical rule, feedback was ineffectual with respect to both sensitivity and bias. A 2 (study status) × 4 (block) × 2 (feedback) ANOVA yielded the expected results: significantly higher recognition of old items than new items [$F(1,27) = 205$, $MS_e = 0.025$, $p < .001$, $\eta_p^2 = .884$], but no modulation of this effect by feedback ($p = .73$). As in the CRR condition, a significant main effect of block captured the tendency for both hits and false alarms to decrease across the course of the test [$F(2,57) = 4.216$, $MS_e = 0.010$, $p < .05$, $\eta_p^2 = .135$], but this trend did not interact with feedback ($p = .85$), and there was no main effect of feedback ($p = .58$). Accordingly, neither A_z nor c_a differed between the feedback and control conditions (both ps $> .53$). Again, there was no evidence of an interaction of feedback and confidence level in c_a analyses ($p = .58$). These results are not surprising, since the SR condition contained no manipulation other than feedback, which has repeatedly produced null effects.

Analysis of RTs followed the established patterns: faster responding to old words than to new words [$F(1,27) = 11.2$, $MS_e = 165{,}976$, $p < .01$, $\eta_p^2 = .293$], increased response speed across test blocks [$F(2,54) = 16.6$, $MS_e = 100{,}958$, $p < .001$, $\eta_p^2 = .381$], and a tendency toward a sharper decrease in RTs to new words than to old words [$F(2,54) = 2.562$, $MS_e = 76{,}718$, $p = .09$, $\eta_p^2 = .087$]. A marginal study status × feedback interaction indexed the trend for RTs in the feedback condition to be identical to those of the control condition for new words but roughly 230 msec slower for old words [$F(1,27) = 3.603$, $MS_e = 165{,}976$, $p = .07$, $\eta_p^2 = .118$]. Overall, feedback participants responded approximately 100 msec slower than did controls, but this difference was not significant ($p = .57$).

Responses to the posttest questions reflected the inability of the participants in the CRR condition to discover the rule separating old and new items. Asked whether they had noticed a difference between items that had been on the study list and those that had not, the percentage of participants answering "yes" was similar in the feedback (50%) and control (53%) conditions. In the SC condition, more feedback participants (95%) than controls (77%) responded "yes," despite the fact that all of the participants in this condition were told via study instructions that such a difference did exist. More strikingly, only 1 participant (5%) in the CRR feedback condition and none in the control condition correctly identified the category-relevant dimension. Including the participants proposing a category rule of places versus objects (an imperfect but diagnostic decision rule) improved these totals only to 2 feedback participants (10%) and 1 control (4%). By contrast, 45% of the SC feedback participants identified the exact category rule, and 90% provided either the exact rule or "places versus objects." In the SC control condition, these totals were 9% and 41%, respectively.

Although feedback exerted no detectable effect on sensitivity in the CRR condition, the results of the RT analyses indicated a compelling advantage for the feedback participants. By contrast, responding in the SR condition was nonsignificantly slower for the feedback participants. These findings suggest that the presence of a category rule allowed the feedback group to speed responses substantially. The connection between feedback and the category rule remains unclear, however: This dramatic increase in recognition speed was not accompanied by a significant increase in accuracy, or conscious identification of the category rule. Moreover, the RT advantage did not increase across blocks (i.e., the feedback × block interaction did not approach significance), as would be expected if it was associated with some implicit form of learning of the category rule.

We do not interpret the results of Experiment 4 as suggesting that there are no circumstances under which feedback could enable participants to learn a category rule in a CRR-like condition. For example, had the rule been an obvious physical distinction between studied and nonstudied words (e.g., length), feedback may well have proven extremely beneficial. By demonstrating that feedback was largely ineffective in exposing a category rule for recognition judgments that was readily learned for categorization judgments, however, the present results suggest substantial resistance of the recognition system to improvements in sensitivity via feedback.

GENERAL DISCUSSION

Our purpose in the present experiments was to examine the effects of trial-by-trial corrective feedback on recognition memory. Although a small extant literature provides scant evidence that feedback improves recognition sensitivity, the present experiments provided a stronger test of its potential efficacy by testing conditions tailored to promote such improvement. The design of these experiments isolated the potential influence of feedback to the test phase, ensuring that any effects observed reflected a difference between participants who did versus did not receive feedback in recognition processing, not in study processing. Our results yielded two primary conclusions. First, feedback can guide appropriate bias shifts when the base rates of old and new items differ sharply from 50–50 (Experiment 2). This result contributes to a recent literature examining the potential for feedback to affect criterion setting in recognition. Although recent studies have established that feedback can lead to bias shifts when item attributes such as trace strength (Verde & Rotello, 2007) and screen location (Rhodes & Jacoby, 2007) are manipulated or when feedback is altered or withheld in a biasing fashion (Han & Dobbins, 2008, 2009), the results of Experiment 2 indicate that neither of these manipulations are necessary; base-rate unevenness alone produced substantial criterion shifts, in contrast to a previously reported finding (Estes & Maddox, 1995).

Second, feedback failed to enhance sensitivity across four experiments. Accuracy was no higher with feedback than without on a test of categorized lists studied with deep or shallow orienting tasks (Experiment 1), in a false-memory paradigm in which feedback repeatedly discouraged endorsing highly familiar critical lures as old (Experiment 3), and in a categorization-like task in which old and new items were separable along a single discernible dimension (Experiment 4, CRR condition). Taken together, these results suggest that the processes of recognition are impenetrable to improvement by feedback. Indeed, there was a small but consistent tendency for A_z to be lower in the feedback condition than in the control condition (a difference that attained statistical significance in Experiment 2). It might be that efforts to use feedback backfired, for example, by encouraging the use of more analytic and less sensitive bases for recognition judgments (see Whittlesea & Price, 2001). Whether a meaningful phenomenon underlies these small and mostly nonsignificant negative effects, they unanimously indicate no hint of a tendency for feedback to increase sensitivity.

With respect to the effect of feedback on recognition RTs, our results are inconclusive. Although feedback was associated with significant increases in response speed in Experiment 2 and the CRR condition of Experiment 4, no differences approaching significance emerged in Experiment 3 or the SR condition of Experiment 4. Furthermore, the two significant RT effects neither grew across blocks nor were accompanied by improvements in accuracy, leaving the locus of these effects unclear. One possibility is that participants may, under some conditions, respond more quickly and decisively when given feedback because of increased interest in the task; indeed, many participants mentioned during debriefing sessions that feedback motivated them to perform as well as they could.

Are the null effects of feedback on recognition accuracy merely Type II errors due to weak manipulations or insufficient power? Estimates of power to detect a medium-sized effect (Cohen's $d = 0.5$) of feedback on recognition sensitivity, calculated using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), averaged .40 across the four experiments. Although power was modest in the individual experiments, several lines of evidence argue against the concern that the present findings are a product of low power. First, the null effects observed do not reflect small tendencies toward feedback effects that failed to reach significance; indeed, in only one instance across four experiments was feedback even directionally advantageous (Experiment 4, CRR condition, old items). Second, feedback did produce substantial criterion shifts in Experiment 2, indicating that it was not misunderstood or ignored by the participants and that there was a sufficient number of observations to detect effects. Third, in an effort to determine whether small modifications in experimental design might create conditions conducive to a feedback advantage, we have conducted at least one variant of each of the experiments reported in this article. In no case was a feedback advantage observed.

We have conducted a number of other studies revealing null effects of feedback on recognition sensitivity with word stimuli. In one study, for example, we sought to determine whether participants attended to and used feedback by repeating some of the test items at lags of four to seven items. Control participants were no more accurate on the second presentation of a repeated item than on the first, whereas feedback participants gave higher recognition ratings to the second presentation of old items and lower ratings to the second presentation of new items, showing that they did indeed learn from the feedback. Nonetheless, the feedback participants were no different from the controls in their responses to once-presented test items, indicating that the feedback did not effect any general increase in old/new discrimination. In another approach (following a suggestion by Asher Koriat), we tested recognition of high- and low-frequency words. Since old/new discrimination of high-frequency words is inferior to that of low-frequency words (i.e., the mirror effect; Glanzer & Adams, 1985), we speculated that feedback could alert participants to the difficulty of recognizing such words, encouraging a change in the processing of these items at test that would support higher accuracy. Neither high- nor low-frequency words were recognized more accurately with feedback, however.

Thus, although there is ample evidence that deliberative, recall-like retrieval processes and sophisticated inferential processes play an important role in at least some recognition memory tasks (e.g., Jacoby, Woloshyn, & Kelley, 1989; Lindsay, 2008), the present results suggest no mechanism by which the information provided in feedback can be harnessed to improve recognition sensitivity. These results suggest an interpretation of previous findings indicating positive effects of feedback on recognition sensitivity (e.g., the digit and letter conditions of Estes & Maddox, 1995; Jennings & Jacoby, 2003; Jou & Foreman, 2007). Given the difficulty of raising sensitivity with feedback at test alone, these past studies likely produced a feedback advantage through changes in study processing. We propose that the format of these studies (multiple study–test cycles or continuous recognition) enabled their participants to translate test feedback into more effective strategies for encoding subsequent study items. This idea bears implications for recognition-based memory training programs such as that of Jennings and Jacoby (2003): Feedback is not likely to enhance performance unless the participants have the opportunity to benefit from feedback during a subsequent study phase.

We close by acknowledging that we have by no means eliminated the possibility that feedback can, under some conditions, improve recognition accuracy. It might be that with more trials, more intense varieties of feedback, or different stimuli, a benefit of feedback on *old–new* discrimination is observable. Our experience suggests, however, that the conditions necessary for feedback to increase recognition accuracy are difficult to find.

AUTHOR NOTE

This research was supported by an Erich and Shelley Mohr Fellowship to J.K. and a Natural Sciences and Engineering Research Council of Canada Discovery grant to D.S.L. We thank Jeffrey Sun, Ben Shiner, Danika Overmars, Leanne Davies, Jill Owen, Claire McEwen, Amanda Ferguson, Julia Kiefer, and Jessica Beveridge for their assistance with data collection. Correspondence concerning this article should be addressed to J. Kantner or D. S. Lindsay, Department of Psychology, University of Victoria, P.O. Box 3050 STN CSC, Victoria, BC, V8W 3P5 Canada (e-mail: jkantner@uvic.ca or slindsay@uvic.ca).

Note—Accepted by the previous editorial team, when Brian H. Ross was Editor.

REFERENCES

Anastasi, J. S., Rhodes, M. G., & Burns, M. C. (2000). Distinguishing between memory illusions and actual memories using phenomenological measurements and explicit warnings. *American Journal of Psychology*, **113**, 1-26. doi:10.2307/1423458

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, **80**, 1-46. doi:10.1037/h0027577

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Clark, S. E. (1999). Recalling to recognize and recognizing recall. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson–Shiffrin model* (pp. 215-243). Mahwah, NJ: Erlbaum.

Clark, W. C., & Greenberg, D. B. (1971). Effects of stress, knowledge of results, and proactive inhibition on verbal recognition memory (d') and response criterion (LX). *Journal of Personality & Social Psychology*, **17**, 42-47. doi:10.1037/h0030422

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-505.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, **11**, 671-684. doi:10.1016/S0022-5371(72)80001-X

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, **58**, 17-22. doi:10.1037/h0046671

Dodson, C. S., & Johnson, M. K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology*, **106**, 541-557. doi:10.2307/1422968

Eng, J. (n.d.). *ROC analysis: Web-based calculator for ROC curves*. Retrieved March 30, 2009, from www.jrocfit.org.

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1075-1095. doi:10.1037/0278-7393.21.5.1075

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, **39**, 175-191.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, **35**, 116-124.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, **4**, 271-276.

Gallo, D. A., Roediger, H. L., III, & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, **8**, 579-586.

Gardner, R. M., Sandoval, Y., & Reyes, B. (1986). Signal-detection analysis of recognition memory of obese subjects. *Perceptual & Motor Skills*, **63**, 227-234.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, **13**, 8-20.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 259-278. doi:10.1037/0278-7393.23.2.259

Han, S., & Dobbins, I. G. (2008). Examining recognition criterion rigidity during testing using a biased-feedback technique: Evidence for adaptive criterion learning. *Memory & Cognition*, **36**, 703-715. doi:10.3758/MC.36.4.703

Han, S., & Dobbins, I. G. (2009). Regulating recognition decisions through incremental reinforcement learning. *Psychonomic Bulletin & Review*, **16**, 469-474. doi:10.3758/PBR.16.3.469

Harvey, L. O., Jr. (2005). Parameter estimation of signal detection models: RscorePlus (Version 5.6.0) [Computer software and manual]. Retrieved November 13, 2009, from http://psych.colorado.edu/~lharvey/html/software.html.

Healy, A. F., & Kubovy, M. (1978). The effects of payoffs and prior probabilities on indices of performance and cutoff location in recognition memory. *Memory & Cognition*, **6**, 544-553.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541. doi:10.1016/0749-596X(91)90025-F

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, **118**, 115-125. doi:10.1037/0096-3445.118.2.115

JACOBY, L. L., YONELINAS, A. P., & JENNINGS, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13-47). Mahwah, NJ: Erlbaum.

JENNINGS, J. M., & JACOBY, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology & Aging*, **8**, 283-293. doi:10.1037/0882-7974.8.2.283

JENNINGS, J. M., & JACOBY, L. L. (2003). Improving memory in older adults: Training recollection. *Neuropsychological Rehabilitation*, **13**, 417-440. doi:10.1080/09602010244000390

JOHNSON, M. K., HASHTROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28. doi:10.1037/0033-2909.114.1.3

JOU, J., & FOREMAN, J. (2007). Transfer of learning in avoiding false memory: The roles of warning, immediate feedback, and incentive. *Quarterly Journal of Experimental Psychology*, **60**, 877-896.

KLUGER, A. N., & DENISI, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, **119**, 254-284. doi:10.1037/0033-2909.119.2.254

KRUSCHKE, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 3-26. doi:10.1037/0278-7393.22.1.3

LANE, S. M., ROUSSEL, C. C., STARNS, J. J., VILLA, D., & ALONZO, J. D. (2008). Providing information about diagnostic features at retrieval reduces false recognition. *Memory*, **16**, 836-851. doi:10.1080/09658210802337734

LANE, S. M., ROUSSEL, C. C., VILLA, D., & MORITA, S. K. (2007). Features and feedback: Enhancing metamnemonic knowledge at retrieval reduces source monitoring errors. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **33**, 1131-1142. doi:10.1037/0278-7393.33.6.1131

LINDSAY, D. S. (2008). Source monitoring. In J. H. Byrne (Series Ed.) & H. L. Roediger III (Volume Ed.), *Learning and memory: A comprehensive reference. Vol. 2: Cognitive psychology of memory* (pp. 325-348). Amsterdam: Elsevier.

MACMILLAN, N. A., & CREELMAN, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271. doi:10.1037/0033-295X.87.3.252

MCCABE, D. P., & SMITH, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, **30**, 1065-1077.

MCCONNELL, M. D., & HUNT, R. R. (2007). Can false memories be corrected by feedback in the DRM paradigm? *Memory & Cognition*, **35**, 999-1006.

MCDERMOTT, K. B., & ROEDIGER, H. L., III (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory & Language*, **39**, 508-520. doi:10.1006/jmla.1998.2582

NEUSCHATZ, J. S., PAYNE, D. G., LAMPINEN, J. M., & TOGLIA, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory*, **9**, 53-71. doi:10.1080/09658210042000076

PARKS, T. E. (1966). Signal-detectability theory of recognition-memory performance. *Psychological Review*, **73**, 44-58. doi:10.1037/h0022662

PASHLER, H., CEPEDA, N. J., WIXTED, J. T., & ROHRER, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 3-8. doi:10.1037/0278-7393.31.1.3

READ, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review*, **3**, 105-111.

RHODES, M. G., & JACOBY, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **33**, 305-320. doi:10.1037/0278-7393.33.2.305

ROEDIGER, H. L., III, & MCDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814. doi:10.1037/0278-7393.21.4.803

SCHNEIDER, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments, & Computers*, **20**, 206-217.

SEAMON, J. G., LUO, C. R., SCHLEGEL, S. E., GREENE, S. E., & GOLDENBERG, A. B. (2000). False memory for categorized pictures and words: The category associates procedure for studying memory errors in children and adults. *Journal of Memory & Language*, **42**, 120-146. doi:10.1006/jmla.1999.2676

SMITH, S. M., WARD, T. B., TINDELL, D. R., SIFONIS, C. M., & WILKENFELD, M. J. (2000). Category structure and created memories. *Memory & Cognition*, **28**, 386-395.

STARNS, J. J., LANE, S. M., ALONZO, J. D., & ROUSSEL, C. C. (2007). Metamnemonic control over the discriminability of memory evidence: A signal detection analysis of warning effects in the associative list paradigm. *Journal of Memory & Language*, **56**, 592-607. doi:10.1016/j.jml.2006.08.013

TITUS, T. G. (1973). Continuous feedback in recognition memory. *Perceptual & Motor Skills*, **37**, 771-776.

UNKELBACH, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, **17**, 339-345. doi:10.1111/j.1467-9280.2006.01708.x

UNKELBACH, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **33**, 219-230. doi:10.1037/0278-7393.33.1.219

VERDE, M. F., MACMILLAN, N. A., & ROTELLO, C. M. (2006). Measures of sensitivity based on a single hit rate and a single false alarm rate: The accuracy, precision, and robustness of d', A_z, and A'. *Perception & Psychophysics*, **68**, 643-654.

VERDE, M. F., & ROTELLO, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition*, **35**, 254-262.

WESTERBERG, C. E., & MARSOLEK, C. J. (2006). Do instructional warnings reduce false recognition? *Applied Cognitive Psychology*, **20**, 97-114. doi:10.1002/acp.1170

WHITTLESEA, B. W. A. (2002a). False memory and the discrepancy-attribution hypothesis: The prototype-familiarity illusion. *Journal of Experimental Psychology: General*, **131**, 96-115. doi:10.1037/0096-3445.131.1.96

WHITTLESEA, B. W. A. (2002b). Two routes to remembering (and another to remembering not). *Journal of Experimental Psychology: General*, **131**, 325-348. doi:10.1037/0096-3445.131.3.325

WHITTLESEA, B. W. A., & PRICE, J. R. (2001). Implicit/explicit memory versus analytic/nonanalytic processing: Rethinking the mere exposure effect. *Memory & Cognition*, **29**, 234-246.

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

NOTES

1. In a related study, Healy and Kubovy (1978) found that adding trial-by-trial feedback to information about prior probabilities or a biasing payoff schedule did not exert any additional influence on criterion placement. Since feedback was not independently manipulated, however, that experiment is of limited relevance to the present work.

2. Although confidence-graded recognition ratings were collected, we report hit and false alarm rates throughout the article for ease of interpretation. Analyses of confidence ratings yielded the same conclusions as those of the hit and false alarm rates in each experiment.

3. Data collection for these norms is ongoing. Norms are available for use by interested researchers and may be requested by contacting either of the authors.

4. We also conducted a version of Experiment 3 in which participants were allowed to respond to test probes whenever they wanted; there too we observed no hint that feedback enabled the participants to avoid the false familiarity effect of critical lures.

5. Prior to the RT analyses in the SC condition, the data of 1 participant who did not provide any correct Category A responses were removed.

(Manuscript received August 29, 2008;
revision accepted for publication December 8, 2009.)