Memory & Cognition
2008, 36 (8), 1415-1428
doi:10.3758/MC.36.8.1415

Perceptual representations in false recognition and priming of pictures

Yana Weinstein and David R. Shanks
University College London, London, England

Using a new procedure, we investigate whether imagination can induce false memory by creating a perceptual representation. Participants studied pictures and words with and without an imagery task and at test performed both a direct recognition test and an indirect perceptual identification test on pictorial stimuli. Corrected false recognition rates were 7% for pictures studied in word form (Experiment 1), 26% for pictures imagined once (Experiment 2), and 48% for pictures imagined multiple times (Experiment 3), although on the indirect test, no priming was found for these items. Furthermore, a perceptual/conceptual imagery manipulation did not affect the tendency to claim that imagined items had been studied as pictures (Experiment 4). These results suggest that the false memories reported on direct tests are not driven by perceptual representations.

In the past three decades, research has shown that people can be induced to believe falsely that they have seen words (Roediger & McDermott, 1995) and even experienced entire events (Loftus & Pickrell, 1995). Various paradigms have been developed to study such false memories. For instance, Loftus's (1975) misinformation paradigm examines how people integrate misleading information into their memory of an event. In a typical misinformation experiment, participants watch a video of an event and then read a description of the same event that includes some inaccuracies. When tested on their memory for the event as depicted in the video clip, participants typically include inaccuracies from the written description. Also widely used is the Deese/Roediger–McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) paradigm, in which participants learn 15-word lists, each of which converges on one associate, the critical lure. For example, the following list is based around the critical lure CHAIR: *table*, *sit*, *legs*, *seat*, *couch*, *desk*, *recliner*, *sofa*, *wood*, *cushion*, *swivel*, *stool*, *sitting*, *rocking*, *bench*. At test, despite never having seen it in the study phase, participants spontaneously produce the critical lure CHAIR in free recall and/or call it "old" in old/new recognition tests. Although these paradigms have received much attention, their focus has been on false memory induction, rather than on specifying the nature of the representation that underlies the false memories.

One process found to contribute significantly to the formation of false memories involves the use of imagination. Henkel and Franklin (1998) investigated imagination inflation with a source-monitoring test. In these experiments, participants are shown drawings of various objects and are asked to imagine other objects. At test, participants are prompted with object labels and are asked whether they perceived or imagined the associated object; in other words, they are asked to identify the source of their memory for this item. One way in which participants might perform this task is to imagine (or reimagine) a picture of the object represented by the label and then compare this picture with any existing memory trace of the item. It is not too hard to see how, in this situation, participants might have trouble identifying whether the memory trace is of a self-generated image or of a picture presented previously. In Johnson, Hashtroudi, and Lindsay's (1993) source-monitoring framework, the first stage of source attribution is to decide whether a memory is of an internally generated event or of an event experienced perceptually in the outside world. Various aspects of the memory might help to distinguish between these two categories; for instance, one might examine the vividness of the image. Thus, incorrect source attribution in the case of an imagined picture could result from the incorrect use of a heuristic that attributes a vivid mental image to prior perception. Source-monitoring failure can be seen to play out in real-life situations—for instance, in the mug shot exposure effect in which eyewitnesses mistakenly pick from a lineup an innocent member whom they recognize from a mug shot rather than from the original event (see Deffenbacher, Bornstein, & Penrod, 2006, for a review)—and in autobiographical memory—for example, in a study by Garry, Manning, Loftus, and Sherman (1996), in which imagining childhood events led to participants' being more likely to claim that these imagined events had occurred.

In the present article, we further investigate the role of imagination in creating a perceptual memory trace that can be mistaken for a true memory. More compelling evidence for this effect would be a situation in which imagination produces increased false *recognition* of pictures. Thus, at test, participants see the picture itself instead of

Y. Weinstein, y.weinstein@ucl.ac.uk

an object label, thereby allowing them to compare the picture directly with any existing memory trace and to make a memory judgment on the stimulus itself. False recognition of nonverbal stimuli was first investigated by Metcalfe and Fisher (1986), who showed false recognition of prototypical dot patterns. A handful of studies have looked at false recognition of pictures (Israel & Schacter, 1997; Koutstaal & Schacter, 1997), but the stimuli have been predominantly line drawings rather than rich pictorial stimuli.

If imagination can produce representations that are similar in nature to those resulting from perception, that would make it very difficult to distinguish pictures that were perceived from pictures that were imagined (i.e., true memories from false ones). There are strong suggestions from neuroimaging work that imagination and perception overlap with respect to neural activity (Kosslyn & Thompson, 2000), and there is evidence that activation of brain regions linked to vivid imagination of objects can predict subsequent false alarms (Gonsalves et al., 2004), although object labels rather than pictures were presented in the procedure used to test this claim.

Other research into memory has used indirect, or implicit, tests, whereby prior exposure to test stimuli facilitates performance on tasks such as lexical decision (Ratcliff, Hockley, & McKoon, 1985) and picture naming (Warren & Morton, 1982). Recently, researchers have investigated whether false memory for critical lures in the DRM paradigm is exhibited on both indirect and direct memory tests. Indirect tests can be divided roughly into those that tap into the perceptual and those that tap into the conceptual details of studied stimuli (Weldon, Roediger, Beitel, & Johnston, 1995). Perceptual indirect tests, such as speeded recognition of degraded stimuli, rely on the identification of perceptual features, although conceptual indirect tests, such as word association, access an item's meaning. Generally it has been found that performance on conceptual indirect tests is facilitated for critical lures as well as for studied items (e.g., McDermott, 1997), whereas the behavior of critical lures on perceptual indirect tests is more controversial (e.g., Hicks & Starns, 2005; McBride, Coane, & Raulerson, 2006).

For perceptual indirect tests, intuition may lead one to hypothesize that the nonpresented items lack perceptual detail and thus should not facilitate performance on an indirect test that relies on perceptual processing. However, if false memories can acquire perceptual details—for instance, by borrowing them from true memories (e.g., Lampinen, Meier, Arnal, & Leding, 2005)—a contrasting hypothesis would be that activation of related items could lead to facilitation (priming) by some mechanism similar to perception. Research in the lexical domain on this topic has produced mixed results, with some studies finding no evidence of facilitation of critical lures on perceptual indirect tests such as lexical decision (e.g., McKone, 2004), and others suggesting that critical lures behave similarly to studied items on indirect tests (e.g., Hancock, Hicks, Marsh, & Ritschel, 2003). Similar to the imagination studies described above, this avenue of research speaks to the nature of the memory trace that leads people to report false memories. So far, indirect memory tests with pictorial stimuli have not been used to study false memory. In order to investigate whether perceptual processes drive false memory for pictures, in the present study, we apply the indirect test methodology to pictures.

With the present experiments, we propose a new procedure for studying false memory for pictures resulting from imagination. In every experiment, participants study pictures and are presented with pictures in the final test. In Experiment 1, we establish a basic false recognition effect: Pictures that appear as words in the study phase are recognized falsely more often than are new pictures. Using a version of the continuous identification procedure (Feustel, Shiffrin, & Salasoo, 1983; Stark & McClelland, 2000) as an indirect memory test, we also verify that studying words in the absence of imagination does not lead to any priming of pictures represented by these words. In other words, there is no cross-form priming on this perceptual identification task. The absence of such priming provides evidence that the task, which has not been used previously for pictures, is highly perceptual in nature (see Experiment 1, Discussion). In Experiments 2 and 3, we obtain increasingly high levels of false recognition of pictures following imagination, and we compare this with performance on the perceptual identification task. To preview the main conclusions, we find no evidence of priming on this test for imagined pictures, which suggests that imagination does not lead to the formation of perceptually rich false memories. To support this claim, we show in Experiment 4 that asking participants to imagine sentences instead of pictures does not reduce the false recognition effect.

EXPERIMENT 1

In the study phase, we presented participants with pictures and words. Then we gave them a combined recognition and perceptual identification test. Typically, it is found that previously perceived pictures are recognized faster than are new pictures (Warren & Morton, 1982), so we expected that studied pictures would be recognized significantly faster than new ones. The transfer-appropriate processing framework (Blaxton, 1989) proposes that performance on perceptual indirect tests should be sensitive to changes in presentation modality. If priming is a function of the match between study and test, a change in form (e.g., verbal to pictorial) should eliminate priming on perceptual indirect tests. Thus, we expected to find a higher rate of false recognition for those pictures that had appeared previously as words than for new pictures, but no perceptual priming.

Method

Participants

Twenty undergraduate students 18–21 years of age from University College London volunteered to participate in return for course credit.

Materials

A set of 110 pictures was created for this and subsequent experiments. Full-color photographs were edited in Adobe Photoshop to produce pictures in five categories: animals, clothes, electrical appliances, fruit and vegetables, and household objects. (see the Appendix for the full list of items). Black-and-white versions of the pictures were produced using the grayscale function (see Figure 1 for an example). Each picture depicted a single object (e.g., a dog, a T-shirt, a washing machine) presented on a 300 × 300 pixel white background with shadows, text, and other extraneous details removed. To obtain agreed-upon labels for the pictures, we asked 4 participants to name a larger pool of pictures, half in color and half in black and white; this was counterbalanced among participants. The final set of 110 pictures included only those that elicited the same response from at least 3 of the 4 participants. For the priming test, two versions of a mask were created by randomly filling a 300 × 300 pixel grid with 36 fragments taken from pictures not shown at test (see Figure 1). All experiments were programmed in Visual Basic 6.0, and a PC with a screen resolution of 1,024 × 768 pixels was used throughout.

Design

The experiment was a repeated measures design with two independent within-subjects variables, item type (studied picture, studied word, new) and test (recognition, priming), and two dependent measures, proportion judged as *old* in the recognition test and reaction time to identify pictures in the priming test.

Critical Items

The 110 items were divided into three groups of 30 critical items, which were rotated through the item–type conditions, and 20 filler items, which appeared in the study phase only. Three counterbalancing conditions were required. Each study list consisted of 30 words and 30 pictures that appeared later in the test and an additional 10 filler items in each presentation format. The test phase consisted of 30 items that had been studied as pictures, 30 items that had been studied as words, and 30 new items. In both the study and test phases, half of the items were in color and half were in black and white. Although the black-and-white/color manipulation was not a factor considered in the present experiment, it played a role in subsequent experiments and so was introduced for consistency.

Procedure

This experiment consisted of two phases: a study phase in which participants were presented with both pictures and words, and a test phase in which participants performed perceptual identification followed by recognition of pictures on each trial.

In the study phase, participants were instructed to study a list of 40 pictures and 40 words carefully. Picture and word trials were randomly intermixed. Picture trials consisted of (1) a 1,000-msec blank interval, (2) a fixation cross appearing at the center of the screen for 1,000 msec, (3) the picture presented at the center of the screen for 250 msec, and (4) a 1,000-msec blank interval. Word trials consisted of a 1,000-msec blank interval followed by the word presented at the center of the screen for 4,000 msec. These timings were used to avoid ceiling effects in true recognition of pictures. Following a 5-min mental arithmetic distractor task, participants were given instructions for the test phase.

In the test phase, for each trial, participants first identified a picture by using the continuous identification procedure described below. Following its identification, the picture was displayed again, and participants made a recognition judgment in their own time. Pictures were presented in the same central location as during the study phase. Previously encountered pictures appeared in the same color scheme as in the study phase, whereas new pictures appeared in black and white or color according to the counterbalancing condition, so that half of the items seen at test were in color and half were in black and white.

The picture identification portion of the test phase was based on the continuous identification (CID) procedure employed by Stark and

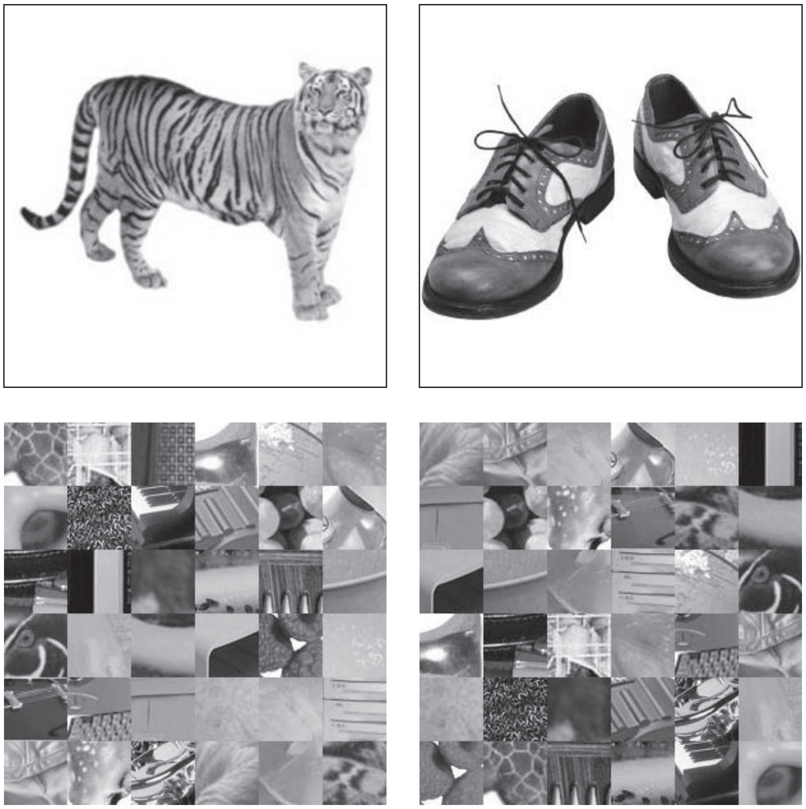

Figure 1. Sample picture stimuli and masks used in the perceptual identification task.

McClelland (2000) and originally formulated for use with words by Feustel et al. (1983). In the present experiment, the procedure was redesigned for pictures. The instructions stated that this part of the experiment was concerned with visual perception and that we were interested in the time it took people to identify pictures accurately. Participants were told that on each trial they would see a picture presented on the screen for a very brief duration, followed immediately by a mask. They were told that the picture would then reappear and continue to flash on the screen for longer and longer durations, making it appear clearer over time. The task was to try to identify the picture as quickly as possible. As soon as participants could make out the picture, they were to press ENTER and, when prompted, to type in the name of the picture. They were informed that, when they pressed ENTER, the mask would reappear and they would not see that picture again until they had entered a response. Participants were told to answer as fast as possible, but to do so only as soon as they were confident that they would identify the picture correctly. When the ENTER key was pressed, RT measurement stopped, and when ENTER was pressed again to complete the response, the recognition portion of the trial was initiated.

We used ExacTicks code for Visual Basic to program the timing of the procedure. Each trial began with a 500-msec presentation of one of the two masks (see the Materials section and Figure 1). The two masks were rotated randomly among stimuli and always appeared in the same color scheme as the corresponding stimulus. Following initial presentation of the mask, the stimulus was presented for 17 msec (the duration of one screen refresh), and the mask followed for 233 msec, making a 250-msec block. This stimulus presentation was repeated, with the duration of the stimulus increasing by 17 msec on every presentation, while total block time remained constant.

Upon identification of a picture, participants were shown the picture again and were asked to judge whether it was *old* or *new*. Recognition judgments were made on a 6-point confidence scale where 1 = *very sure new* and 6 = *very sure old*. Participants were instructed specifically to use the *old* response only if they had seen that picture in the study phase. Instructions stated that, if participants remembered seeing the word in the study phase, the correct response was *new*. Participants made the recognition judgment in their own time while viewing the picture. Before beginning the test, participants performed a practice block of seven identification trials, although no recognition judgments were required. No feedback was given to participants about their performance during the test.

Results

Test Phase

Recognition. Recognition responses given on the 6-point scale were collapsed into *old* responses (ratings of 4–6) and *new* responses (ratings of 1–3). Figure 2 presents these data for Experiments 1–3. A within-subjects ANOVA showed a significant effect of item type (studied picture, studied word, new) on recognition [$F(1.44, 27.44) = 240.08$, $MS_e = 0.02$, $p < .001$, $\eta_p^2 = .93$]. Mauchly's test indicated that the assumption of sphericity had been violated for this main effect [$\chi^2(2) = 8.75$, $p < .05$]; the Greenhouse–Geisser correction ($\varepsilon = .72$) was used. The proportion of hits to pictures that had appeared in the study phase ($M = .86$) as well as false alarms to pictures that had appeared as words in the study phase ($M = .21$) were both significantly higher than the proportion of false alarms to new pictures ($M = .14$) [$t(19) = 17.52$, $SEM = .04$, $p < .001$, $r = .94$; and $t(19) = 3.09$, $SEM = .02$, $p < .01$, $r = .24$, respectively].

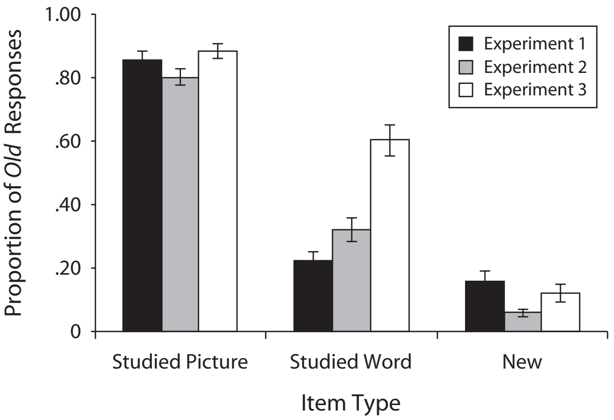

Figure 2. Experiments 1–3: Mean proportions of *old* responses by item type on the direct test. The error bars represent standard errors of the means in all figures.

The same pattern of recognition data emerged when the high-confidence (confidence level 6, *very sure old*) responses were analyzed separately: A repeated measures ANOVA showed a significant effect of item type [$F(1.15, 21.85) = 145.07$, $MS_e = 0.03$, $p < .001$, $\eta_p^2 = .88$]. Mauchly's test indicated that the assumption of sphericity had been violated [$\chi^2(2) = 24.19$, $p < .001$]; the Greenhouse–Geisser correction ($\varepsilon = .58$) was used. High-confidence hits were made at a rate of .68, whereas high-confidence false alarms to studied word items were made at a mean rate of .07, which was significantly higher than the false alarm rate of .04 to new items [$t(19) = 2.15$, $SEM = .02$, $p < .05$; $r = .19$]; only 7 participants made any high-confidence false alarms to new items, although 13 of the 20 participants made high-confidence false alarms to studied word items. The average confidence of *old* responses was compared among the three conditions for the 10 participants who made at least four *old* responses in each category. A within-subjects ANOVA confirmed that confidence levels differed between conditions [$F(2,18) = 25.40$, $MS_e = 0.09$, $p < .001$, $\eta_p^2 = .74$]. Whereas hits to studied picture items were made with higher confidence ($M = 5.69$) than were false alarms to new items ($M = 4.78$) [$t(9) = 5.68$, $SEM = .16$, $p < .001$; $r = .71$], confidence levels did not differ significantly between new-item and studied-word-item false alarms ($M = 4.98$) [$t(9) = 1.86$, $SEM = 0.11$, $p = .10$; $r = .18$].

In all experiments, we looked at recognition sensitivity and bias for two comparisons: studied picture versus studied word items and studied picture versus new items. The sensitivity results are shown in Table 1. For all sensitivity analyses, d' and C were calculated from hits and false alarm rates. To accommodate hit rates of 1 and false alarm rates of 0, we added 0.5 to each data point and divided the result by $n+1$, where n is the number of trials of that type (Snodgrass & Corwin, 1988).[1] A repeated measures ANOVA confirmed that discriminability was lower for studied picture items when compared with studied word items than for studied picture items when compared with new items [$F(1,19) = 7.85$, $MS_e = 0.13$, $p < .05$, $\eta_p^2 = .29$]. Participants appear to be relatively unbiased in the studied-picture/new-item comparison, and significantly more liberal in the studied-picture/studied-word comparison [$F(1,19) = 7.98$, $MS_e = 0.03$, $p < .05$, $\eta_p^2 = .30$].

Table 1
Measures of Sensitivity and Bias in Experiments 1–4

	Studied Pictures/Studied Words				Studied Pictures/New			
	Sensitivity (d')		Bias (C)		Sensitivity (d')		Bias (C)	
Experiment	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
1	2.00	.17	−.17	.09	2.32	.20	−.01	.10
2	1.41	.14	−.16	.07	2.25	.10	.25	.05
3	0.86	.16	−.71	.09	2.28	.14	−.00	.07
4	1.87	.13	.25	.07	2.06	.13	.35	.06

Priming. Individuals' median reaction times (RTs) by item type were used in the analyses. RTs were excluded for responses that did not match the picture label or any of its synonyms. Identification accuracy was high (.94) across all conditions; a within-subjects ANOVA indicated that accuracy differed between conditions [$F(2,38) = 4.55$, $MS_e = .01$, $p < .05$, $\eta_p^2 = .19$]. Participants were more accurate in identifying studied picture items (.96) than in identifying new items (.94) [$t(19) = 2.40$, $SEM = .01$, $p < .05$; $r = .29$] and equally accurate in the studied word (.93) and new ($p = .25$) conditions. Of the 1,800 trials across all participants, 103 were excluded due to inaccurate identification. RTs also were excluded on the 7 trials where the RT exceeded the length of the CID cycle (3,500 msec). Overall, these exclusion criteria resulted in a loss of 6.1% of the data. Figure 3 shows median RTs in milliseconds by item type for Experiments 1–3. Priming is defined by a decrease in RT relative to baseline performance on new items. Item type (studied picture, studied word, new) had a significant effect on RTs [$F(2,38) = 11.38$, $MS_e = 10{,}087$, $p < .001$, $\eta_p^2 = .38$]. However, although priming of studied pictures was significant [$t(19) = 4.13$, $SEM = 33$, $p < .01$; $r = .19$], there was no significant priming of pictures that had appeared as words in the study phase ($t < 1$).[2] Furthermore, only 9 of 20 participants showed the effect (i.e., they were on average faster in responding to studied word items than in responding to new items, allowing for a 17-msec measurement error, the equivalent of one screen refresh). A paired-samples *t* test confirmed that priming differed significantly between the studied picture and studied word conditions [$t(19) = 3.84$, $SEM = 33$, $p = .001$; $r = .41$] (see Table 2).

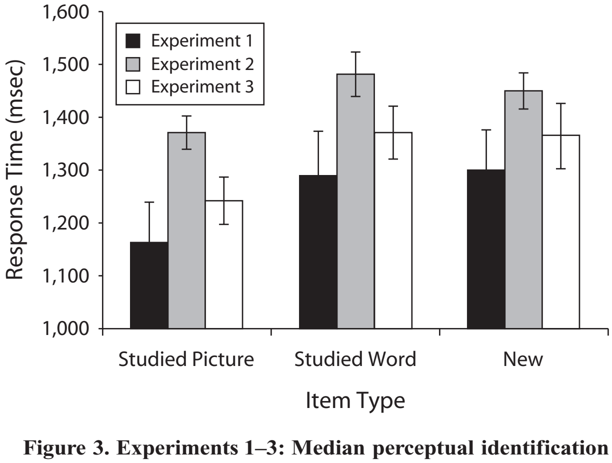

Figure 3. Experiments 1–3: Median perceptual identification response times by item type on the indirect test.

Discussion

In this experiment, pictures that had been studied were recognized at a high rate in the direct test and were identified faster than new pictures on the indirect test. Participants also made more false alarms to pictures that had appeared in the study phase as words than to new pictures. However, pictures that had appeared in the study phase as words were identified no faster on the indirect test than were new pictures. Although most studies have found that a change from verbal stimuli in the study phase to pictorial stimuli in the test phase eliminates perceptual priming completely (Hirshman, Snodgrass, Mindes, & Feenan, 1990; Warren & Morton, 1982; Weldon et al., 1995), some other studies (Durso & Johnson, 1979; Park & Gabrieli, 1995; Srinivas, 1993) have found significant cross-form priming, suggesting that certain perceptual priming tasks can include a nonperceptual component. In Experiment 1 we have demonstrated that our implicit task is sensitive to variations in visual features between study and test, shows no detectable cross-form priming, and hence is perceptually driven (Jacoby & Hayman, 1987).

In Experiments 2 and 3, we modified the procedure in an attempt to increase perceptual activation of the studied word items. The new procedure consisted of a study phase involving only pictures, the same test procedure used in Experiment 1, and an intervening imagery phase in which participants were instructed to form mental images of objects in response to words. Most of the words appearing in this phase corresponded to previously studied pictures, but 10 critical items were replaced with new words. This false memory induction was expected to produce high levels of false alarms to items that had appeared only as words in the imagery phase, as compared with new items from the same semantic categories. If creating a mental image has an effect on perceptual processing equivalent to (if lower than) that of actually seeing a picture, we would expect to find perceptual priming of items that had been studied in word form and imagined.

To some extent, whether participants show priming after imagination depends on the perceptual overlap between the images they create and the test stimuli. Biederman's (1987) recognition-by-components theory proposes that priming relies on the overlap of geometric components (*geons*) between study and test. According to this theory,

Table 2
Comparison of Direct and Indirect Tests in Experiments 1–3

	Direct Test: Proportion "Old" Responses		Indirect Test: Priming Effect (msec)	
Experiment	True Memory	False Memory	True Memory	False Memory
1	.71	.07	136	10
2	.74	.26	75	−37
3	.76	.48	123	−6

perceptual priming would be preserved despite differences between study and test stimuli, as long as these changes do not affect the structure of the objects (Biederman & Cooper, 1991). Manipulations of object size (Biederman & Cooper, 1992) and viewing angle (Biederman & Gerhardstein, 1993) have confirmed that this is the case. More relevant to our research, Bartram (1974) showed that naming an object facilitates naming another object (or token) corresponding to the same label (or type), though to a lesser extent than does naming the same object. A representation created through imagery can be seen as another token of the type embodied by the associated object label. The few studies that have looked specifically at whether imagery can lead to perceptual priming have produced mixed results. McDermott and Roediger (1994) found priming in a perceptual task that involved the identification of pictures from fragments presented for 100/200 msec following imagination of these pictures from their word labels. On the other hand, on an object identification task, Michelon and Koenig (2002, Experiment 2) found significant priming only after perception and not after imagery. In Experiments 2 and 3, we were interested in whether false recognition of pictures could be in part attributed to perceptual overlap between the test stimulus and a representation that has been formed as a result of an experience other than perception. If this is so, we would expect participants to show both false recognition and perceptual priming of studied word items following imagination.

EXPERIMENT 2

Method

Participants

Fifty-one 1st-year undergraduates (6 males, 45 females) 18–29 years of age ($M = 18.8$, $SD = 1.8$) participated in the experiment as part of a course requirement.

Materials

The set of 110 pictures described in Experiment 1 was divided into five 22-item categories: animals, clothes, electrical appliances, fruit and vegetables, and household objects (see the Appendix for the full list of items). As far as possible, the selection of items in each list was based on semantic category norms (Van Overschelde, Rawson, & Dunlosky, 2004), but to some extent was constrained by the availability of appropriate pictorial stimuli. To achieve the required list length, Van Overschelde et al.'s separate *fruit* and *vegetable* categories were merged into one list, *kitchen utensils* and *carpenter's tools* were combined to create the *household objects* list, and a list of *electrical appliances* was compiled. Categories did not overlap; for instance, none of the household objects could be powered by electricity. An additional 14-item *furniture* category was created for use in the practice phase. Ten items from other categories were used as filler items in the test phase.

On the basis of semantic category norms, six critical items were chosen from each of the four categories for which norms were available. In all cases, the item was one of the top seven most popular items in each category. For the new category of electrical appliances, norms were produced by asking 20 people to list six exemplars; the six most popular exemplars were used as critical items for the electrical appliances category.

Design

With respect to manipulated variables, the experiment was identical to Experiment 1: A repeated measures design was used with two independent within-subjects variables, item type (studied picture, studied word, new) and test (recognition, priming), and two dependent measures, proportion judged as *old* in the recognition test and reaction time to identify pictures in the priming test.

Critical Items

Of the 110 pictures, 30 were critical items (6 from each of the five semantic categories). There were three groups of critical items: 10 studied pictures, 10 studied words, and 10 new items. These item labels were adopted for easy comparison with Experiment 1, although in Experiments 2 and 3, the studied word items also were imagined. In the present experiment, *studied pictures* were those that were shown to participants as pictures during the study phase and as words in the imagery phase, *studied words* were those items that appeared only as words in the study phase and once again as words in the imagery phase, and *new items* did not appear in either phase. The three sets of 10 critical items (2 from each of the five semantic categories) were rotated among three counterbalancing conditions (studied picture, studied word, and new) and were constructed in such a way that mean identification rates of pictures in each group on the perceptual identification task (see the Test Phase section below) were roughly similar. We collected identification rates by pretesting the test phase alone on 6 participants. Apart from the critical items, each condition thus had a total of 104 items in common: 14 furniture pictures for the practice phase, 10 filler items for the test phase, and 16 each of animals, clothes, electrical appliances, fruit and vegetables, and household objects. In each of these conditions, half of the items appeared in color and half appeared in black and white in both the study and test phases. Three additional conditions were created to counterbalance this, producing six counterbalancing conditions in total.

Procedure

Experiments 2 and 3 involved three phases: a study phase in which participants were presented with pictures, an imagery phase in which they were presented with words referring to these pictures along with some additional words, and a test phase identical to that in Experiment 1, in which participants performed picture perceptual identification followed by recognition on each trial.

In the study phase, participants were shown 90 pictures, 80 that were common to all conditions and 10 from the critical items list (studied pictures in the test). The task was to choose the correct word for each picture from a choice of two. The two words appeared on the screen first for 4,000 msec, one on each side of the screen. During this time, participants were asked to imagine pictures for each of these words. They were not given any specific instructions for this imagery task (e.g., whether to imagine the pictures in black and white or color). Next, a picture corresponding to one of the words appeared in the center of the screen for 250 msec, and then the words disappeared. Following this brief presentation, participants had to make a choice by pressing either the backslash key "\" to select the left word or the forward-slash key "/" to select the right word. As a reminder, these symbols appeared in place of the words when it was time to make a choice. As soon as participants made their selection, a low-pitched beep sounded in the case of an incorrect response, and the next trial began immediately.

The word corresponding to the picture appeared as one of the choices on each trial. The alternative word corresponded to a different picture from the same category. For example, participants may have been shown the words *cat* and *dog* followed by a picture of a dog. The incorrect word always corresponded to another picture that participants would see during the experiment. The order of trials was randomized across all categories, and the program randomly chose word pairs with the constraints above.

Participants read on-screen instructions and completed a practice block of 14 items, after which they were told their score before the real experiment began. Of the 90 pictures shown to participants, 45 appeared in color and 45 appeared in black and white, but there was nothing in the instructions relating to this. There was also no mention of a memory test, and participants were not asked explicitly to remember the pictures or anything about them.

In the imagery phase, after a 5-min mental arithmetic distractor task, participants were shown 90 words and for each word were asked to indicate whether they had seen the corresponding picture in black and white or color during the first part of the experiment. This task was used in order to encourage participants to create and examine a mental image of the picture. Each word appeared in the center of the screen for 4,000 msec, during which time participants were asked to form a mental image of the corresponding picture in order to answer the question. As soon as the word disappeared, participants made their choices using the same keys as in the study phase. Participants were asked to use the number keypad to rate how confident they were in their answer following each response on a 5-point scale, where 1 = *not at all confident* and 5 = *extremely confident*.

The words that appeared in this phase referred to 70 of the 80 pictures seen by all participants in the study phase (the same 10 pictures were excluded for all participants), and the items that later appeared as studied pictures (10 words) and studied words (10 words) in the test phase. Thus, 80 of the words corresponded to 80 of the pictures seen by participants in the study phase, whereas the other 10 had not been encountered previously. The instructions stated that the words in this phase corresponded to pictures seen in the previous part of the experiment; in other words, participants were not informed that some of the words had not appeared as pictures. However, participants were instructed to try to form mental images for any words they could not remember having seen as pictures in the study phase and make their choices on the basis of those mental images in the absence of a memory. The format of this phase was similar to that of the study phase, with instructions appearing on-screen. There was no practice block and no feedback in response to participants' answers.

In the test phase, the format and instructions were identical to those of Experiment 1: Participants identified a picture on each trial and then indicated whether this picture was *old* (i.e., whether they had seen the picture in the study phase) or *new* on a confidence scale from 1 to 6, where 1 = *very sure new* and 6 = *very sure old*. Participants were instructed specifically to use the *old* response only when they had seen the picture itself earlier in the experiment. All participants were tested on the same 40 items. Ten of these were filler items from novel categories, and the other 30 were critical items: 10 studied pictures, 10 studied words, and 10 new items. Therefore, of the 40 pictures, participants had seen only 10 during the study phase. As in Experiment 1, studied pictures appeared in the same color scheme as in the study phase, and participants completed seven practice identification trials prior to the test.

Results

Study Phase

Performance in the study phase ranged from 88.9% to 100% correct identification of pictures from a choice of two words, with a mean score of 97.5%. In other words, over half of the participants made fewer than 3 identification errors in the 90 trials, and no one made more than 10 errors.

Imagery Phase

Performance in the imagery phase was based on the 80 items that had appeared in the study phase; the 10 studied word items were excluded from the analysis because no correct response was possible. In this task, participants had to indicate whether they had seen each picture in black and white or color in the study phase. Scores ranged from 52.5% to 83.8%, with a mean score of 65.7% (SD = 8.2), which was significantly different from the chance level of 50% [$t(50) = 13.61, p < .001$]. On the 5-point confidence scale, participants were significantly more confident on their answers to the 80 items that they had actually seen as pictures (M = 3.41) than on their answers to the 10 studied word items for which there was no correct response, because these items had not been seen (M = 2.00) [$t(50)$ = 12.81, SEM = .11, $p < .001$; r = .67]. The fact that performance on this task was consistently above chance indicates that participants were paying attention to the task and following instructions.

Test Phase

Recognition. For simplicity, responses given on the 6-point scale were collapsed into *old* responses (items rated 4–6) and *new* responses (items rated 1–3). Figure 2 shows the proportion of *old* responses to studied pictures, studied words, and new items. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of the within-subjects variable, item type [$\chi^2(2) = 13.9, p < .01$]; a Greenhouse–Geisser estimate of sphericity was used to correct degrees of freedom (ε = .80). A repeated measures ANOVA showed that there was a significant effect of item type [$F(1.61, 80.2) = 233.35$, $MS_e = 0.04, p < .001, \eta_p^2 = .82$]. Therefore, the likelihood of participants' responding *old* to an item depended on its status. The proportions of hits to pictures that had appeared in the study phase (M = .80) and of false alarms to pictures that had appeared as words in the imagery phase (M = .32) were both significantly higher than the proportion of false alarms to new pictures (M = .06) [$t(50)$ = 28.41, SEM = .03, $p < .001$, r = .94; and $t(50)$ = 7.41, SEM = .04, $p < .001$, r = .55, respectively], and 40 of the 51 participants showed the latter effect (i.e., their studied word false alarm rate was at least one item higher than their new item false alarm rate).

Looking at high-confidence responses only, a repeated measures ANOVA showed a significant effect of item type [$F(1.47, 73.47) = 207.87$, $MS_e = 0.13$, $p < .001$, $\eta_p^2 = .81$]. Mauchly's test indicated that the assumption of sphericity had been violated [$\chi^2(2) = 21.95$, $p < .001$]; the Greenhouse–Geisser correction ($\varepsilon = .74$) was used. High-confidence hits to studied picture items were made at a rate of .57, whereas high-confidence false alarms to studied word items were made at a mean rate of .09, significantly higher than the false alarm rate of .03 to new items [$t(50) = 3.68$, $SEM = .02$, $p = .001$; $r = .31$]; only 11 participants made any high-confidence false alarms to new items, whereas 26 of the 51 participants made high-confidence false alarms to studied word items. Because of the low number of false alarms to new items, average confidence of *old* responses was compared between the studied picture and studied word conditions only for the 23 participants who made at least four *old* responses in each of the two categories. A paired-smaple *t* test confirmed that hits to studied picture items were made with higher confidence than were false alarms to studied word items ($M = 5.60$ and $M = 4.90$, respectively) [$t(22) = 9.28$, $SEM = 0.08$, $p < .001$, $r = .61$].

A repeated measures ANOVA confirmed that discriminability was lower for studied-picture when compared with studied-word items than for studied-picture when compared with new items [$F(1,50) = 62.72$, $MS_e = 0.43$, $p < .001$, $\eta_p^2 = .56$]. Participants tended to respond conservatively in the studied-picture–new-item comparison, and more liberally in the studied-picture–studied-word comparison [$F(1,50) = 62.82$, $MS_e = 0.11$, $p < .001$, $\eta_p^2 = .56$] (see Table 1).

Priming. The same exclusion criteria were applied to RTs as those used in Experiment 1. Identification accuracy was high (.95) and did not differ between conditions ($p > .4$). Of the 1,530 trials across all participants, 70 trials were excluded due to inaccurate identification, and 5 trials were excluded because the RT exceeded the length of the CID cycle (3,500 msec). Overall, these exclusion criteria resulted in a loss of 4.9% of the data. Figure 3 shows median RTs in milliseconds by item type. A repeated measures ANOVA performed on studied-picture/studied-word/new RTs produced a significant effect of the within-subjects variable, item type, on perceptual identification [$F(2,100) = 9.12$, $MS_e = 17{,}977$, $p < .001$, $\eta_p^2 = .15$]. Thus, RTs were dependent on the item's status. Planned comparisons revealed that, whereas studied picture items were recognized significantly faster than new items [$t(50) = 3.03$, $SEM = 25$, $p < .01$; $r = .16$], there was no significant difference in RTs between studied words and new items [$t(50) = -1.18$, $SEM = 29$, $p = .21$; $r = -.07$].[3] In fact, the numerical difference was in the direction opposite to that predicted, and only 20 of the 51 participants showed the pattern of results that would be indicative of priming (as in Experiment 1, the criterion was a minimum difference in RTs of 17 msec). A paired-samples *t* test confirmed that priming differed significantly between the studied picture and studied word conditions [$t(50) = 4.26$, $SEM = 26$, $p < .001$; $r = .28$] (see Table 2).

Discussion

This experiment, unlike Experiment 1, included a false memory induction with a perceptual component (imagery). If this perceptual component increases false alarms to studied word items, we might have predicted that it would also lead to priming of these items on the perceptual identification test. However, no such priming was found. The absence of priming for pictures of imagined objects on a perceptual identification task is in line with the Michelon and Koenig (2002) finding of no perceptual priming following imagination, the Hicks and Starns (2005) finding of no priming for critical lures in the DRM paradigm on a verbal perceptual identification task, and the studies that showed no priming for critical lures on a lexical decision task (McKone, 2004; Zeelenberg & Pecher, 2002). The findings are also in line with studies that have found no cross-form priming from words to pictures (Hirshman et al., 1990; Warren & Morton, 1982; Weldon et al., 1995). However, results of Experiment 2 contradict McDermott and Roediger (1994), who found priming in a perceptual task that involved the identification of pictures from fragments presented for 100/200 msec following imagination of these pictures from their word labels.

EXPERIMENT 3

It is possible that the priming effect is weaker than the false recognition because priming responses are intrinsically noisier than direct measures (Berry, Henson, & Shanks, 2006; Buchner & Brandt, 2003). In this experiment, we attempted to make the false memory induction even stronger in order to create optimal conditions for detecting a priming effect. Goff and Roediger (1998) examined false recognition of imagined actions and found that increasing the number of times participants imagined performing an action led to an increased tendency to report having performed the action. Following this rationale, we increased the number of times participants were exposed to, and were asked to imagine, the studied word items from one to four repetitions.

Method

Participants

Twenty-four volunteers (8 males) 18–43 years of age ($M = 23.0$) were recruited from within University College London to participate in the experiment and were reimbursed for their time.

Procedure

The procedure was almost identical to that of Experiment 2, except for the number of times participants encountered the critical studied word items, which was increased from once in Experiment 2 (in the imagery phase only) to four times in the present experiment (once in the imagery phase and three times in the study phase). In order to do this, we altered the word pool from which the incorrect words were drawn in the study phase labeling task. In Experiment 2, the incorrect word always corresponded to another picture participants would see during the study phase. In the present experiment, 12 of the 18 trials in each semantic category followed this design. For the other six trials in each category, however, incorrect words corresponded to the two critical studied word items that were not presented as pictures in the study phase but appeared as words in the imagery phase. This was done to increase activation of these items in an attempt to increase false memory for these items at test. The imagery phase was identical to that of Experiment 2. Finally, the test phase was also identical to that of Experiment 2, except that, on the direct test, participants simply had to indicate whether a picture was *old* or *new* instead of making a 1–6 confidence rating.

Results

Study and Imagery Phases

In the study phase, participants identified images from a choice of two words with a mean correct score of 97.0% (SD = 4.65). Performance in the imagery phase, where participants reported the color format of pictures seen in the study phase, ranged from 55.0% to 83.4%, with a mean score of 66.7% (SD = 8.9), which was significantly above the chance level of 50% [$t(23) = 9.16$, $p < .001$]. As in Experiment 2, in the imagery phase, participants were significantly more confident on the 5-point scale in their answers to the 80 pictures from the study phase (M = 3.56) than to the 10 studied word items (M = 2.59) [$t(23) = 7.01$, $SEM = .14$, $p < .001$; $r = .59$]. In a between-experiments comparison of Experiments 2 and 3, although participants were equally confident in their answers to items from the study phase ($p = .39$), those in the present experiment were more confident in their answers to the studied word items [$t(73) = 2.99$, $SEM = .20$, $p < .01$; $r = .35$]. This difference can be attributed to the fact that participants in the present experiment had encountered these novel items previously as words in the study phase. Aside from this, there were no significant differences in performance in any of the measures above between Experiment 3 and Experiment 2.

Test Phase

Recognition. Figure 2 shows the proportion of *old* responses to the three item types. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of the within-subjects variable, item type [$\chi^2(2) = 6.85$, $p < .05$]; a Greenhouse–Geisser estimate of sphericity was used to correct degrees of freedom (ε = .79). A repeated measures ANOVA showed that there was a significant effect of item type [$F(1.58, 36.3) = 166.01$, $MS_e = 0.03$, $p < .001$, $\eta_p^2 = .88$]. The proportions of hits to pictures that had appeared in the study phase (M = .88) and of false alarms to pictures that had appeared as words in the study and imagery phases (M = .60) were both significantly higher than the proportion of false alarms to new pictures (M = .12) [$t(23) = 22.82$, $SEM = .03$, $p < .001$, $r = .95$; and $t(23) = 12.15$, $SEM = .04$, $p < .001$, $r = .78$, respectively], and all 24 participants showed the latter effect. A repeated measures ANOVA confirmed that discriminability was lower for studied picture items when compared with studied word items than for studied picture items when compared with new items [$F(1,23) = 168.70$, $MS_e = 0.14$, $p < .001$, $\eta_p^2 = .88$]. Participants tended to respond conservatively in the studied-picture–new-item comparison, and more liberally in the studied-picture–studied-word comparison [$F(1,23) = 168.69$, $MS_e = 0.04$, $p < .001$, $\eta_p^2 = .88$] (see Table 1).

Priming. Individuals' median RTs by item type were analyzed with the same exclusion criteria as in Experiments 1 and 2. Identification accuracy was very high (.97) and did not differ between conditions; of the 720 trials across all participants, 30 trials (4.2%) were excluded due to inaccurate identification. Figure 3 shows median RTs in milliseconds by item type. A repeated measures ANOVA performed on studied-picture/studied-word/new item RTs produced a significant effect of the within-subjects variable, item type [$F(2,46) = 7.16$, $MS_e = 17{,}660$, $p < .01$, $\eta_p^2 = .24$]. Planned comparisons revealed that whereas studied picture items were recognized significantly faster than new items [$t(23) = 3.08$, $SEM = 40$, $p < .01$; $r = .23$], with 21 of 24 participants showing the effect, there was no significant difference in RTs between studied word and new items ($t < 1$);[4] once again, the numerical effect was in the direction opposite to that expected if performance on the indirect test mirrored that of the direct test, and only 11 of the 24 participants showed the effect. A paired-samples t test confirmed that priming differed significantly between the studied picture and studied word conditions [$t(23) = 4.03$, $SEM = 32$, $p < .001$; $r = .30$] (see Table 2).

Discussion

In Experiments 1–3, we showed a consistent false memory effect for studied word items in the direct test and a consistent lack of priming of these items on the perceptual identification test. Figure 2 shows that although hits to studied picture items and false alarms to new items remained stable between the three experiments, the false alarm rate to studied word items depended on the procedure and, consistent with Goff and Roediger's (1998) imagination inflation study, was highest when participants were asked to imagine items four times. This false alarm rate was always significantly above that of new items. Figure 3, on the other hand, shows that the procedure did not affect how fast studied word items were identified relative to new items; studied picture items were always identified significantly faster than new items, whereas studied word items were identified at roughly the same speed as new items. Note that participants performed the identification task at different speeds in the three experiments. Experiment 1 yielded the fastest overall RTs, and this is most probably a result of participants' speeding up across the 90 items in the test phase; in Experiments 2 and 3, the test phase consisted of only 40 items. Participants also performed the task faster in Experiment 3 than in Experiment 2, perhaps because the recognition decision made on each trial (*old*/*new* as opposed to a 1–6 confidence judgment) required less time and effort and thus interfered less with the identification task. The important result, however, is that despite these variations in speed, the pattern of RTs was consistent across all three experiments. Table 2 presents a comparison of performance on the direct and indirect tests in Experiments 1–3. It is clear from these data that although false recognition of studied word items increased with imagery instructions, no manipulation produced any perceptual priming of these items.

The lack of priming for studied word items on the perceptual identification task in Experiments 2 and 3 suggests that there is little perceptual overlap between the representations of falsely recognized pictures and the actual stimuli. Nonetheless, false recognition for these items is high following one imagination event (Experiment 2) and even higher following multiple imagination events (Experiment 3). Of course, we cannot be certain that participants did in fact follow instructions to imagine objects. However, the imagery task fulfilled its function of inflating the false alarm rate on the direct test by increasing participants' interaction with visual representations of the objects; whatever the specific mental processes participants engage in during imagery, the conclusion that false recognition on the direct task is unaccompanied by perceptual priming remains unaffected. On the other hand, the pattern of results does not preclude the possibility that participants do create perceptually rich representations during imagery, but does suggest that something other than perceptual overlap between internal representations and test stimuli must be driving the effect. In fact, any mismatch between the two ought to make it easier for participants to reject unstudied items. If this were the case, then we would expect a lower rate of false recognition when objects were imagined multiple times, strengthening the internal representations and making more salient any differences between them and the presented pictures. Instead, we see an increase in the false alarm rate, suggesting that nonperceptual processes are responsible for the effect.

EXPERIMENT 4

In Experiments 1–3, we showed that studying a word and imagining an object can lead to increased false alarms to the associated picture when it is presented in a direct recognition test in the absence of the analogous effect on an indirect perceptual identification test. One criticism of these experiments is that, in the recognition test, participants could be simply responding *old* to pictures when they had studied the associated word, either because they are misunderstanding the instructions or because they are relying on familiarity and endorsing any familiar item as *old*. A source-monitoring test would give participants the option to indicate that they previously had encountered a word without the associated picture. Given this additional response option, any false alarms to studied word items above the new item base rate could be characterized as faulty discrimination between imagination and memory, as suggested by the source-monitoring framework (Johnson et al., 1993). The first aim of Experiment 4 was thus to extend the procedure used in Experiments 2 and 3 to accommodate a more conservative direct memory task.

The second aim of Experiment 4 was to further address the issue of the nature of false memory representations. Studying words and imagining associated objects (Experiments 2 and 3) led to much greater corrected false alarm rates than simply studying words (Experiment 1). One way in which this imagination process might increase false memory is by creating a perceptual representation that at test is mistaken for a memory of having perceived the item. However, the lack of priming for studied word items in all three experiments suggests that this representation is not matched closely to the actual stimulus in terms of perceptual features. Nonperceptual activation of the critical studied word items could instead be responsible for the increased level of false alarms to these items. Thus, the results so far seem to imply that there is little perceptual processing involved.

To further test this possibility, we compared two conditions in which everything except imagery-phase instructions was held constant. In the perceptual condition, participants imagined pictures in response to words, whereas in the conceptual condition, participants imagined sentences in response to the same words. If perceptual activation plays a role in the false recognition effect observed so far, participants in the perceptual condition should claim to have perceived more of the pictures that they only studied as words than participants in the conceptual condition.

Method

Participants

Thirty-five undergraduate students 18–21 years of age from University College London took part in this study and were reimbursed for their time; 21 were assigned to the perceptual condition, and 14 were assigned to the conceptual condition.

Design and Procedure

The materials, design, and procedure were based on those of Experiment 2; the differences are detailed below. The experiment was a mixed design, with item type (studied picture, studied word, new) as the within-subjects variable and imagery (perceptual, conceptual) as the between-subjects variable. All participants performed the same source-monitoring test, but, unlike in the previous experiments, there was no perceptual identification component.

The study phase was identical to that of Experiment 2. Although greater false recognition rates were achieved in Experiment 3, participants encountered studied word items in the study phase where they were asked to form images in response to word labels, and this would have interfered with the imagery manipulation in the present experiment.

The general format of the imagery phase, including presentation times, was similar to those of Experiments 2 and 3: Participants made rating judgments on words, which were presented for 4,000 msec each. As in Experiments 2 and 3, participants were informed that the words that they were about to see related to the pictures that they had seen earlier. In the perceptual condition, participants were asked to imagine the picture that the word referred to, and to rate the pleasantness of this image. In the conceptual condition, participants were asked to think of a sentence using the word and to rate the pleasantness of the sentence. Aside from these instructions, there was no difference between conditions. In both conditions, participants made responses on the same scale, from 1 = *not at all pleasant* to 5 = *extremely pleasant*.

The test phase did not differ between conditions, and test items were identical to those in Experiments 2 and 3. Participants were shown the 40 pictures one by one and were asked to make a source judgment. They were instructed to press the key labeled "P" if they had studied the picture and its label, "L" if they had studied only its label, or "N" if they had not encountered it previously in the experiment. Participants made judgments in their own time while viewing the picture, with no practice phase or feedback.

Results

Participants in the two conditions did not differ significantly on any measures of performance in the study and imagery phases. Thus, the data summarized below represent the performance of all 35 participants.

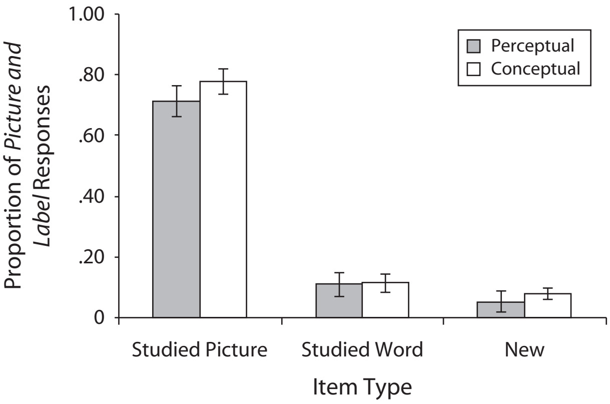

Figure 4. Experiment 4: *Picture and label* responses by item type following perceptual and conceptual imagery.

Study and Imagery Phases

Accuracy in the study phase was high as in the previous experiments, with a mean score of 97.0% across the two imagery conditions; there were no differences in this measure between conditions ($p > .6$). Pleasantness ratings in the imagery task were almost identical across the two imagery conditions ($M = 3.18$, $p > .9$) and were not affected by whether participants had seen pictures referring to the words in the study phase ($p > .5$).

Test Phase

Figure 4 shows the proportion of *picture and label* responses to items in the source-monitoring test by imagery-condition. We conducted an overall analysis on these responses using a mixed ANOVA, with item type (studied picture, studied word, new) as the within-subjects variable and imagery (perceptual, conceptual) as the between-subjects variable. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of the within-subjects variable, item type [$\chi^2(2) = 16.06$, $p < .001$]; a Greenhouse–Geisser estimate of sphericity was used to correct degrees of freedom ($\varepsilon = .72$). Across the two conditions, the likelihood of participants' responding *picture and label* to an item was dependent on its status [$F(1.43, 47.33) = 269.38$, $MS_e = 0.03$, $p < .001$, $\eta_p^2 = .89$]. Both the proportion of hits to pictures that had appeared in the study phase ($M = .74$) and the proportion of false alarms to pictures that had appeared as words in the imagery phase ($M = .11$) were significantly higher than the proportion of false alarms to new pictures ($M = .06$) [$t(34) = 18.46$, $SEM = .04$, $p < .001$, $r = .91$; and $t(34) = 2.51$, $SEM = .02$, $p < .05$, $r = .19$, respectively]. A repeated measures ANOVA confirmed that discriminability was lower for studied picture items when compared with studied word items than for studied picture items when compared with new items [$F(1,34) = 7.61$, $MS_e = 0.09$, $p < .01$, $\eta_p^2 = .18$]. Given the additional option to indicate that they had studied the label only, participants responded conservatively in both comparisons, although significantly more so in the studied-picture–new-item comparison [$F(1,34) = 7.60$, $MS_e = 0.02$, $p < .01$, $\eta_p^2 = .18$] (see Table 1). The imagery manipulation did not have an effect on responses, and there was no interaction between imagery and item type (Fs < 1). In other words, whether participants imagined pictures (perceptual) or sentences (conceptual) in the imagery phase did not affect the proportion of false alarms they made to these items on the picture source-monitoring test.

Discussion

Experiment 4 fulfilled two aims: It extended our false memory induction to include a more conservative direct test and provided further evidence that the effect is not driven by perceptual processes. With respect to the first aim, we can now assert that the false memory for studied word items in our procedure cannot be attributed simply to a relaxed criterion or misunderstanding of instructions. When given the option to indicate that an item had been studied only in word form, participants still claimed to have perceived significantly more of the pictures that had appeared in word form compared with new items from the same categories.

In the source-monitoring test, participants presumably attempt to match the attributes of the stimulus to their memory of the item and indicate that they had previously perceived the picture if it matches a memory trace. There are thus two possible reasons why participants may make false alarms in this case. One is that the false memory is a result of conceptual activation of the item that would arise from studying a word. Another is that, in the perceptual condition, the image participants create in response to a particular word is sufficiently similar to the picture to produce a strong match between the stimulus and the image stored in memory, leading participants to believe that they had perceived the picture. Evidence for the former comes from the fact that participants made no fewer *picture and label* responses to studied word items when they were instructed to imagine sentences instead of pictures, thus eliminating the perceptual component of the task. This result also ties in neatly with the lack of perceptual priming obtained for studied word items in Experiments 2 and 3.

GENERAL DISCUSSION

The present article makes three contributions, two of which are methodological. First, we developed a new false memory induction procedure that elicits high levels of false recognition of pictorial stimuli following imagination (Experiments 2 and 3). Second, we examined false memory for pictorial stimuli on an indirect test (Experiments 1–3), a technique previously employed only with word stimuli. Finally, we made use of these methodological advances to answer our theoretical question: whether false memories that arise from imagination are predominantly perceptual or conceptual in nature.

In Experiments 1–3, participants demonstrated a clear tendency to endorse pictures that had been studied in word form as *old* more frequently than they did new items. The relatively low baseline false alarm rate indicates that participants had no problem distinguishing between items they had studied and new items from the same category. The elevated false recognition rate for the critical studied word items was thus specifically related to studying the item in its verbal form (Experiment 1) and imagining it pictorially (Experiments 2 and 3). In Experiments 2 and 3, we presented a procedure that produced corrected false recognition rates of up to 48% (Experiment 3) for rich pictorial stimuli presented only 15 min earlier. Unlike most previous studies looking into false recognition at short delays (e.g., Israel & Schacter, 1997), we used rich pictorial stimuli and always presented the pictures themselves as retrieval cues. The distinctiveness of the pictures should have aided participants in rejecting unseen stimuli, and yet we achieved higher corrected false recognition rates than those in studies that used line drawings with the DRM paradigm (Israel & Schacter, 1997) and with the category associates procedure that used a 3-day delay between study and test (Koutstaal & Schacter, 1997). Furthermore, we have shown that the effect of imagery on false recognition is not driven by criterion shifts: Participants showed a reduced ability to discriminate between studied picture and studied word items following imagination in Experiment 2, despite responding with the same criterion as in Experiment 1 (see Table 1).

Experiment 4 also showed that false recognition of studied word items remained higher than false recognition of new items, even when the response options at test allowed participants to distinguish between pictures that had actually been perceived and those that had been presented only in word form. In Experiments 1–3, participants may have made false alarms to studied word items because they misinterpreted the *old* response option on the recognition test and called pictures *old* if they remembered having studied the associated word. The results of Experiment 4 discredit the hypothesis that participants always are aware of not having studied the pictures. In other words, there is evidence to suggest that in at least some cases, when shown a picture that had appeared only as a word, participants do have a false memory of the picture itself. However, at least some of the effect is likely to be driven by diffuse familiarity, as evidenced by the difference in false alarms to studied word items in Experiment 4 and Experiment 2, which is comparable in terms of encoding.

McKone (2004) has suggested that perceptual indirect tests can help distinguish between true and false memories; more specifically, priming on indirect memory tests that are thought to be driven largely by perceptual processing should occur only for items that actually are studied. In Experiments 1–3, we add further support to such a view: Priming on a perceptual identification test was found only for pictures that were actually studied, but not for pictures that had appeared in word form (Experiment 1) or for pictures that were imagined once (Experiment 2) or imagined multiple times (Experiment 3). We showed this absence of priming in the presence of an increasingly large false recognition effect. This suggests that the representations of the falsely recognized items are dissimilar in nature to those of truly perceived items. Such an interpretation stems from the view that the direct and indirect tests tap a common memory representation. Of course, another possibility is that they are based on distinct memory systems (Squire, 1994).

The perceptual/conceptual dichotomy can help characterize the representations of true and false memories. One hypothesis is that a false memory of a nonpresented item, much like a true memory of an item that has actually been perceived, is based on a memory trace that contains perceptual features that partially match those of the test stimulus. This view is supported by McDermott and Roediger's (1994) finding of perceptual priming of pictures following imagination, although this study was not designed specifically to examine false memory. The opposite hypothesis is that the relevant concept is activated, but that no perceptual trace exists. Evidence for this view comes from Michelon and Koenig (2002), who found no perceptual priming of pictures following imagination, studies that have reported no priming of critical lures on perceptual indirect tests (Hicks & Starns, 2005; McKone, 2004), and studies that have found that participants recall more distinctive, item-specific perceptual details in relation to true memories than to false memories (e.g., Norman & Schacter, 1997). A false memory of a nonpresented item may also contain perceptual features that differ from those of the actual stimuli. This mismatch ought to serve as a cue to reject unstudied pictures; in the present experiments, however, the more times participants encountered object names, the more false alarms they made (Experiment 3 compared with Experiment 2), rendering this explanation unlikely.

In our experiments, recognition can be seen as a conceptual task, and identification can be seen as a data-driven or perceptual task. This distinction is based on the finding that changes in perceptual features affect priming but not recognition (Jacoby & Dallas, 1981). According to this reasoning, the absence of priming for studied word items in our experiments would indicate that the underlying memory that leads participants to report having perceived a picture when actually they had studied only the associated word is conceptual in nature. In other words, there is a memory of having studied a concept, but there is no memory of having encoded a picture. Henkel and Franklin (1998) addressed this issue with an imagination inflation study. Participants imagined objects that were related either perceptually or conceptually to other objects that they were shown as pictures, although, unlike in our procedure, the test involved making a source judgment in response to the object labels rather than to pictures of the objects themselves. Perceptually related items were physically similar in form (e.g., *magnifying glass* and *lollipop*), whereas conceptually related items came from the same functional category (e.g., *apple* and *banana*). Henkel and Franklin (Experiment 2) found that conceptual and perceptual relatedness of perceived and imagined items contributed equally to inflating incorrect claims of having perceived items that had been imagined. The results of our Experiment 4 accord nicely with this finding: When participants were asked to imagine sentences in response to word labels (conceptual imagery), at test, they incorrectly claimed to have studied the associated picture as often as when they were asked to imagine pictures in response to the word labels (perceptual imagery).

The present article is the first to have directly applied an indirect test to the study of false memory for stimuli other than words. There is scope for more work within this framework; it is possible that, although studying words and imagining objects does not lead to perceptual priming of associated pictures, performance on other, more conceptual priming tasks could be facilitated. For instance, although participants may not recognize the physical properties of a picture imagined previously any faster than those of a new picture, they may be faster at processing the meaning of the picture, just as they would be if they had studied the picture itself. Further research with the new procedure we propose could help clarify the similarities and differences between true and false memories.

AUTHOR NOTE

This research was supported by BBSRC Studentship 07322/-101970 awarded to Y.W. Correspondence concerning this article should be addressed to Y. Weinstein, Division of Psychology and Language Sciences, University College London, Gower Street, London WC1E 6BT, England (e-mail: y.weinstein@ucl.ac.uk).

REFERENCES

Bartram, D. J. (1974). The role of visual and semantic codes in object naming. *Cognitive Psychology*, **6**, 325-356.

Berry, C. J., Henson, R. N. A., & Shanks, D. R. (2006). On the relationship between repetition priming and recognition memory: Insights from a computational model. *Journal of Memory & Language*, **55**, 515-533.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, **20**, 585-593.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 121-133.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 1162-1182.

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 657-668.

Buchner, A., & Brandt, M. (2003). Further evidence for systematic reliability differences between explicit and implicit memory tests. *Quarterly Journal of Experimental Psychology*, **56A**, 193-209.

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

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law & Human Behavior*, **30**, 287-307.

Durso, F. T., & Johnson, M. K. (1979). Facilitation in naming and categorizing repeated pictures and words. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 449-459.

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General*, **112**, 309-346.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, **3**, 208-214.

Goff, L. M., & Roediger, H. L., III (1998). Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory & Cognition*, **26**, 20-33.

Gonsalves, B., Reber, P. J., Gitelman, D. R., Parrish, T. B., Mesulam, M.-M., & Paller, K. A. (2004). Neural evidence that vivid imagining can lead to false remembering. *Psychological Science*, **15**, 655-660.

Hancock, T. W., Hicks, J. L., Marsh, R. L., & Ritschel, L. (2003). Measuring the activation level of critical lures in the Deese–Roediger–McDermott paradigm. *American Journal of Psychology*, **116**, 1-14.

Henkel, L. A., & Franklin, N. (1998). Reality monitoring of physically similar and conceptually related objects. *Memory & Cognition*, **26**, 659-673.

Hicks, J. L., & Starns, J. J. (2005). False memories lack perceptual detail: Evidence from implicit word-stem completion and perceptual identification tests. *Journal of Memory & Language*, **52**, 309-321.

Hirshman, E., Snodgrass, J. G., Mindes, J., & Feenan, K. (1990). Conceptual priming in fragment completion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 634-647.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, **4**, 577-581.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **3**, 306-340.

Jacoby, L. L., & Hayman, C. A. G. (1987). Specific visual transfer in word identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 456-463.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.

Kosslyn, S. M., & Thompson, W. L. (2000). Shared mechanisms in visual imagery and visual perception: Insights from cognitive neuroscience. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 975-985). Cambridge, MA: MIT Press.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory & Language*, **37**, 555-583.

Lampinen, J. M., Meier, C. R., Arnal, J. D., & Leding, J. K. (2005). Compelling untruths: Content borrowing and vivid false memories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 954-963.

Loftus, E. F. (1975). Leading questions and eyewitness report. *Cognitive Psychology*, **7**, 560-572.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, **25**, 720-725.

McBride, D. M., Coane, J. H., & Raulerson, B. A., III (2006). An investigation of false memory in perceptual implicit tasks. *Acta Psychologica*, **123**, 240-260.

McDermott, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin & Review*, **4**, 582-586.

McDermott, K. B., & Roediger, H. L., III (1994). Effects of imagery on perceptual implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1379-1390.

McKone, E. (2004). Distinguishing true from false memories via lexical decision as a perceptual implicit test. *Australian Journal of Psychology*, **56**, 42-49.

Metcalfe, J., & Fisher, R. P. (1986). The relation between recognition memory and classification learning. *Memory & Cognition*, **14**, 164-173.

Michelon, P., & Koenig, O. (2002). On the relationship between visual imagery and visual perception: Evidence from priming studies. *European Journal of Cognitive Psychology*, **14**, 161-184.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, **25**, 838-848.

Park, S. M., & Gabrieli, J. D. E. (1995). Perceptual and nonperceptual components of implicit memory for pictures. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1583-1594.

Ratcliff, R., Hockley, W., & McKoon, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. *Journal of Experimental Psychology: General*, **114**, 435-450.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.

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

Squire, L. R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 203-231). Cambridge, MA: MIT Press.

Srinivas, K. (1993). Perceptual specificity in nonverbal priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 582-602.

Stark, C. E. L., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 945-972.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory & Language*, **50**, 289-335.

Warren, C., & Morton, J. (1982). The effects of priming on picture recognition. *British Journal of Psychology*, **73**, 117-129.

Weldon, M. S., Roediger, H. L., III, Beitel, D. A., & Johnston, T. R. (1995). Perceptual and conceptual processes in implicit and explicit tests with picture fragment and word fragment cues. *Journal of Memory & Language*, **34**, 268-285.

Zeelenberg, R., & Pecher, D. (2002). False memories and lexical decision: Even twelve primes do not cause long-term semantic priming. *Acta Psychologica*, **109**, 269-284.

NOTES

1. Sensitivity and bias were calculated from hit and false alarm rates rather than from the raw confidence data, due to the low number of high-confidence false alarms to new items.

2. The same analysis was carried out on the subset of studied word and new items to which participants made false alarms. In order to obtain a reliable measure of RT, only the 9 participants for whom at least four false alarm RT observations were available in both the studied word and new item conditions were included in the analysis. Participants were not reliably faster at identifying studied word items (M = 1,173 msec) than at identifying new items (M = 1,196 msec) ($t < .5$).

3. It was not possible to look at the subset of studied word and new items that were accompanied by an old response, due to the low rate of false alarms to new items. Instead, RTs of studied word items were compared with the new item RT baseline. The 20 participants for whom at least four false alarm RT observations were available in the studied word condition were included in the analysis. The effect was nonsignificant and in the direction opposite to that predicted, with studied word items (M = 1,552) identified slower than new items (M = 1,482) [$t(19) = -1.60$, $p = .13$].

4. As for Experiment 2, false alarms to new items were rare, so RTs of studied word items were compared with the new item RT baseline. The 18 participants for whom at least four false alarm RT observations were available in the studied word condition were not reliably faster at identifying those items (M = 1,401 msec) than at identifying new items (M = 1,420 msec) ($t < .5$).

APPENDIX
List of Items by Category

Animals	Clothes	Electrical Appliances	Fruit/ Vegetables	Household Objects
tiger	*shirt*	*computer*	*carrot*	*screwdriver*
horse	*shorts*	*tv*	*apple*	*drill*
bear	*trousers*	*toaster*	*lettuce*	*spoon*
cow	bag	dishwasher	aubergine	binoculars
deer	belt	fridge	cherry	bottle
duck	bow tie	iron	grapes	bowl
eagle	bra	laptop	lemon	comb
flamingo	scarf	printer	raspberry	globe
shark	tie	radio	strawberry	jug
snake	watch	sewing machine	sweetcorn	pen
tortoise	zip	vacuum	tomato	umbrella
cat	*shoes*	*hairdryer*	*orange*	*knife*
dog	*socks*	*washing machine*	*broccoli*	*pan*
lion	*hat*	*kettle*	*banana*	*hammer*
camel	coat	camera	avocado	candle
elephant	dress	clock	courgette	cup
frog	gloves	iPod	kiwi	glasses
giraffe	sandals	lamp	nectarine	hanger
kangaroo	skirt	microwave	pear	lightbulb
peacock	slippers	phone	pepper	rolling pin
rabbit	suit	photocopier	pineapple	scissors
wolf	t-shirt	razor	potato	stapler

Note—Italics denote critical items.

(Manuscript received March 3, 2008;
revision accepted for publication July 25, 2008.)