Illusory recollection and dual-process models of recognition memory

Philip A. Higham
University of Southampton, Southampton, UK

John R. Vokey
University of Lethbridge, Lethbridge, Alberta, Canada

P. A. Higham and J. R. Vokey (2000, Exps. 1 & 3) demonstrated that a slight increase in the display duration of a briefly presented word prior to displaying it in the clear for a recognition response increased the bias to respond “old”. In the current research, three experiments investigated the phenomenology associated with this illusion of memory using the standard remember–know procedure and a new, independent-scales methodology. Contrary to expectations based on the fluency heuristic, which predicts effects of display duration on subjective familiarity only, the results indicated that the illusion was reported as both familiarity and recollection. Furthermore, manipulations of prime duration induced reports of false recollection in all experiments. The results—in particular, the implications of illusory recollection—are discussed in terms of dual-process, fuzzy-trace, two-criteria signal detection models and attribution models of recognition memory.

Over the past 30 years, dual-process theory (e.g., Atkinson & Juola, 1973, 1974; Jacoby, 1991; Jacoby & Dallas, 1981; Mandler, 1979, 1980; Yonelinas, 1994, 1997) has remained a dominant account of recognition memory. The theory proposes that recognition memory is based on two qualitatively and quantitatively distinct processes, commonly referred to as recollection and familiarity (see Yonelinas, 2002, for a recent review). Recollection (with or without familiarity) involves the conscious retrieval of veridical episodic information from an earlier encounter with a stimulus and gives rise to a feeling of reliving a past event. On the other hand, familiarity is associated with fluent conceptual and perceptual processing, stimulus similarity, and a vague, source nonspecific feeling of remembrance. Thus, it follows from dual-process
theory that recognition hits (accepting studied targets as “old”) may be caused by familiarity, recollection, or both processes, whereas recognition false alarms (accepting new distractors as “old”) are based only on familiarity.

Because participants’ phenomenology is associated with the processes underlying recognition memory, procedures have been developed to assist researchers in gaining access to participants’ mental experiences during memory tasks. One of these is the remember–know (R–K) procedure, introduced by Tulving (1985). With this procedure, as it applies to recognition memory, participants are instructed to indicate whether each test item rated “old” was “remembered” (R) or “known” (K). If the item is given an R rating, that is said to indicate that some aspect(s) of the context surrounding the item’s study episode has been retrieved. K ratings, on the other hand, are thought to indicate that the person believed that the test item was in the study list (high familiarity), but that the details regarding that previous encoding context of the item have not been retrieved (no recollection).

It is also widely assumed in the dual-process literature that R judgements are an accurate measure of recollection and do not reflect variations in familiarity (e.g., Gardiner, 1988; Jacoby, 1998; Rajaram, 1993; Yonelinas & Jacoby, 1995)— an assumption that we refer to as the recollection–remember identity assumption. The recollection–remember identity assumption, together with the exclusion of recollection as a source of false alarms, produces a serious problem for most dual-process accounts of recognition: Such theories have no mechanism to account for systematic variation in false R ratings. Indeed, false R errors are sometimes simply dismissed as noise. For example, in his review of dual-process theory, Yonelinas (2002) estimated recollection by subtracting R false alarms from R hits. This assumes that the observed R hit rate is composed of “true” recollection plus some error (which should be removed), which, in this case, is estimated by the R false alarm rate. This simple, but crucial, gap in the way that false R data are interpreted in dual-process theory has important ramifications, and it is one of the central issues in the research reported in this article.

Dual-process models and false recollection

Probably because of the belief that R judgements can only arise from mechanisms applied to or initiated by old items, false R judgements are usually not the focus of R–K research motivated from a dual-process theoretical viewpoint. To illustrate, consider first the fact that analyses of R data in published research have concentrated principally on the R hit rate (although see Donaldson, 1996; Gardiner, Richardson-Klavehn, & Ramponi, 1997; Hicks & Marsh, 1999; Hirshman & Henzler, 1998; Hirshman & Master, 1997; Inoue & Bellezza, 1998; Strack & Förster, 1995; Xu & Bellezza, 2001, and a few others for exceptions). The R false alarm rate is assumed (and often shown) to be minimal and to reflect a negligible number of responses. As a result, analyses of false alarms, or even discrimination measures (e.g., \(d'\) or \(A'\)), are rare. As Donaldson has stated, “Recently, there has been an increased tendency [in recognition memory research] to use the hit rate as if it accurately measures memory” (Donaldson, 1996, p. 523). Second, most researchers attempting to manipulate R judgements have done so by varying the kind of encoding that old items undergo at study. These variations include the level of processing of the study items (Gardiner, 1988; Gardiner, Gregg, Mashru, & Thaman, 2001; Rajaram, 1993), reading versus generating study words (Gardiner, 1988; Java, 1994), the attentional resources available during study (Gardiner et al., 2001; Gardiner & Parkin, 1990),
the orthographic distinctiveness of the study words (Rajaram, 1998), the conceptual salience of the study words (Rajaram, 1998), the perceptual form of the study items (picture versus word: Dewhurst & Conway, 1994; Rajaram, 1993, 1996; Wagner, Gabrieli, & Verfaellie, 1997), and the number of study time presentations (Jacoby, Jones, & Dolan, 1998), to name a few. The focus on old item encoding at study when it comes to manipulating R judgements suggests that investigating false R judgements is of low priority in these lines of research.

The failure on the part of dual-process models to account for false R judgements is particularly problematic with respect to the growing body of research that has used the R–K procedure, or some variant, to examine the nature of false memories (e.g., Dalla Barba, 1993; Dewhurst, 2001; Higham, 1998; Holmes, Waters, & Rajaram, 1998; Lane & Zaragoza, 1995; Norman & Schacter, 1997; Payne, Elie, Blackwell, & Neuschatz, 1996; Read, 1996; Roediger, Jacoby, & McDermott, 1996; Roediger & McDermott, 1995; Schacter, Verfaellie, & Anes, 1997; Whittlesea, 2002). Rather than the R false alarm rate being negligible, this research has found that participants quite often rate their false memories as “remembered”, rather than merely “known”. A number of explanations have been advanced to explain such false R data, including inappropriate binding of episodic content to new items (Holmes et al., 1998; Reyna & Lloyd, 1997), failures of source monitoring (Higham, 1998; Johnson, Hashtroudi, & Lindsay, 1993; Lane & Zaragoza, 1995) and phantom recollection (Brainerd, Wright, Reyna, & Mojardin, 2001). Our point, however, is not that there are no mechanisms that can account for false R data. It is that dual-process theory, as it is typically understood, cannot.

To illustrate further the problem of false R data for dual-process theory, consider the following two different cases in which it has been clearly highlighted. Both cases are taken from the literature on word frequency effects in recognition.

**Case 1.** Typically in recognition experiments, high-frequency words receive lower hits, but higher false alarms, than low-frequency words, an effect dubbed the *mirror effect* (e.g., Glanzer & Bowles, 1976; Higham & Brooks, 1997). Reder et al. (2000) applied the R–K procedure to recognition experiments in which the preexperimental frequency of the words used as stimuli was varied. They found that both K hits and false alarms were higher for high-frequency words than for low-frequency words. In contrast, the R hit rate was higher for low-frequency words than for high-frequency words.

Reder et al. (2000) argued that their results were consistent with their quantitative *source of activation confusion* model—essentially a dual-process model—in which words are represented in two ways. First, there are word nodes, which are abstract representations of preexperiential experiences with the word. Links fan out from the word nodes, with greater fan for high- than for low-frequency words (more associations). Activating a word node, either through reading a (test) word, or through activation spreading from other nodes via links, generates K responding, and is analogous to familiarity. In addition to word nodes, there are also *episode nodes* in the source of activation confusion model. These nodes represent the contextual details of each within-experiment study encounter with each word. Links form between episode nodes and word nodes at study so that activation can spread to and from a word node’s associated episode node. If reading a previously studied word at test causes enough activation of the episode node, an R response is generated, and recollection has occurred.

The source of activation confusion model accounts well for the pattern of results that Reder et al. (2000) obtained. However, because R judgements derive from episode nodes, and
because new items have no associated episode nodes, no mechanism is in place in the source of activation confusion model to account for false R judgements. Consequently, any systematic variation (from independent variables) in the R false alarm rate is not and cannot be predicted by the model. Such systematic variation was found in their Experiment 1; the rate of false R judgements was significantly lower for low-frequency than for high-frequency words, despite false R rates being near floor.

Case 2. Other research on word frequency effects in recognition that also makes the recollection–remember identity assumption characteristic of dual-process theory has revealed larger effects on false R judgements. For example, Joordens and Hockley (2000, Exps. 1A & 1B) limited recollection by extending the retention interval between study and test in a standard recognition experiment using the R–K procedure. They predicted that the hit rate portion of the mirror effect would reduce, in response to the retention interval increase, because it was based primarily on recollection. Furthermore, the effect of retention interval on hits should be reflected primarily in R judgements because of the recollection–remember identity assumption. In contrast, they predicted that the false alarm portion of the mirror effect, because it was based on familiarity, would remain intact across manipulations of retention interval and would be reflected primarily in K judgements.

Joordens and Hockley (2000) made very similar assumptions in their predictions as Reder et al. (2000); that is, they assumed that R judgements reflect recollection, which supports hits, not false alarms. However, their independent variables had systematic effects on false R judgements; across two experiments, both low- and high-frequency words showed more than double the rate of R false alarms in the delay condition than in the immediate condition. Additionally, across four different comparisons, high-frequency words received more R false alarms than did low-frequency words. As with Reder et al.’s source of activation confusion model, these effects cannot be explained by Joordens and Hockley’s dual-process model because of the absence of any mechanism to account for systematic variation in R false alarms.

Overview of the experiments

The procedure we used in our experiments was based on one we developed elsewhere (Higham & Vokey, 2000; see also Watkins & Gibson, 1988). In those experiments, participants first studied a long list of words and later were administered a recognition memory test. However, prior to the presentation of each test item (target) for a recognition judgement, the same item (prime) was presented very briefly and then masked with ampersands (e.g., the target “table” was primed with a brief presentation of “table”). Participants were asked to try to identify the prime and then to make a recognition response to the target subsequently presented in the clear. Unbeknownst to participants, we varied the duration of the prime presentation; primes in the long duration condition were presented for 30 ms longer than primes in the short duration condition. For the brief durations we used in our prior research,

1We also included experiments in which the long and short prime display durations were separated by over 230 ms, which yielded very different results from those described in the text (Higham & Vokey, 2000; Exp 2 and the 16. 67–250-ms condition of Exp. 3). However, the methodology of the current experiments was more similar to that used in Experiment 1, so when we discuss our previous research on memory illusions induced by prime identification, we are referring to that experiment.
we found that both old and new targets following long duration primes were more likely to be rated “old” than were targets following short duration primes. Consequently, given that we used similar durations in the experiments reported here, we expected a similar illusory memory effect of prime duration on recognition. However, our main concern here was with the effect of prime duration on false alarms and, in particular, with how the associated phenomenology would be rated by participants. Is the effect on new items manifested principally on familiarity ratings and K judgements as anticipated by dual-process models, or can we demonstrate similar effects on recollection ratings and R judgements?

The second purpose of the current research was to introduce a new methodology for investigating participants’ phenomenology during recognition. In Experiments 2A and 2B, instead of using the standard R–K procedure, we asked participants to make two independent memory judgements for each item. After being taught the distinction between familiarity and recollection by adapting instructions that are now standard in the R–K literature, participants were asked to indicate how familiar the item was and, separately, to indicate how much the item was recollected. A different 4-point scale was used to make each judgement (1 = low; 4 = high, for both scales). As we discuss subsequently, this methodology avoids a number of undesirable side-effects of the R–K procedure, such as a necessary arithmetic dependence between the R and K rates, and its inability to isolate confident correct rejections.

EXPERIMENT 1

Other research has been conducted that has used prime manipulations similar to ours. For example, Rajaram (1993) replicated an illusion of memory first demonstrated by Jacoby and Whitehouse (1989). The illusion occurs when participants increase their bias to respond “old” to a recognition stimulus that follows a brief presentation of itself versus a brief presentation of some other, unrelated word. Rajaram found that when the R–K procedure was used to measure participants’ phenomenology, this illusion of memory was manifested in K judgements. Rajaram maintained that her results supported Jacoby and Whitehouse’s claim that the perceptual fluency of the target was enhanced by the matching prime compared to the mismatching prime. This enhanced fluency resulted in a feeling of familiarity, which, in turn, enhanced the likelihood of a K judgement.

The Jacoby–Whitehouse illusion investigated by Rajaram (1993) is not the same as the one reported in Higham and Vokey (2000) because her comparison was between prime–target match and mismatch trials, not between different prime display durations for prime–target match trials. However, to explain the latter Higham and Vokey effect, it could similarly be argued that prime duration enhances the fluency with which the subsequent matching target is processed and that participants experience this fluency as familiarity, which then increases their bias to respond “old”. From a fluency perspective, then, any effect of prime duration should be manifested in familiarity (K) judgements. However, we argued (Higham & Vokey, 2000) that the memory illusion we observed was due to use of an identification heuristic, and not target-processing fluency; that is, participants partially based their recognition decisions on their ability to identify primes, using successful identification as indicator that the item was probably old. Participants may rely on identification performance because old items are generally identified better than new items (perceptual priming), so such performance is in fact predictive of prior presentation. However, the memory illusion arises when participants fail to
consider (consciously or otherwise) other sources of identification ability (such as the slight, but effective, difference in prime duration) and overattribute identification success to prior presentation. For current purposes, if this memory illusion is, indeed, based on prime identification success and not target-processing fluency, it is not clear what the phenomenology associated with the memory illusion should be, or how participants would translate that experience into memory judgements. To find out, in Experiment 1, we induced the memory illusion using a prime display duration manipulation at test and investigated participants’ phenomenology using standard R–K methodology.

Method

Participants

Participants were 40 undergraduate psychology students who participated in return for course credit. Participants were tested individually.

Apparatus

All displays were controlled by Pentium PCs using the default 25 × 80, white-on-black MS-DOS text mode with 640 × 400-pixel displays. The vertical refresh rate was 70 Hz. All training items were displayed in lower case, centred vertically and horizontally on the screen.

Design and materials

Study phase. A pool of 148 five-letter English words of medium lexical frequency (M = 50 occurrences/million; SD = 52) was used. The pool was divided into two lists of 74 items that were balanced in terms of lexical frequency. Each participant was shown words from only one of the lists during the training phase. Words from the exposed list acted as “old” words on the recognition test, whereas words from the unexposed list acted as “new” words. Assignment of lists to the old and new conditions was counterbalanced across participants.

Test phase. All 148 words from both training lists were presented in the test phase of the experiment. Each test trial consisted of (1) a fixation stimulus (*) presented for 1000 ms, (2) presentation of the prime for either short or long duration, (3) presentation of a mask (&&&&&) that remained in view until the space bar was depressed, and (4) presentation of the target (which always matched the prime) until a memory judgement was made. Half the words on the test list were presented for short duration (6–34 ms) whereas the other half were presented for long duration (36–64 ms), and this factor was crossed with prior presentation (old vs. new). Durations of 6–34 ms and 36–64 ms were achieved by programming durations of 20 ms and 50 ms, but taking into account the refresh rate of the monitors that we were using. In Experiments 1 and 2A, we used 70-Hz monitors, which means that the displays were refreshed every 14.3 ms. Depending upon the location of the raster gun when the video memory was changed by the program, some amount of time, from 0 to 14.3 ms, was added or subtracted from the programmed duration. Hence, participants experienced an almost perfectly uniform range of prime display durations from 6 to 64 ms, but with displays 6–34 ms defined as “short” and displays 36–64 ms defined as “long”. Thus, there were four experimental conditions: old–short, old–long, new–short, and new–long. Assignment of words to the long and short duration conditions was also counterbalanced across participants.
Procedure

Study phase. Participants were instructed to try to memorize the word list that would be presented to them on the computer screen by reading each word aloud. To reinforce the instructions, a message to this effect appeared on the computer screen for participants to read after they had received the instructions verbally. After participants indicated that they understood the instructions, they pressed the space bar, and each word from the 74-word training list was presented once to each participant, one at a time, for 2 s each in a random order. When all 74 words had been presented, a message was displayed that indicated that the training phase had ended and that they should wait for instructions.

Test phase. Participants were informed that the test phase consisted of a series of trials. Each trial began with the presentation of a fixation stimulus (*), which indicated where the rest of the events in the trial would occur, and participants were instructed to look at it while it was in view. Following the fixation stimulus, a word was briefly presented and then masked with ampersands (&&&&&). These ampersands remained in view until participants tried to identify the word by writing it down on paper that was provided, guessing if necessary. To ensure that participants’ identification attempts applied to the appropriate test trial, the numbers 1 to 148 were printed on the paper that participants used for prime identification, and the trial number (1–148) was displayed on the computer monitor for each test trial. After attempting to identify the prime, participants were instructed to press the space bar to reveal the test item. Memory judgements were also made using paper; next to each word written as an identification attempt were two columns. After viewing the target item in the clear, participants were instructed to insert either an “O” (for “old”) or “N” (for “new”) next to the item in the first column and then either an “R” (for “remember”) or “K” (for “know”) in the second column for those items rated “old”. The particular instructions used to describe the R–K procedure to participants were taken from Rajaram (1993) and are shown in Appendix A.

Results and discussion

In all experiments, the recognition ratings were analysed first (both raw scores and signal detection measures of discrimination and bias), followed up with an analysis of the identification performance. Our primary statistical tool was analysis of variance (ANOVA). An alpha level of .05 was adopted for all analyses, and all significant effects from the ANOVAs are reported. However, for the sake of brevity, nonsignificant effects are not, unless they are considered important for exposition.

Recognition ratings. The mean likelihoods of “old”, R, and K ratings are shown in Table 1. The recognition, R ratings, and K ratings were submitted to three separate 2 (prior presentation: old, new) × 2 (duration: short, long) within-subjects ANOVAs. All ANOVAs revealed main effects of prior presentation: recognition, old = .72, new = .29, F(1, 39) = 707.39, MSE = 0.010; R, old = .40, new = .07, F(1, 39) = 361.36, MSE = 0.012; K, old = .32, new = .22, F(1, 39) = 27.52, MSE = 0.014. Additionally, a significant effect of duration was revealed for the analysis of the R judgements, F(1, 39) = 4.26, MSE = 0.004; items preceded with primes presented for long duration were judged to be “remembered” more often (.25) than items preceded by primes presented for short duration (.23).

Given the importance of establishing that duration specifically influenced false recollection, the new items were analysed separately. A within-subjects one-way ANOVA indicated that there were significantly more recognition false alarms in the long duration condition (.32)
than in the short duration condition (.26), $F(1, 39) = 5.01$, $MSE = 0.012$. More important, lengthening the prime display duration reliably induced false R judgements; although the effect was small, new items preceded by a matching prime displayed for long duration were significantly more likely to receive false R judgements (.09) than were new items preceded by a short duration prime (.06), $F(1, 39) = 7.29$, $MSE = 0.002$ (see Table 1).

The recognition data were further analysed by determining nonparametric indices of old–new discrimination ($A'$, Grier, 1971) and response bias ($B''_D$, Donaldson, 1992). These indices are shown in the top panel of Table 2. Six 1-way, within-subjects ANOVAs were completed in total. All tested the effect of prime display duration on recognition, K, and R judgements, with the first three testing $A'$ and the latter three testing $B''_D$. Only the effect of duration on $B''_D$ for R judgements was significant, $F(1, 39) = 6.97$, $MSE = 0.011$. Although generally quite conservative, response bias for R judgements was significantly more liberal in the long duration condition (.82) than in the short duration condition (.89). No significant effects were found for discrimination.

### Table 1

Mean likelihoods of recognition, R, and K judgements in Experiment 1 as a function of prior presentation and prime duration

<table>
<thead>
<tr>
<th>Memory measure</th>
<th>Experimental condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition</td>
<td>.71</td>
</tr>
<tr>
<td>R</td>
<td>.40</td>
</tr>
<tr>
<td>K</td>
<td>.32</td>
</tr>
</tbody>
</table>

*Note: R and K means may not sum to the mean for recognition because of rounding error.*

### Table 2

Old–new discrimination and response bias in Experiments 1, 2A and 2B, as a function of response type and duration

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Memory measure</th>
<th>$A'$</th>
<th>$B''_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Short</td>
<td>Long</td>
</tr>
<tr>
<td>1</td>
<td>Recognition</td>
<td>.81</td>
<td>.79</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>.61</td>
<td>.59</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>.79</td>
<td>.78</td>
</tr>
<tr>
<td>2A</td>
<td>Familiarity</td>
<td>.67</td>
<td>.65</td>
</tr>
<tr>
<td></td>
<td>Recollection</td>
<td>.73</td>
<td>.74</td>
</tr>
<tr>
<td>2B</td>
<td>Familiarity</td>
<td>.71</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>Recollection</td>
<td>.76</td>
<td>.72</td>
</tr>
</tbody>
</table>

Identification performance. The mean likelihoods of identifying primes in Experiment 1, as a function of prior presentation and duration, are shown in Table 3. A 2 (prior presentation: old, new) × 2 (duration: short, long) within-subjects ANOVA revealed main effects of prior presentation, $F(1, 39) = 84.14$, $MSE = 0.004$, and duration, $F(1, 39) = 316.57$, $MSE = 0.035$. Old primes were more likely to be identified (.64) than new items (.55; perceptual priming, Jacoby & Dallas, 1981), and primes presented for long duration were more likely to be identified (.86) than primes presented for short duration (.33).

Overall, the main results of Experiment 1 were the following:

1. Longer prime display duration significantly increased the likelihood of R judgements, and significantly reduced R bias, relative to shorter prime duration.
2. For new items, the longer prime display duration significantly increased the likelihood of both recognition false alarms (“old” responses) and the associated subset of R false alarms, relative to the shorter prime duration. In contrast, the K false-alarm rate was not significantly affected by the prime display duration manipulation. This pattern of results suggests that participants sometimes experience the illusion of memory brought about by varying prime duration as something they would rather attribute to recollection than to familiarity.
3. Old items were identified better than new items (perceptual priming, Jacoby & Dallas, 1981).

**EXPERIMENT 2**

A serious problem with the standard R–K procedure is that the effect of prime duration on R judgements that we observed in Experiment 1 could be due to the fact that any effect on recognition must be reflected in R, K, or both. Indeed, inclusion of a third “guess” category in an R–K paradigm has been shown to change the pattern of R–K results obtained (e.g., Gardiner et al., 1997), suggesting that the constraints imposed by the either–or nature of the standard R–K procedure are problematic. Although, strictly speaking, R judgements are not forced to increase if “old” responding increases (i.e., the effect could be manifested in K judgements), certainly an experimental effect on R would be more impressive if it was free to vary independently of old–new recognition decisions. Consequently, instead of using the standard R–K procedure in Experiment 2, we introduced an independent scales methodology to investigate our participants’ reported phenomenology during the recognition task. Participants were not asked for an overt recognition response, but instead were asked to rate how much each test...
item was “familiar” and separately, how much each item was “recollected” using two separate 4-point rating scales. The distinction between recollection and familiarity was made explicit using amended R–K instructions. However, unlike the R–K instructions, participants were persuaded to use the two scales independently. For example, they were provided with examples that were meant to represent all four combinations of familiarity (Y–N) and recollection (Y–N), and they were persuaded that all four combinations were possible (see Appendix B for instructions). Two different versions of Experiment 2 were run, differing in how the prime displays were controlled.

EXPERIMENT 2A

Method

Participants

Participants were 28 undergraduate psychology students who participated in return for course credit. Participants were tested individually.

Design and materials

The design and materials were the same as those in Experiment 1 except that, to test the generality of the effect of prime identification on recognition, the prime display durations were decreased slightly to 1–29 ms in the short condition and to 31–49 ms in the long condition.

Procedure

All aspects of the procedure were the same as those in Experiment 1, except for the following differences:

1. The recognition rating: After viewing the target item in the clear, participants were instructed to write an integer between 1 and 4 (where 1 = definitely no, 2 = probably no, 3 = probably yes, 4 = definitely yes) in each of the two spaces provided next to each identification attempt. The first value was meant to represent the degree to which the target item was familiar, and the second the degree to which the target item was recollected. No overt yes–no recognition response was required. The detailed instructions used to prepare participants for the recognition ratings may be found in Appendix B.

2. Clarification of the instructions: Steps were taken in an attempt to ensure that participants fully understood the distinction between familiarity and recollection. First, a printout of the instructions for the test phase of the scales condition was administered to all participants; they were asked to read the instructions silently while the experimenter read them aloud. Second, after being instructed about the familiarity–recollection distinction, the experimenter asked participants to explain the distinction in their own words. When participants had displayed a satisfactory level of understanding of these concepts, the experimenter explained the 4-rating scales and read aloud examples of the different types of rating that the participants would be completing. These examples were meant to highlight the fact that the two scales could be used completely independently.

Following these examples, participants were administered a four-item questionnaire to ensure that the relevant distinctions were clear to them. The questionnaire contained four scenarios, each meant to exemplify (1) both recollection and familiarity, (2) neither recollection nor familiarity, (3) familiarity without recollection, and (4) recollection without familiarity. The participants’ task was to read each
scenario and decide which of these four combinations applied. Participants were able to proceed with the actual experiment only if they achieved a perfect score on the questionnaire.

Results and discussion

Recognition ratings. The recognition data in each experimental condition were first analysed by examining participants’ mean ratings on the recollection and familiarity scales (shown in the top panel of Table 4). A 2 (scale: recollection, familiarity) × 2 (prior presentation: old, new) × 2 (duration: short, long) within-subjects ANOVA revealed main effects of prior presentation, $F(1, 27) = 92.27, \text{MSE} = 0.295$, and duration, $F(1, 27) = 15.45, \text{MSE} = 0.386$. Old items were rated higher (2.70) than new items (2.01). More important, items displayed for long duration were rated higher (2.52) than items presented for short duration (2.19). The effect of duration was significant whether it was tested as a main effect with the data from each scale pooled, as in the previous analysis, or as orthogonal contrasts (i.e., simple effects) for each scale independently using the scale by duration interaction error term: $\text{MSE} = 0.051$; familiarity: $F(1, 27) = 51.41$; recollection: $F(1, 27) = 66.68$. The analysis that pooled the data from both scales indicated that there was a significant prior presentation by scale interaction, $F(1, 27) = 9.61, \text{MSE} = 0.107$. Orthogonal contrasts using this interaction error term revealed that new item ratings were significantly different between the scales, $F(1, 27) = 18.80$ (familiarity = 2.14 vs. recollection = 1.87), whereas old item ratings did not differ, $F<1$ (familiarity = 2.70 vs. recollection = 2.70). The fact that duration did not interact with either prior presentation, $F<1$, or scale, $F<1$, indicates that it had a comparable effect on both old and new items and on both rated recollection and familiarity.

It is possible that the previous effects on mean ratings occurred because of variations of unsure and guess responses, but not definitely yes or no responses. That is, our manipulations may only have affected participants’ use of scale values 2 and 3, leaving high-confidence judgements (1 and 4) unaffected. To investigate this possibility, the frequency distributions across the four scale values for the recollection and familiarity scales were determined. Initial examination of these distributions indicated that scale values 2 and 3 were actually the least sensitive to our manipulations, and they all seemed to hover around $.18–.20$. Accordingly, we eliminated these data from the following analysis and focused on only the proportions of “definitely yes” (4) and “definitely no” (1) responses. There was no necessary arithmetic dependence between the proportions associated with 1 and 4 (unlike the R–K procedure, where $R+K = \text{old}$), so “value” was included in a $2 \times 2 \times 2 \times 2$ within-subjects ANOVA as a factor with two

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Memory measure</th>
<th>Short–old</th>
<th>Long–old</th>
<th>Short–new</th>
<th>Long–new</th>
</tr>
</thead>
<tbody>
<tr>
<td>2A Familiarity</td>
<td>2.56</td>
<td>2.84</td>
<td>1.98</td>
<td>2.30</td>
<td></td>
</tr>
<tr>
<td>2A Recollection</td>
<td>2.53</td>
<td>2.88</td>
<td>1.70</td>
<td>2.04</td>
<td></td>
</tr>
<tr>
<td>2B Familiarity</td>
<td>2.58</td>
<td>2.96</td>
<td>1.94</td>
<td>2.28</td>
<td></td>
</tr>
<tr>
<td>2B Recollection</td>
<td>2.52</td>
<td>2.87</td>
<td>1.67</td>
<td>2.05</td>
<td></td>
</tr>
</tbody>
</table>
levels (1, 4), along with scale (recollection, familiarity), prior presentation (old, new), and duration (short, long). Because the experimental manipulations (duration and prior presentation) were expected to affect participants’ tendency to make positive responses, they were predicted to have opposite effects on the likelihood of choosing scale value 1 versus 4. Consequently, the effects of prior presentation and duration should be manifested as two-way interactions involving value.

The $2 \times 2 \times 2 \times 2$ ANOVA revealed a main effect of value, $F(1, 27) = 5.74$, $MSE = 0.174$; participants rated more items as 1 (.35) than 4 (.26). However, as expected, value interacted with both duration, $F(1, 27) = 17.66$, $MSE = 0.068$, and prior presentation, $F(1, 27) = 93.93$, $MSE = 0.058$. Orthogonal contrasts using the interaction error term indicated that the duration by value interaction occurred because items following long duration primes were assigned 4 significantly more often (.31) than items following short duration primes (.21), $F(1, 27) = 8.69$, but this effect was significantly reversed for 1 ratings, $F(1, 27) = 8.98$ (.30 and .40 for long and short duration trials, respectively). Similar orthogonal contrasts using the interaction error term revealed that the prior presentation by value interaction occurred because old items were assigned 4 significantly more often (.37) than new items (.15), $F(1, 27) = 45.29$, but this effect was significantly reversed for 1 ratings, $F(1, 27) = 48.67$ (.24 and .47 for old and new items, respectively). However, the value by prior presentation two-way interaction was qualified by a significant three-way interaction including scale, $F(1, 27) = 11.38$, $MSE = 0.019$. That is, the opposite effect of prior presentation on the 1 and 4 scale values (two-way) was greater with the recollection scale than the familiarity scale (three-way).

Generally speaking, the pattern of ratings associated with old items remained fairly constant across the recollection and familiarity scales. However, the new item ratings differed substantially. In particular, new items were more likely to be correctly rejected with the recollection scale than the familiarity scale. The only other significant effect from the ANOVA was a two-way interaction involving prior presentation and duration, $F(1, 27) = 9.41$, $MSE = 0.006$. Old items in the long duration condition had a higher response proportion (.31) than old items in the short duration condition (.29), but the trend was reversed for new items (long, .30; short, .32). This interaction was not anticipated and is not discussed further. In short, the analysis limited to confident “yes” and “no” responses revealed the same pattern of results as the analysis on mean ratings, eliminating the possibility that the experimental manipulations were merely affecting unsure and guess responses.

Because of the current emphasis on false recollection, new item recollection ratings were analysed separately. In three analyses—mean ratings, the proportion of 1 (definitely no) responses, and the proportion of 4 (definitely yes) responses—the effect of duration was significant, $F(1, 27) = 16.21$, $MSE = 0.103$, $F(1, 27) = 19.98$, $MSE = 0.015$, and $F(1, 27) = 7.28$, $MSE = 0.013$, respectively. New items following primes displayed for long duration were assigned higher mean ratings (2.04) and were more likely to receive a response of “definitely yes” (.15), than new items following primes presented for short duration (1.70 and .07, respectively). However, new items were less likely to receive a response of “definitely no” in the long (.44) than short (.58) condition (see Table 4).

As in Experiment 1, discrimination ($A'$) and bias ($B''_{1D}$) were analysed. The results are shown in the middle panel of Table 2. The previous analyses suggest that the recollection scale should have a higher discrimination index because it had lower new item ratings, and about equal old item ratings, than the familiarity scale. In contrast, prime display duration should
have an effect on the bias measure (i.e., long duration more liberal bias than short) because it
touched both old and new item ratings about equally. To conduct the analyses, responses of
“3” or “4” on the 4-point rating scale were scored as “yes” responses so that hit and false alarm
rates could be computed for each participant. A 2 (scale: recollection, familiarity) \(\times 2\) (dura-
tion: short, long) within-subjects ANOVA on \(A'\) revealed a significant effect of scale type, \(F(1, 27) = 9.80, MSE = 0.016\). As expected, old-new discrimination using the recollection scale
was greater (.73) than with the familiarity scale (.66). To determine whether discrimination
was different between the scales because of variations in the hit rate, false alarm rate, or both,
the hit and false alarm rates were entered into a 2 (scale: recollection, familiarity) \(\times 2\) (response
type: hit, false alarm) \(\times 2\) (duration: short, long) within-subjects ANOVA. The ANOVA
revealed main effects of response type, \(F(1, 27) = 85.11, MSE = 0.041\), and duration, \(F(1, 27) = 12.04, MSE = 0.062\), but more important, a response type by scale interaction, \(F(1, 27) = 6.79, MSE = 0.017\). Orthogonal contrasts using the interaction error term indicated that
discrimination differed between the scales because the recollection scale had significantly
fewer false alarms (.28) than the familiarity scale (.38), \(F(1, 27) = 15.89\), whereas the number of
hits was constant, \(F < 1\) (recollection = .57; familiarity = .58). Also as expected, a 2 (scale:
recollection, familiarity) \(\times 2\) (duration: short-long) within subjects ANOVA on \(B''\), revealed
a significant effect of duration, \(F(1, 27) = 12.64, MSE = 0.273\). Response bias was more liberal
in the long duration condition (.01) than in the short duration condition (.36). This effect of
duration on response bias was comparable to the duration effect found for R response bias in
Experiment 1.

The final analysis on the recognition ratings involved deriving R and K responses from the
independent scale data. The proportion of R judgements was estimated as the proportion of
items that were assigned “yes” on the recollection scale, regardless of the associated familiarity
value. In contrast, the proportion of K judgements was estimated as the proportion of items
assigned “yes” on the familiarity scale and “no” on the recollection scale. Four 2 (recollection:
Y, N) \(\times 2\) (familiarity: Y, N) contingency tables are presented in the top panel of Table 5, one
corresponding to each experimental condition in the experiment (old–new \(\times\) short–long),
along with the derived R and K responses.

As in Experiment 1, the proportions of derived K and R responses were analysed with sepa-
rate 2 (duration: short, long) \(\times 2\) (prior presentation: old, new) within-subject ANOVAs. The
analysis on derived R responses yielded results that were similar to those on actual R responses
in Experiment 1: significant main effects of prior presentation, \(F(1, 27) = 97.99, MSE = 0.026\),
and duration, \(F(1, 27) = 12.91, MSE = 0.035\), were found. Old items were more likely to be
remembered (.57) than new items (.27), and items preceded by long duration primes (.48)
were more likely to be remembered than items following short duration primes (.36). Thus,
although the absolute scores differed somewhat in that both the R hit rate and the R false alarm
rate were higher in Experiment 2A than in Experiment 1 (see Tables 1 and 5), actual R judge-
ments and derived R judgements appear to be sensitive to similar variables. These similarities
occurred despite methodological differences between Experiments 1 and 2A, which included
different recognition memory test formats and slightly different display durations.

In contrast, the analogous ANOVA on the derived K responses revealed a very different
pattern of results from that of the actual K responses in Experiment 1. In addition to a
counterintuitive marginal effect of prior presentation, \(F(1, 27) = 3.28, MSE = 0.022, p < .09\),
where more new items (.22) than old items (.17) were rated as familiar-but-not-recollected, a
significant prior presentation by duration interaction was obtained, \(F(1, 27) = 9.00, \text{MSE} = 0.005\). Orthogonal contrasts using the interaction error term indicated that derived K ratings were significantly higher for new items (.24) than old items (.15) in the long duration condition, \(F(1, 27) = 24.33\), but no such difference existed for targets in the short duration condition, \(F < 1\) (old = .19; new = .20). Based on the results from Experiment 1, no interaction of any nature was expected on K judgements, but particularly not one involving higher familiarity for new items than old items!

Although this result was unexpected, others have also found that manipulations that were expected to increase knowing (familiarity), actually decrease it when the R–K procedure is used. For example, Yonelinas and Jacoby (1995) found that matching the size of stimuli (perceptual reinstatement of context) between study and a recognition test decreased knowing relative to a condition where sizes were mismatched (see also Rajaram & Coslett, 1992). This result was counterintuitive in that reinstating the perceptual qualities of the stimuli between study and test should enhance the fluency of processing at test, resulting in more K responses. However, they attributed the counterintuitive pattern of results to the nature of the standard R–K task; because participants cannot make both an R and a K judgement to the same item, and because the presence of recollection always leads to an R judgement, R effectively limits K. Indeed, they found that if K was divided by one minus R (such that R and K were independent), the adjusted K proportion increased in response to size matching.

Because the independent scales method eliminates the either–or nature of responding, it also alleviates the problem of R judgements limiting K judgements—indeed, this is one reason that we designed the method in the first place. However, the problem is only alleviated if the rating on the familiarity scale is taken as a measure of the underlying familiarity process. If, instead, the estimate of the underlying process is a rating of familiarity—in-the-absence-of-

<table>
<thead>
<tr>
<th>TABLE 5</th>
<th>Familiarity–recollection contingency likelihoods and derived remember and know ratings in Experiments 2A and 2B as a function of prior presentation and prime duration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental condition</strong></td>
<td><strong>Short–old</strong></td>
</tr>
<tr>
<td><strong>Experiment</strong></td>
<td><strong>Contingency</strong></td>
</tr>
<tr>
<td>2A</td>
<td>FY–RY</td>
</tr>
<tr>
<td></td>
<td>FY–RN (derived K)</td>
</tr>
<tr>
<td></td>
<td>FN–RY</td>
</tr>
<tr>
<td></td>
<td>FN–RN</td>
</tr>
<tr>
<td></td>
<td>RY (derived R)</td>
</tr>
<tr>
<td></td>
<td>Derived independent K</td>
</tr>
<tr>
<td>2B</td>
<td>FY–RY</td>
</tr>
<tr>
<td></td>
<td>FY–RN (derived K)</td>
</tr>
<tr>
<td></td>
<td>FN–RY</td>
</tr>
<tr>
<td></td>
<td>FN–RN</td>
</tr>
<tr>
<td></td>
<td>RY (derived R)</td>
</tr>
<tr>
<td></td>
<td>Derived independent K</td>
</tr>
</tbody>
</table>

*Note: FY = familiarity–yes; RY = recollection–yes; FN = familiarity–no; RN = recollection–no; R = remember; K = know. The contingencies FY–RY, FY–RN, FN–RY, and FN–RN may not sum to one, and/or FY–RY and FN–RY may not sum to derived R because of rounding error.*
recollection, then the R–K problem identified by Yonelinas and Jacoby (1995) is also a problem with the independent scales method: The recollection rate limits the proportion of available responses that can be rated familiar-but-not-recollected.

Following Yonelinas and Jacoby (1995), we divided the derived K proportion by one minus the derived R proportion to gain an unbiased (by R) estimate of familiarity.\(^2\) The results are shown in the bottom row of the top panel of Table 5. Indeed, the adjustment to K rendered a pattern of familiarity that was much more reasonable; a 2 (prior presentation: old, new) × 2 (duration: short, long) within-subjects ANOVA on the independent estimate of familiarity revealed main effects of prior presentation, \(F(1, 27) = 4.33, MSE = 0.035\), and duration, \(F(1, 27) = 4.73, MSE = 0.028\). Familiarity was higher for old items (.39) than new items (.32), and higher for items following long duration primes (.39) than short duration primes (.32). Thus, the odd results obtained with derived K in the previous analysis, which suggested higher familiarity for new items than old items, were probably due to recollection limiting the number of familiar-but-not-recollected judgements. The interaction from the ANOVA on the adjusted K proportions was also significant, \(F(1, 27) = 4.31, MSE = 0.014\). Orthogonal contrasts using the interaction error term indicated that duration had a significant effect on new items, \(F(1, 27) = 13.56\) (short = .26 vs. long = .37), but not on old items, \(F < 1\) (short = .38 vs. long = .40). The effect of duration specifically on new items (which also resulted in an overall main effect of duration in the ANOVA) was the only point of departure between the two experiments, although there was a trend for new items to receive more K responses in Experiment 1 also.

Identification performance. The mean likelihoods of identifying primes in Experiment 2A, as a function of prior presentation and duration, are shown in Table 3. A 2 (prior presentation: old, new) × 2 (duration: short, long) within-subjects ANOVA revealed main effects of prior presentation, \(F(1, 27) = 52.87, MSE = 0.003\), and duration, \(F(1, 27) = 689.74, MSE = 0.021\). Old items were more likely to be identified (.49) than new items (.41; perceptual priming, Jacoby & Dallas, 1981), and primes presented for long duration were more likely to be identified (.81) than primes presented for short duration (.09). Additionally, the analysis revealed a significant interaction, \(F(1, 27) = 9.34, MSE = 0.006\); the size of the perceptual priming effect was larger for primes presented for long duration (.12) than for primes presented for short duration (.03). This result is probably due to a floor effect, as identification likelihoods in the short duration condition were less than 10% (see Table 3).

**EXPERIMENT 2B**

As explained earlier, there was some (uniform) variability of the prime display durations within the short and long conditions of Experiments 1 and 2A. Also, because the short and long duration conditions were only separated by an average of 30 ms, there was virtually no difference between the longest duration in the short condition and the shortest duration in the

\(^2\)We also adjusted the K proportions in Experiment 1 by dividing them by one minus R and reanalysed the data. Unlike Experiment 2A, the results of the analyses on the adjusted K proportions were the same as those for the raw K proportions.
long duration condition. The lack of a marked division between the two duration categories within each experiment may have increased the likelihood that participants misattributed prime identification performance to the prior exposure of the target items. That is, the uniform distributions may have rendered display duration a less salient attributional source of identification performance, resulting in more misattributions to prior presentation. To test for this possibility, in this next experiment, the uniformity of the short and long prime display duration distributions was eliminated by synchronizing the displays with the vertical raster-scan of the computer monitors.

Experiment 2B was intended as a replication of the independent scales procedure of Experiment 2A, but with fixed display durations. As 60-Hz video monitors were used in this experiment, prime display durations had to be some multiple of 1/60 s, so as to be synchronized with the raster-scan of the monitors. Accordingly, prime display durations of 1/60 and 3/60 s (16.67 and 50 ms) were used as the short and long display durations, respectively.

Method

Participants

A total of 20 undergraduate psychology students participated in exchange for course credit. Participants were tested individually and in groups of 2–4 at individual workstations. A total of 5 participants were assigned at random to each of four counterbalancing conditions produced by rotating the items through the two levels of prior presentation and the two levels of prime display duration.

Apparatus

To achieve synchronization with the raster-scan, Apple //GS computers and monitors were used. These monitors had vertical refresh rates of 60 Hz. All items were displayed in lower case using the default 40 × 24, white-on-black text mode.

Design and materials

The design and materials were similar to those of Experiment 2A. The principal difference was that the prime displays were synchronized to the 60-Hz raster-scan of the video monitors by detecting the vertical blanking interval (or VBLINT) of the video monitors and updating the video memory and, hence, the display only during this interval. One further difference was that participants were prompted for their recognition judgements by the presentation of two response fields on the computer monitor, one labelled “familiarity” and one labelled “recollection”, into which they were to type their ratings. Only scale ratings of 1–4 were accepted, and participants had to type a response into both fields before they were allowed to proceed to the next trial.

---

3 The VBLINT is the period during which the raster-gun of the video monitor returns from the bottom-right to the top-left of the display without affecting the display, after which it again re-paints the display with the contents of video memory. Altering the contents of video memory only during this interval synchronizes the display of the contents of the video memory with the raster-scan of the monitor.
Procedure

The procedure was the same as that in Experiment 2A, except that participants in this experiment read the training list silently rather than aloud because some participants were tested in small groups.

Results and discussion

Recognition ratings. As in Experiment 2A, the recognition data in each experimental condition were first analysed by examining participants’ mean ratings (shown in the bottom panel of Table 4) on the recollection and familiarity scales. Old items received significantly higher ratings (2.73) than did new items (1.98), \( F(1, 19) = 48.50, MSE = 0.465 \), and items displayed for long duration received higher ratings (2.54) than did items presented for short duration (2.18), \( F(1, 19) = 20.24, MSE = 0.259 \). The effect of duration was also significant when effects of duration were tested as orthogonal contrasts (i.e., simple effects) for each scale independently using the scale by duration interaction error term: \( MSE = 0.028 \); familiarity, \( F(1, 19) = 92.60 \); recollection, \( F(1, 19) = 96.59 \). As in Experiment 2A, there was a prior presentation by scale interaction, \( F(1, 19) = 6.69, MSE = 0.043 \). Orthogonal contrasts using the interaction error term revealed that ratings to new items were significantly different between the scales, \( F(1, 19) = 28.57 \) (familiarity = 2.11 vs. recollection = 1.86), whereas ratings to old items did not differ significantly, \( F(1, 19) = 2.84 \) (familiarity = 2.77 vs. recollection = 2.69). As in Experiment 2A, duration did not interact significantly with either prior presentation, \( F < 1 \), or scale, \( F < 1 \), indicating that duration had comparable effects on both old and new items and on both rated recollection and familiarity.

As in Experiment 2A, the relative frequencies of the two extreme scale values (1 = definitely no, and 4 = definitely yes) for the recollection and familiarity scales were analysed, along with prior presentation (old–new), and duration (short–long), in a 2 × 2 × 2 × 2 within-subjects ANOVA. The main effect of scale value was marginally significant, \( F(1, 19) = 3.63, MSE = 0.180, p = .07 \). Participants rated marginally more items as “definitely no” (.34) than “definitely yes” (.25). However, as in Experiment 2A, scale value interacted with both duration, \( F(1, 19) = 17.90, MSE = 0.053 \), and prior presentation, \( F(1, 19) = 44.94, MSE = 0.096 \). An orthogonal contrast using the interaction error term indicated that the duration by value interaction occurred because items following long duration primes were assigned “definitely yes” ratings significantly more often (.30) than items following short duration primes (.20), \( F(1, 19) = 7.40 \), but this effect was significantly reversed for “definitely no” ratings, \( F(1, 19) = 10.65 \) (.28 and .40 for long and short duration primes, respectively). A similar orthogonal contrast revealed that the prior presentation by scale value interaction occurred because old items were assigned ratings of “definitely yes” significantly more often (.38) than new items (.11), \( F(1, 19) = 28.89 \), but this effect was significantly reversed for ratings of “definitely no”, \( F(1, 19) = 16.86 \) (.24 and .44 for old and new items, respectively). However, again as in Experiment 2A, this cross-over interaction of scale value and prior presentation was significantly larger for the recollection scale than the familiarity scale, \( F(1, 19) = 7.41, MSE = 0.009 \). Unlike in Experiment 2A, however, there was no significant interaction between prior presentation and duration, \( F(1, 19) = 2.02, MSE = 0.007, p > .17 \).

As in Experiment 2A, new item recollection ratings were analysed separately to assess the degree of false recollection. The effect of duration was significant for all three analyses: mean ratings, \( F(1, 19) = 19.25, MSE = 0.074 \), the proportion of “definitely no” responses, \( F(1, 19) = 20.24, MSE = 0.259 \), and items displayed for long duration received higher ratings (2.54) than did items presented for short duration (2.18), \( F(1, 19) = 20.24, MSE = 0.259 \). The effect of duration was also significant when effects of duration were tested as orthogonal contrasts (i.e., simple effects) for each scale independently using the scale by duration interaction error term: \( MSE = 0.028 \); familiarity, \( F(1, 19) = 92.60 \); recollection, \( F(1, 19) = 96.59 \). As in Experiment 2A, there was a prior presentation by scale interaction, \( F(1, 19) = 6.69, MSE = 0.043 \). Orthogonal contrasts using the interaction error term revealed that ratings to new items were significantly different between the scales, \( F(1, 19) = 28.57 \) (familiarity = 2.11 vs. recollection = 1.86), whereas ratings to old items did not differ significantly, \( F(1, 19) = 2.84 \) (familiarity = 2.77 vs. recollection = 2.69). As in Experiment 2A, duration did not interact significantly with either prior presentation, \( F < 1 \), or scale, \( F < 1 \), indicating that duration had comparable effects on both old and new items and on both rated recollection and familiarity.

As in Experiment 2A, the relative frequencies of the two extreme scale values (1 = definitely no, and 4 = definitely yes) for the recollection and familiarity scales were analysed, along with prior presentation (old–new), and duration (short–long), in a 2 × 2 × 2 × 2 within-subjects ANOVA. The main effect of scale value was marginally significant, \( F(1, 19) = 3.63, MSE = 0.180, p = .07 \). Participants rated marginally more items as “definitely no” (.34) than “definitely yes” (.25). However, as in Experiment 2A, scale value interacted with both duration, \( F(1, 19) = 17.90, MSE = 0.053 \), and prior presentation, \( F(1, 19) = 44.94, MSE = 0.096 \). An orthogonal contrast using the interaction error term indicated that the duration by value interaction occurred because items following long duration primes were assigned “definitely yes” ratings significantly more often (.30) than items following short duration primes (.20), \( F(1, 19) = 7.40 \), but this effect was significantly reversed for “definitely no” ratings, \( F(1, 19) = 10.65 \) (.28 and .40 for long and short duration primes, respectively). A similar orthogonal contrast revealed that the prior presentation by scale value interaction occurred because old items were assigned ratings of “definitely yes” significantly more often (.38) than new items (.11), \( F(1, 19) = 28.89 \), but this effect was significantly reversed for ratings of “definitely no”, \( F(1, 19) = 16.86 \) (.24 and .44 for old and new items, respectively). However, again as in Experiment 2A, this cross-over interaction of scale value and prior presentation was significantly larger for the recollection scale than the familiarity scale, \( F(1, 19) = 7.41, MSE = 0.009 \). Unlike in Experiment 2A, however, there was no significant interaction between prior presentation and duration, \( F(1, 19) = 2.02, MSE = 0.007, p > .17 \).

As in Experiment 2A, new item recollection ratings were analysed separately to assess the degree of false recollection. The effect of duration was significant for all three analyses: mean ratings, \( F(1, 19) = 19.25, MSE = 0.074 \), the proportion of “definitely no” responses, \( F(1, 19) =
11.18,\( MSE = 0.013\), and the proportion of “definitely yes” responses, \(F(1, 19) = 8.05, MSE = .009\). New items following primes displayed for long duration were assigned higher mean ratings (2.05) and were more likely to receive a response of “definitely yes” (.14) than new items following primes presented for short duration (1.67 and .05, respectively). However, new items were less likely to receive a response of “definitely no” in the long (.44) than in the short (.55) condition.

The recognition ratings were again analysed in terms of the signal detection indices of discrimination \((A')\) and bias \((B''D)\) as described for Experiment 2A. The results are shown in the bottom panel of Table 2. As in Experiment 2A, old-new discrimination \((A')\) using the recollection scale was significantly greater (.73) than that with the familiarity scale (.71), \(F(1, 19) = 6.73, MSE = 0.003\). Similarly, response bias \((B''D)\) was more liberal in the long duration condition (–.02) than in the short duration condition (.44), \(F(1, 19) = 19.95, MSE = 0.215\). To determine whether discrimination was different between the scales because of variations in the hit rate, false alarm rate, or both, the hit and false alarm rates were analysed as a function of duration (short–long), and scale (familiarity–recollection). As in Experiment 2A, there were significant main effects of response type (hit vs. false alarm), \(F(1, 19) = 50.87, MSE = 0.065\), and duration, \(F(1, 19) = 20.87, MSE = 0.040\), and, more important, a prior exposure by scale interaction, \(F(1, 19) = 4.50, MSE = 0.007\). Orthogonal contrasts using the interaction error term showed that, as in Experiment 2A, discrimination differed between the scales primarily because the recollection scale had a significantly lower false alarm rate (.26) than did the familiarity scale (.35), \(F(1, 19) = 25.79, MSE = 0.26\), whereas the hit rates associated with each scale were only marginally different, \(F(1, 19) = 4.31, p < .06\) (recollection = .57 vs. familiarity = .61).

The derived R and K rates were computed and analysed as in Experiment 2A. The results are shown in the bottom panel of Table 5. For derived R rates, as in Experiment 2A, significantly more old items were remembered (.57) than new items (.26), \(F(1, 19) = 51.72, MSE = 0.038\), and items preceded by long duration primes (.49) were significantly more likely to be remembered than items following short duration primes (.34), \(F(1, 19) = 19.98, MSE = 0.021\). For derived K rates, as in Experiment 2A, there was a counterintuitive marginal effect of prior presentation, \(F(1, 19) = 3.73, MSE = 0.012, p < .07\), in which more new items (.18) than old items (.13) were rated as familiar—but-not-recollected. However, as in Experiment 2A, dividing the K rates by one minus R (Yonelinas & Jacoby, 1995) rendered a more reasonable pattern of results; the bottom row of the bottom panel of Table 5 shows that the derived independent K rate was significantly higher for old items (.32) than new items (.24), \(F(1, 19) = 5.98, MSE = 0.021\) and significantly higher for items following long duration primes (.31) than short duration primes (.24), \(F(1, 19) = 5.20, MSE = 0.016\). Unlike Experiment 2A, there was no significant interaction of prior presentation and prime duration (\(F < 1\)) on derived independent K estimates: The effect of duration did not differ significantly between old and new items, as in Experiment 1.

**Identification performance.** The mean likelihoods of identifying primes in Experiment 2B, as a function of prior presentation and duration, are shown in Table 3. A 2 (prior presentation: old, new) \(\times\) 2 (duration: short, long) within-subjects ANOVA revealed main effects of prior presentation, \(F(1, 19) = 24.18, MSE = 0.003\), and duration, \(F(1, 19) = 226.51, MSE = 0.046\). Old items were more likely to be identified (.46) than new items (.40; perceptual priming, Jacoby & Dallas, 1981), and primes presented for long duration were more likely to
be identified (.80) than primes presented for short duration (.07). As in Experiment 2A, the analysis revealed a significant interaction, $F(1, 19) = 4.49, MSE = 0.003$; the size of the perceptual priming effect was larger for primes presented for long duration (.08) than for primes presented for short duration (.03). As we suggested for Experiment 2A, this interaction is most likely due to a floor effect as identifications in the short duration condition were less than 10% (see Table 3).

Overall summary of Experiments 2A and 2B

In summary, the pattern of results from Experiment 2B replicated that of Experiment 2A almost exactly:

1. Relative to shorter prime duration, longer prime display duration significantly increased mean recognition ratings, the likelihood of confident acceptance (i.e., choosing scale value 4), and the likelihood of responding “yes” more generally, whereas it decreased response bias and the likelihood of an item being assigned 1 (confident rejection). These effects of duration were comparable between old and new items, and between the recollection and familiarity scales.

2. When examining recollection and new items separately, longer prime duration (relative to shorter prime duration) significantly increased mean recollection ratings, recollection false alarms, and the likelihood of an item being confidently accepted (4) on the recollection scale, but it decreased the likelihood of an item being confidently rejected (1) on the recollection scale. That is, the duration manipulation affected the degree of rated false recollection.

3. The recollection scale had better old–new discrimination than the familiarity scale, but this effect was attributable specifically to lower new item ratings (false alarm rate), not higher old item ratings (hit rate). (Indeed, the recollection hit rate was marginally lower than the familiarity hit rate in Experiment 2B.) This pattern of results was not limited to unsure and guess responses, but rather was mostly apparent in “definitely yes” (4) and “definitely no” (1) responses.

4. R judgements derived from the recollection and familiarity scales showed similar sensitivity to the experimental manipulations as did actual R judgements in Experiment 1.

5. K judgements derived from the scales were not sensitive in the same way to the same experimental manipulations as were actual K judgements in Experiment 1 and showed a counterintuitive pattern of results (e.g., $K_{\text{new}} > K_{\text{old}}$). However, if the K rate was divided by one minus R to render it independent of the R rate (Yonelinas & Jacoby, 1995), the pattern of results was more reasonable. This suggests that the R rate was limiting the uncorrected K rate.

6. Old items were identified better than new items (perceptual priming; Jacoby & Dallas, 1981).
GENERAL DISCUSSION

The current research has replicated the memory illusion reported by Higham and Vokey (2000), but has extended that earlier work by exploring the attributions associated with the illusion. In short, increasing the duration, and hence identification, of a briefly presented stimulus immediately prior to judging the same stimulus presented in the clear for recognition enhanced both reported recollection (Experiments 1, 2A, and 2B) and familiarity (Experiments 2A and 2B) of that stimulus relative to misidentifying it. This result was obtained both with the standard R–K methodology and with our new independent-scales methodology. The identification effect on subjective familiarity and recollection was not limited to hits, but occurred also for false alarms.

Probably because dual-process theory arose largely as a theory of target recognition, provisions are simply not in place in the theory to account for recollection of false alarms. Although this gap in dual-process theory has been alluded to in the literature (e.g., see Xu & Bellezza, 2001), and some theorists have even attempted to model false recollection in a dual-process framework (e.g., Brainerd et al., 2001), we think its seriousness for dual-process models has been underemphasized in recognition memory research. Accordingly, we now consider two alternative revisions to the assumptions underlying dual-process theory that directly address the false-recollection problem.

Revisions to dual-process theory

Solution 1: Change the relationship between recollection (the process) and items. Instead of viewing recollection as something that can only support hits, dual-process theory might be altered to have the process of recollection occur for both hits and false alarms. This approach is similar to Higham’s (1997) account of specific similarity effects in artificial grammar learning. He found that test items that were specifically similar to single training items were more likely to be endorsed as “grammatical”, and he argued that this effect was recollection based because it was eliminated when attention was divided at test. Following this example, then, independent evidence for the existence of recollection is derived from the effect of manipulations (e.g., dividing attention, lengthening the retention interval, and so on) that, a priori, are expected to affect recollection. Evidence for recollection is not based on the type of item (old vs. new) associated with the response. However, it is not clear how straightforward it would be for dual-process theory to incorporate such a modification to its assumptions given that recollection has always been conceived as something that reflects a veridical encounter with an item. Certainly, rather drastic amendments would have to be made to process dissociation equations (Jacoby, 1991) and formal dual-process models (e.g., source of activation confusion model, Reder et al., 2000) to implement such a change.

Solution 2: Change the relationship between recollection (the process) and recollection judgements. A second, perhaps preferable, solution would involve changing the assumed identity between the recollection process and recollection judgements (the recollection–remember identity assumption) by allowing the familiarity process to affect such judgements. Over 10 years ago, Jacoby (1991) argued that memory researchers were making a “process purity” assumption whereby it was assumed that performance on certain tasks was a pure measure of the underlying process. In our view, a similar, equally questionable “process purity” assump-
tion is being made in R–K research today, only the assumption is that a particular response (R judgement) is a pure measure of the underlying process (recollection). Indeed, a number of false R effects seem to be best explained by variations in familiarity. For example, the fact that new high-frequency words receive higher R ratings than new low-frequency words, whereas the opposite is true of old words (Joordens & Hockley, 2000; Reder et al., 2000), suggests that familiarity is “leaking” into R judgments of new items.

Fuzzy-trace theory

Fuzzy-trace theory (e.g., Brainerd, Reyna, & Brandse, 1995; Reyna & Brainerd, 1992, 1995), although probably better described as a dual-trace theory, shares some characteristics with dual-process theory. It assumes that encountering a stimulus (at study) produces two representations in memory: a verbatim representation, which contains literal details of the original encoding incident, and a gist representation, which contains semantic content. At test, old items typically initiate retrieval of both verbatim and gist representations, whereas new items typically result in the retrieval of gist representations. Old–new recognition is based on the extent to which the retrieved representation(s) is similar to the test item’s surface form and/or gist. In terms of the R–K distinction, usually R judgements result from retrieval of verbatim representations, whereas gist representations drive K judgements. Thus, fuzzy–trace theory greatly resembles Reder et al.’s source of activation confusion model: Source of activation confusion’s word nodes are analogous to fuzzy–trace theory’s gist representations, and source of activation confusion’s episode nodes are analogous to fuzzy–trace theory’s verbatim representations. Despite the similarity, however, fuzzy–trace theory is not necessarily subject to the same problems as Reder’s source of activation confusion model because, in fuzzy–trace theory, R judgements can be based on processes other than veridical recollection.4 For example, Reyna and Lloyd (1997) have argued that false R judgements might occur if a retrieved gist representation is erroneously “ascribed to experience” (p. 99)—that is, if a new item’s gist representation is strongly instantiated at study, and if the new item is an excellent retrieval cue for that gist memory, false recollection can occur. More recently, Brainerd and colleagues (Brainerd, Reyna, & Mojardin, 1999; Brainerd et al., 2001) have dubbed this type of gist–based memory error “phantom recollection” and have modelled its effects by including it as a separate parameter in their multinomial model. Because phantom recollection occurs when high familiarity (gist) “leaks” into R judgements, the inclusion of this mechanism in fuzzy–trace theory means that the theory meets the conditions of Solution 2 above.

4Recently, Arndt and Reder (2003) have proposed a model of recognition that is similar to Reder et al.’s (2000) source of activation confusion model, except that theme nodes are included in addition to episode and concept (word) nodes. During study, activation of concept nodes that are consistent with a similar theme can spread to theme nodes. Theme nodes are associated with the experimental context (as are episode nodes), so their activation can generate R judgements during test. Such a model is useful for explaining false R judgements based on the instantiation of strong semantic gist during study (e.g., Deese, 1959; Roediger & McDermott, 1995). However, it is unclear how such a model could account for the test-based prime duration effects generated in the current experiments, or even if the model can still be considered a dual-process model given the presence of three distinct node types.
Phantom recollection is useful for explaining semantic false R effects, such as those seen in the Deese–Roediger–McDermott paradigm (e.g., Deese, 1959; Dewhurst, 2001; Roediger & McDermott, 1995). False-recollection errors occur in the paradigm when a critical lure’s meaning is primed by the presentation of several related targets at study. Similarly, Reyna and Kiernan (1995; see also 1994) have found in sentence recognition that children produce more false alarms to never-presented true inferences (that are derived from the studied sentences), than to never-presented false inferences, despite the instruction to only accept a literal copy of the original sentence. Both of these false recollection results might be caused by phantom recollection because both effects occur in the context of strong gist instantiation at study and excellent retrieval of that gist at test.

The false R effects obtained in our paradigm also might be produced by phantom recollection. First, assume that identifying a new word at test, before making a recognition response, causes a gist representation of the word to be created. Retrieval of this gist representation should be excellent during the recognition component of the test task, given the temporal proximity of the identification and recognition tasks and the fact that the recognition stimulus always matches the newly formed gist representation. Thus, phantom recollection is more likely for identified new words than misidentified ones. However, phantom recollection cannot explain data from some of our other research using this paradigm. For example, in two experiments (Higham & Vokey, 2000, Exp. 2 and the 16.67–250 ms condition of Exp. 3), we made the manipulation of prime duration extreme; the manipulated difference in prime duration was over 200 ms, rather than 30 ms as used in our current research. Presumably, fuzzy-trace theory predicts that gist representations of identified new words with the extreme manipulation would be even more accessible during recognition than those of misidentified words because, generally, they were presented for longer as primes (i.e., over 200 ms). Thus, fuzzy-trace theory with phantom recollection predicts that the prime identification illusion that we demonstrated in our current research should be more pronounced with the more extreme manipulation of prime duration. However, we found the opposite; under these extreme circumstances, identification of the prime reduced the likelihood of an “old” response compared to misidentification, a result that is directly counter to the prediction of fuzzy-trace theory.

Besides phantom recollection, Reyna and Lloyd (1997; see also Brainerd & Reyna, 1998) have argued that false R judgements might also derive from retrieval of verbatim memories “from the wrong context” (p. 100). Note that, although a verbatim representation has been accessed, it does not count as veridical recollection because the details associated with the memory are incorrect. Verbatim-based false recollection errors might occur in misinformation studies (e.g., Higham, 1998; Loftus, Miller, & Burns, 1978) in which erroneous details, suggested to participants after an event, are “remembered” as having occurred in the original event (wrong context).

In summary, fuzzy-trace theory assumes that the experience of false recollection can be based on either the retrieval of gist or verbatim representations, which marks it as different from other dual-process models. Because judgements of false recollection in fuzzy-trace theory are not limited to the retrieval of veridical study detail, the requirements of Solution 2 above are met. Hence, fuzzy-trace theory is not subject to the false recollection problem in the same way that other dual-process models are. However, this flexibility comes at the cost of precision and testability; there is a lot of room in fuzzy-trace theory for post hoc descriptions of
the data. With more than one mechanism available to account for false recollection, there is the temptation to simply call upon one of these mechanisms after the fact, choosing the one that best “explains” the pattern of results. Of course, such accounts have little value because they are difficult, if not impossible, to disprove.

Recollection versus familiarity

Both Experiments 2A and 2B showed that the false-alarm rate differed between the familiarity and recollection scales, but the hit rate did not. The false-alarm portion of this difference would be expected from a dual-process perspective: For example, it could be argued that the absence of retrieval of veridical study information rendered recollection ratings near floor, whereas fluent processing (intraitem integration) of some new items renders somewhat higher familiarity ratings.

Although the recollection–familiarity false-alarm difference is interpretable with dual-process theory, the similarity in the hit rates, particularly in Experiment 2A, is more awkward to explain. Dual-process theory holds that there are two bases supporting hits, with the recollection basis presumably measured by the recollection scale and the familiarity basis measured by the familiarity scale. The central question is: If the recollection and familiarity scales are measuring different components of memory, then why were the old item ratings virtually indistinguishable between the scales? One possibility is that the similarity was mere coincidence. That is, although the memory substrate measured by the two scales was different, participants chose the same scale values for a given item, such that the pattern of responding between the scales remained invariant. The temporal proximity of the recollection and familiarity ratings may have contributed to the use of similar scale values. However, this argument leaves unexplained why new item ratings differed between the scales. Additionally, the credibility of such an argument is stretched to the limit if one also wants to dismiss as coincidence the fact that prime display duration had almost exactly the same size effect on familiarity and recollection ratings to old items, and this similarity was maintained across two different experiments (see Table 4). A more plausible possibility is that the scales were actually measuring the same thing—that is, participants simply did not distinguish between recollection and familiarity in the presence of studied items. This conclusion is clearly at odds with dual-process theory and with an essential assumption underlying the R–K methodology but, nonetheless, it is the most parsimonious explanation of the current data.

Our finding that recollection correct rejections form the basis of the distinction between the scales indicates a clear advantage of using the independent-scales methodology over the more standard R–K methodology. The independent-scales methodology can identify and separate cases of “probably” (3) and “definitely” (4) not recollected, regardless of the level of familiarity. This possibility is denied with standard R–K methodology. The R–K method was designed only to flag cases where the presence of veridical study information leads to a positive response, not cases leading to confident rejection. The nature of the R–K design means that both “definitely not” and “probably not” recollection judgements are assigned ambiguously to the default categories of “new” and “known”, making it impossible to pinpoint confident recollection correct rejections. Elsewhere, we have argued that rejection mechanisms form an important basis of responding in recognition and classification tasks (Brooks, Vokey, & Higham, 1997; Higham & Brooks, 1997; Higham, Vokey, & Pritchard, 2000; Vokey & Illusory Recollection 23
Higham, 1999; see also, Johns & Mewhort, 2002; Mewhort & Johns, 2000; Whittlesea, 2002; Wright & Burton, 1995). By using the independent-scales methodology, the importance of such mechanisms has been highlighted yet again, suggesting that they form a central basis of the distinction between judgements of recollection and familiarity, at least in the current experimental context.

**R–K judgements and Donaldson’s (1996) two-criteria signal detection model**

Donaldson (1996; see also Hirshman & Master, 1997; Inoue & Bellezza, 1998; Xu & Bellezza, 2001) proposed that most R–K data could be accounted for with a two-criteria signal detection model with a single dimension of familiarity. Specifically, he suggested that R judgements are made when a test item possesses familiarity above a response criterion that is more conservative than the recognition (old–new) criterion. K judgements, on the other hand, are made when a test item possesses familiarity above the recognition criterion but below the more conservative R criterion, and “new” judgements are made to test items with familiarity below both criteria.

This signal detection model of R and K judgements has the clear advantage over dual-process theory in its ability to account for false R data. Such judgements in signal detection models are based on the fact that the tail of new-item distribution often falls above the R criterion. Nonetheless, such a model cannot easily explain some of our data derived from the independent-scales technique. For example, familiarity and recollection ratings in a two-criteria signal detection model would most plausibly be represented by two different criteria placed on the same underlying memory dimension, much like R judgements are thought to be represented by a more conservative criterion than are K judgements in the original model. Importantly, however, the signal-detection model predicts that discrimination should not differ between the scales because each is assumed to be measuring the same underlying strength dimension. However, the analyses on the signal detection parameters in Experiments 2A and 2B indicated that this prediction was not supported: The recollection scale yielded better discrimination than the familiarity scale.

**An attributional account**

The reliable effect of prime duration (identification) on R judgements in Experiment 1 and on recollection ratings in Experiments 2A and 2B may be surprising to some readers. Prime duration is likely to have affected the perceptual fluency of the recognition target, and Rajaram (1993) has shown that enhancing the perceptual fluency of recognition targets by preceding them with briefly presented matching versus mismatching primes resulted in more subjective familiarity (K), but not more subjective recollection (R). Furthermore, our prime duration manipulation was unlikely to have affected the conceptual or elaborative encoding of the study items, the traditional underpinning of conscious recollection processes from the dual-process perspective (e.g., Jacoby & Dallas, 1981; Mandler, 1979, 1980), nor was it likely to have affected the distinctiveness or salience of the encoded items (Rajaram, 1998). Indeed, the fact that we observed prime duration effects on false R data renders all explanations based on differential study encoding of old words logically unfeasible.
We believe the discrepancy between our results and those of Rajaram (1993), and the different pattern of results obtained with mild versus extreme manipulations of prime duration (cf. Higham & Vokey, 2000), may be best understood in terms of the attributional processes involved. We have argued that participants faced with 100% matching prime–target pairs may have been using an identification heuristic, whereby prime identification performance was used as a basis of recognition (i.e., identified = “old”; misidentified = “new”, with more items identified in the long duration condition than in the short duration condition). In contrast, participants faced with a mixture of matching and mismatching prime–target pairs, as in Rajaram’s research, may have relied more on the perceptual fluency of the target processing (i.e., fluent = “old”; nonfluent = “new”, with more targets fluently processed in the matching condition than in the mismatching condition). If, in fact, Rajaram’s participants were using the fluency heuristic, whereas our participants were relying on the identification heuristic, then there is no reason to expect that either responding, or the phenomenology associated with responding, would be the same.

By this attributional account, particular judgements (e.g., K, R, “familiar”, or “recollected”) are not directly tied to retrieval of particular traces or the activation of specific processes or systems. Rather, we believe that judgements of this sort are made in much the same way that Johnson and colleagues have suggested that people make source attributions (e.g., Johnson et al., 1993; although see below for points of departure between our and Johnson and colleagues’ viewpoints). During a memory test, participants have available to them myriad kinds of information. This information includes not just retrieved study details, but retrieved pre-experimental experiences, aspects of current processing, information regarding recent performance, and other available details. The nature of the phenomenological judgement, in our view, will depend on which components of the information complex participants “assess” to complete the task at hand and how these components are interpreted. The interpretation, in turn, may depend on lay theories about the general workings of memory and on more personal knowledge regarding the working of participants’ own memory. Which information is salient and how it is interpreted will depend on the particular task and on the test context. For example, if the task involves reality monitoring, then highly salient aspects of this information complex will include memory for perceptual details (perhaps leading to a external source attribution) and memory for cognitive operations performed on the material (perhaps leading to an internal source attribution). On the other hand, if the task involves judgements of “familiar” or “recollected”, then any source-specifying details in the information complex will be made salient and will probably lead to a recollection judgement. Similarly, our current research suggests that the lack of such details will support a “definitely not recollected” judgement.

Our account of recollection and familiarity judgements is similar to the functionalist account forwarded by Gruppuso, Lindsay, and Kelley (1997; see also Bodner & Lindsay, 2003, and Whittlesea, 2002). Like us, they argued that recollection and familiarity are not necessarily tied to the operation or activation of particular processes or systems, but what counts as recollection or familiarity depends on the situation and context. In particular:

Recollection allows one to exert control based on retrieval of memory for particular aspects or attributes of a prior event, and familiarity refers to retrieval of information that falls short of the specificity required by the task (Gruppuso et al., 1997, p. 273).
We agree wholeheartedly with the notion that participants’ definitions of recollection and familiarity are not fixed and are contextually dependent. However, in our view, Gruppuso et al.’s (1997) functionalist account, as well as the source-monitoring framework more generally, is still limited in that it, like dual-process models, assumes that recollection is based primarily on retrieval of (source-specifying) veridical study details. We concur that the presence of veridical study details in the information complex available at test will certainly be important for supporting recollection judgements, as long as participants recognize them as such and incorporate them into their judgements (i.e., monitor them correctly). However, more than this is needed. The functionalist account does not explain, for example, participants’ translation of successful test–time identification performance into attributions of recollection when that success has occurred in the context of hard-to-identify primes.

With this attribution account in mind, consider again the opposite (to the current findings) pattern of results obtained when the prime duration manipulation was made more extreme (Higham & Vokey, 2000, Exps. 2 & 3), which posed problems for fuzzy-trace theory. We argued that with an extreme manipulation of prime duration, participants no longer relied on the identification heuristic to make recognition judgements. This was so because identification performance was no longer predictive of prior presentation; for example, items were identified correctly at the very long duration (over 200 ms) regardless of whether they were old or new. Instead, participants adopted the fluency heuristic and made attributions regarding target-processing fluency to various sources. When primes were identified, a ready source of target processing fluency was made available (prime identification itself), and so misattributions of high target-processing fluency to prior presentation were less likely than if the prime was misidentified. Consequently, the effect of prime duration, seen in the current experiments, reverses with more extreme manipulations of prime duration.

The identification heuristic is probably just one strategy of many that participants use in a recognition memory experiment. Recognition memory research has been primarily focused on better understanding of participants’ use of fluency, not on discovering the various heuristics and strategies that participants might use across different experimental paradigms (although see Whittlesea, 2002; Whittlesea & Leboe, 2000, 2002, for exceptions). Nonetheless, participants’ use of the identification heuristic in the current setting, at the very least, highlights the insufficiency of veridical retrieval, source specifying or not, as the sole basis of recollection judgements.

In summary, fuzzy-trace theory with phantom recollection, two-criteria signal detection models, and particularly dual-process theory with the recollection–remember identity assumption all have difficulty accounting for the results we have obtained in this and associated research. In contrast, an attributional account, which argues that participants assess the evidence available to them at the time of the recognition response (e.g., prime identification success; target processing fluency), assess it for its predictive value, and make (mis)attributions accordingly, is able to explain the extant data generated by this paradigm.

REFERENCES


*Original manuscript received 2 August 2001
Accepted revision received 8 April 2003*
APPENDIX A

Instructions for the R–K condition of Experiment 1

During this phase, words will be flashed one at a time on the computer monitor. Once a word is flashed, your first job is first to try to identify it and write it on paper in front of you. After that, press the space bar and the word that was flashed will appear in the clear. At this point, regardless of whether you identified the word correctly or not, rate whether you think the word occurred in the training phase or not. Write either “O” (for “old”; it did occur in training) or “N” (for “new”; it did not occur in training) in the space provided next to the word.

Remember judgements

If you decide that an item is “old” (i.e., it occurred in the training list), then you will need to make a further judgment as well. If your recognition of the word is accompanied by a conscious recollection of its prior occurrence in the study list, then put down “R” (for “remember”). “Remember” is the ability to become consciously aware again of some aspect or aspects of what happened or what was experienced at the time the word was presented (e.g., aspects of the physical appearance of the word, or of something that happened in the room, or of what you were thinking and doing at the time). In other words, the “remembered” word should bring back to mind a particular association, image, or something more personal from the time of study, or something about its appearance or position (i.e., what came before or after the word).

Know judgements

“Know” responses should be made when you recognise that the word was in the list of words that you read aloud, but you cannot consciously recollect anything about its actual occurrence. In other words, put down “K” (for “know”) when you believe you recognise the word, but the word fails to evoke any specific conscious recollection from the study list.

To further clarify the difference between these two judgments, (i.e., “R” versus “K”), here are a few examples. If someone asks for your name, you would typically respond in the “know” sense, without becoming consciously aware of anything about a particular event or experience; however, when asked about the last movie you saw, you would typically respond in the “remember” sense, that is, becoming consciously aware again of some aspects of the experience. If you have any questions regarding these judgments, please ask the experimenter.

APPENDIX B

Instructions for the independent scales condition of Experiment 2A

In the test phase, you will be presented with more words, one at a time on the computer monitor. However, at first, each word will be displayed very briefly and then covered up. Your first job is to try to identify the word and to write it down on the paper provided. Even though you may not have been able to see the word, guess what it is. It is important that you write down something on every trial. To help you keep track of what trial you are on, the trial number will be displayed on the monitor. Make sure that the trial number you are at on the test sheet is the same as the number displayed on the monitor.

After you have attempted to identify the word and have written it down on the paper, press the space bar to find out if you were right. Once the word is presented in the clear, then two ratings are required. One is a recollection rating and the other is a familiarity rating.

Recollection rating

If the word is accompanied by a conscious memory of its prior occurrence in the study list, then you are recollecting it. “Recollection” is the ability to become consciously aware again of some aspect or aspects of what happened or what was experienced at the time the word was presented (e.g., aspects of the physical appearance of the
word, or of something that happened in the room, or of what you were thinking or doing at the time). In other words, the “recollected” word should bring back to mind a particular association, image, or something more personal from the time of study, or something about its appearance or position (i.e., what came before or after that word). One half of the items in the test list were presented in the study list and one half were not.

_Familiarity rating_

Sometimes you may know a word occurred in the study list because it provides a feeling of familiarity. This feeling can be thought to occur independently of recollection. A word might seem familiar whether or not you recollect anything from the time you studied it. Likewise, recollection can occur with or without a feeling of familiarity. As stated above, one half of the items in the test list were presented in the study list and one half were not.

To clarify the difference between these two ratings, consider the following examples.

**High recollection–high familiarity**

If a word evokes a feeling of familiarity, and you can recollect something about the word’s occurrence in the study list, then you should rate both recollection and familiarity as high.

**High recollection–low familiarity**

For any particular test item, you might recollect your encounter with the test word, but it does not seem familiar. For example, you might remember coughing when this word was presented earlier, but you have no feeling of familiarity associated with this memory. If this is the case, recollection should be rated high but familiarity should be rated low.

**Low recollection–high familiarity**

For any particular test item, you might have a strong feeling of familiarity associated with it, but not recollect anything about your encounter with it in the study phase. If this is the case, recollection should be rated low but familiarity should be rated high.

**Low recollection–low familiarity**

Items that evoke no feeling of familiarity or recollection should be rated low on both recollection and familiarity. Make the familiarity and recollection ratings for each word by inserting a number from 1 to 4 in the space provided where:

1 = definitely no
2 = probably no
3 = probably yes
4 = definitely yes

This rating scale is shown at the top of the first page of your answer sheets.

If you have any questions regarding these judgments, please ask the experimenter.