Journal of Memory and Language 63 (2010) 400–415

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Unintended effects of memory on decision making: A breakdown in access control

Michael S. Humphreys [a,*], Jason M. Tangen [a], T. Bettina Cornwell [b], Emerald A. Quinn [a], Krista L. Murray [a]

[a] School of Psychology, The University of Queensland, Australia
[b] Sport Management Program, Division of Kinesiology, University of Michigan, United States

ARTICLE INFO

Article history:
Received 13 April 2010
revision received 3 June 2010
Available online 17 July 2010

Keywords:
Memory
Decision making
Source monitoring
Access control
Evaluative conditioning

ABSTRACT

A hybrid evaluative-conditioning/source-monitoring paradigm is put forward as an alternative to the standard evaluative-conditioning paradigm. The first experiment paired brand names with a small number of attractive or unattractive female faces and used a likeability rating task as well as a source monitoring task. The second experiment paired words which differed along a masculine–feminine dimension with male and female faces, and used a speeded judgment about whether words were stereotypically masculine or feminine. The third experiment paired words that differed along an active–inactive dimension with male and female faces and used a variation of the Implicit Association Test where judgments about whether words were active or inactive were mixed with judgments about whether faces were male or female. In all three experiments, we observed transfer between the recently acquired information and the judgment task. In addition, the three experiments progressively reduce the probability of demand characteristics. We explain the results in this paradigm, and in many other paradigms, as a breakdown in access control. We also point to several similarities between existing theories of evaluative conditioning and memory phenomena/theories that have gone unnoticed in the evaluative conditioning literature.

Introduction

There has been a long history of attempts to show that attitudes can be changed by the repeated pairing of a neutral stimulus with valenced stimuli. For example, Staats and Staats (1958) paired each of two national names with either positively or negatively valenced words. After repeated pairings, the national name paired with the positive words was rated more favorably than the national name paired with the negative words. Similarly, Levey and Martin (1975; also see Baeyens, Eelen, & Van den Bergh, 1990) paired neutral pictures with valenced pictures.

Work of this type has been conceptualized as involving classical conditioning or implicit learning/memory and it has had a major impact on thinking about attitude change in social psychology and marketing (Allen & Janiszewski, 1989; Cacioppo, Marshell-Goodell, Tassinary, & Petty, 1992; Lee, 2002; Olson & Fazio, 2001; Olson & Fazio, 2006; Shimp, Stuart, & Engle, 1991). Work to understand the phenomenon continues as new marketing techniques based on repeated pairings, such as corporate sponsorship of sport (Speed & Thompson, 2000) and brand placement in movies, proliferate (Yang & Roskos-Ewoldsen, 2007); and as predispositions built on repeated exposures such as racial prejudice persist (Olson & Fazio, 2006). There is no question that such exposures change affective ratings, but there is considerable debate over whether this is a direct transfer of the valence or a demand characteristic (Field, 2000, p. 1974; Shanks & St. John, 1994).

* Corresponding author. Address: Centre for Human Factors, University of Queensland, St. Lucia, Queensland 4072, Australia. Fax: +61 07 3365 3542.
E-mail address: mh@humanfactors.uq.edu.au (M.S. Humphreys).

0749-596X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2010.06.006

In a simple evaluative conditioning experiment, two neutral stimuli are used: one is paired with a small set of positively valenced stimuli, and the other is paired with a small set of negatively valenced stimuli. Because only 2 cues are used, it would generally be quite easy for someone to become aware of the relationship between the cues and their paired stimuli or to recover some or all of this information at test. Similarly, in discrimination learning, very few people fail immediately to learn that one class of stimuli results in an unpleasant event, such as an electric shock or puff of air to the eye, while the other does not. For this reason, a rather considerable effort is often made to "complicate" the learning situation to prevent participants from becoming aware of these contingencies (Lovibond & Shanks, 2002; Olson & Fazio, 2001).

In a further extension of the traditional evaluative-conditioning paradigm, Olson and Fazio (2006) set out to change racial prejudice against African Americans with repeated conditioned stimulus–unconditioned stimulus (CS–US) pairings of eight different photographs of black individuals with positively valenced words and images, and eight different photographs of white individuals with negatively valenced words and images. During the learning phase, participants were required to press a button whenever a pre-specified (neutral filler) target appeared. Interspersed during the exposure were the focal black-positive and white-negative CS–US pairings. In order to prevent awareness of this contingency, the task was embedded among several hundred images and words that appeared simultaneously and sequentially in various spatial and temporal locations over six blocks, and the photographs of the black and white individuals were engaged in various occupational roles. Olson and Fazio used multiple CSs to determine whether conditioning at the level of the 16 photograph exemplars would transfer to the attitudes of their participants at the level of the category. At test, participants were presented with a subset of four black and white photographs, each paired with two positive and two negative stimuli, and they were asked to estimate whether the two images appeared together during the learning phase. They found that photos paired with same-valence stimuli (i.e., white-negative and black-positive) were not endorsed more often than photos paired with different-valence stimuli (i.e., white-positive and black-negative). In a second experiment, they used the same conditioning procedure along with a priming measure of racial attitudes where photos of new individuals were presented just prior to presenting positively or negatively valenced adjectives. Over the course of the test, each photo was paired with two positively valenced adjectives and two negatively valenced adjectives. The participants were instructed to attend to the photo in anticipation of a subsequent recognition test and to categorize each adjective as either positive or negative. There was a significant priming effect consistent with a direct transfer of valence to previously unseen photographs of black and white individuals.

The results of Olson and Fazio (2006) are impressive. There are, however, some fundamental problems in determining whether there is ever chance performance on an explicit test and greater than chance performance on an attitude change test. With respect to the Olson and Fazio procedure, one wonders whether the pair recognition task used to assess awareness was sensitive enough to show an effect given the complicated procedures and the interference that would have been caused by repeating the photos four times during the test. More generally, Mitchell, De Houwer, and Lovibond (2009) have argued that the procedure of using a post experimental questionnaire or interview to assess contingency awareness is flawed. They note that it is easy to find evidence for learning without awareness simply by using a noisy measure, and point to the general difficulty in accepting the null hypothesis that participants are unaware.

There is also a problem with inferring that someone was aware, while they were taking the attitude change task, from an answer obtained on a post experimental questionnaire. The alternative is that one or more of the questions on the questionnaire triggered memory retrieval of a specific learning instance, a thought or reaction to a specific instance, or a more generalized memory (Humphreys, Murray, & Maguire, 2009). Because the specific cues provided by the questions are not necessarily present during the attitude change task, it is possible that the memories that are retrieved will differ in the two situations.

Another issue that needs addressing is the question as to why a few minutes of laboratory training should change long standing attitudes. That is, why should a few pairings of black individuals with positive stimuli change "racial prejudice" in the Olson and Fazio (2006) procedure? Our starting point here is the assumption that we are dealing with associative learning/memory. As Jacoby, Shimizu, Daniels, and Rhodes (2005) noted, memory can be controlled in two ways: (1) memory output can be examined to determine if it is appropriate, or (2) retrieval can be constrained so that only sought after material comes to mind (also see Humphreys et al., 2003). It seems possible that the intrusion of recent memories into an evaluative judgment can be seen as a breakdown in access control. That is, the participant is trying to retrieve personal memories/reactions, but without deliberate intent – some of the recent learning (or some aspects of the recent learning) intrude, thereby altering the judgment. If correct, this would be important because we know something about the conditions in which a breakdown in access control is likely to occur and something about its properties.

Our first point about access control is the observation that cross talk between two different memory retrieval processes can be observed when one has to switch between performing one access process and the other. For example, Nelson, McEvoy, and Friedrich (1982) investigated associative interference by placing both an associate of a cue and a rhyme of a cue in a study list. The cue itself was not studied. The list that contained both a rhyme and an associate of the cue, produced interference when participants were required to randomly switch between using one of the cues to recall an associate and using another one of the cues to recall a rhyme. Interference was not observed, however, when participants used the cues to recall associates only, or if they used the cues to recall rhymes only. Logan and Delheimer (2001) provide another example of cross talk between two different retrieval processes when the participant is required to perform them in quick succession. These results seem to be particularly pertinent to the question as to why evidence for evaluative conditioning can be found in priming paradigms (Olson & Fazio, 2006). In these paradigms, the retrieval process that is required for the second cue (i.e., decide whether this word or picture is good or bad) may induce participants to retrieve good or bad memories to the first cue. A very similar explanation can be applied to the Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998). In this procedure, the participant is required to rapidly switch between performing two different retrieval operations (e.g., classifying pictures of faces as black or white and classifying words/pictures as good or bad).

The second example of the loss of access control shows that we do not always retrieve an entire episode. Instead, with the right cues, a portion of one episode can emerge in the memory for another episode. Thus, if a neutral stimulus has been paired with a positively evaluated picture, then under some conditions, the memory for that pairing may not be retrieved. Instead, a memory for a thought or reaction that one had to the pairing might be retrieved, and this retrieval would not necessarily be recognizable as being of episodic or recent origin. Humphreys and Cornwell et al. (2010, Experiment 2) looked at the effect of providing information about an ambushing attempt and/or counter-ambushing information on memory for the sponsor of the event. Ambushing includes the use of phrases and images associated with an event or activity within the event broadcast or presence in and around the venue, where a non-sponsoring company associates itself with the event, without paying for sponsorship rights. In this experiment (see Fig. 1 for the design of the two between-subject conditions, which are relevant to this discussion), participants first received 16 messages about real companies sponsoring fictitious events. For one group, each of the 16 sponsorship messages was followed on the same day by either an ambushing message linking a competitor of the sponsor to the event or an unrelated filler message. The next day, they received a different message depending on what they received on Day 1. If on Day 1, they had received an ambushing message, the Day 2 message was either a counter-ambushing message, announcing that a competitor was attempting to ambush the event, or an unrelated filler message. Similarly, if on Day 1 they received a filler message, then the Day 2 message was either a counter-ambushing message or another filler message. Following the sponsorship announcements, the second group received either a counter-ambushing message or a filler message corresponding to each of the sponsorship announcements. Then on Day 2, if they had received a counter-ambushing message on Day 1, the Day 2 message was either an ambushing message or a filler message. Similarly, if they had received a filler message on Day 1, then the Day 2 message was either a counter-ambushing message or a filler message. Finally on Day 2, both groups were given the 16 events as cues and asked to recall the sponsor. The ambushing message linked a competitor of the sponsor to the event, but was ambiguous about whether or not the competitor was a sponsor of the event. If, when cued with the event, a participant recalled an ambushing message, they may well have been confused and assumed that the ambusher was in fact the sponsor. The counter-ambushing message was a message from the event organizers. It named the competitor and the event and explicitly stated that the competitor was attempting to ambush the event. If, when cued with the event, a participant recalled the entire counter-ambushing message or the central theme of the counter-ambushing message (the “not sponsor” information), they would have known that the ambusher was not the sponsor. In this situation then, if the counter-ambushing message reduces sponsor recall and/or increases the recall of the ambusher, then we can be sure that the entire episode and/or its central theme is not being retrieved.

There was a small reduction in sponsor recall and a major increase in intrusions of the ambusher when the counter-ambushing message occurs on Day 2 (it follows either an ambushing or a filler message on Day 1) than when it occurs on Day 1 (it is followed by either a filler or ambushing message on Day 2). Because of recency, participants should have been more capable of recalling either the entire counter-ambushing message or the “not sponsor” information provided in the counter-ambushing message when it occurred on Day 2 than when it occurred on Day 1. Instead, it appears that the linking of the competitor to the event, which occurred in the counter-ambushing message, was having an effect on intrusions in the absence of a memory for the “not sponsor” information that was the central theme of the counter-ambushing message. Humphreys and Cornwell et al. (2010) argued that this occurred because the cues that were being used (the event, the Day 1 context, the concept of a brand name) were more likely to cue the name of the competitor than they were to retrieve the “not sponsor” information.

When a participant has strong pre-existing associations with the cues used in an evaluative-conditioning paradigm (e.g., the national names used by Staats & Staats, 1958), it looks like an episodic memory (what has been learned during the evaluative conditioning procedure) is intruding into a semantic memory (the pre-existing associations or knowledge about the cues). Access control also breaks down in these situations as episodic memories acquired over multiple occasions can intrude in semantic tasks. For example, Humphreys et al. (2009) had participants learn multiple 4-pair lists. Each list was constructed from a set of 6 cues. Across the short lists, each cue was paired with two different targets, where the high frequency target occurred twice as often as the low frequency target. Each of the short lists was followed by the presentation of 2 cues. Initially, participants were simply asked to recall the target paired with that cue in the last list. In later trials, participants were sometimes asked to produce the first of the list targets that came to mind. As a final task, participants were given a free association task where they were given a cue and asked to respond with the first word that came to mind. In this task, the first 6 cues were new and the next 12 contained 6 new cues as well as the 6 cues that had been used to construct the study lists. Participants provided normal associative responses to the new cues (e.g., to “salt”, “pepper” might be supplied). However, 51% and 19% of the responses to the 6 old cues were respectively, the dominant and non-dominant targets from the

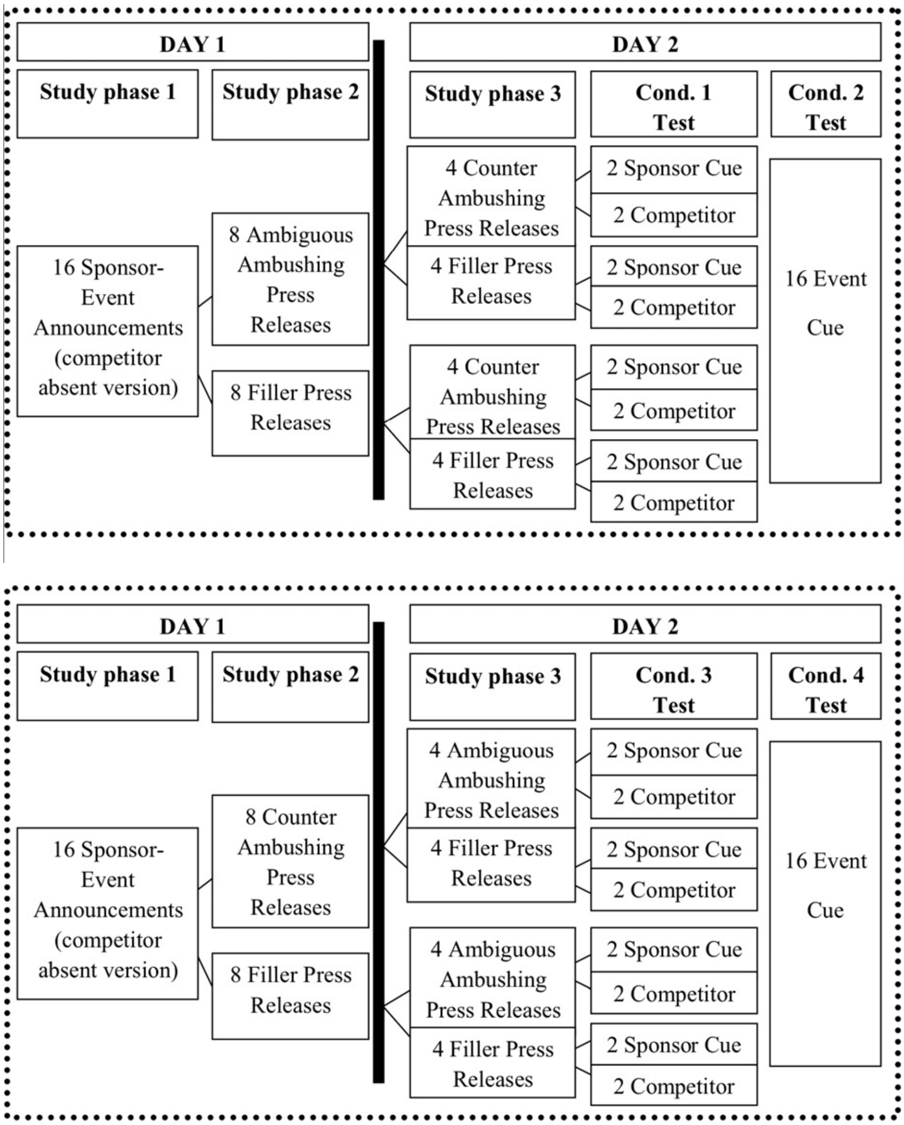

Fig. 1. Design of Humphreys and Cornwell et al. (2010, Experiment 2).

experiment rather than well established word associates. Furthermore, three of the participants who were graduate students and friends of the experimenter spontaneously reported that they were surprised when the first item they thought of on the free association task was one of the unrelated targets that had been paired with that cue during the learning phase. Thus, under these conditions, there was transfer from recent learning to a test of semantic knowledge (free association). The participants would have been aware of the intrusion of the list targets in the free association task, but based on their spontaneous reports, it appears that this intrusion was unbidden (non-deliberate). In addition, the procedures in Humphreys et al. (2009) share some characteristics with procedures commonly used in evaluative-conditioning paradigms. That is, learning about the pairing of a neutral stimulus with a valenced word or picture occurs over multiple presentations, which occur at irregular intervals in somewhat different contexts.

If we are correct in attributing the attitude change following evaluative conditioning to a breakdown in access control, then there is little reason to look for a change in attitude that occurs without performance on a final explicit test that reveals some level of awareness of the pairings between the neutral cues and the valenced stimuli. Because we assume that the same memories are involved, we expect that there will be a substantial correlation between performance on the explicit and implicit tests. Second, the crucial concern for a theory about a breakdown in access control is to rule out intentional retrieval. That is, for the theory to be correct, the retrieval of the associations formed during the study phase must be unintentional or unbidden. We expect that at times people will be aware that they are recalling associations formed during the study phase (e.g., Humphreys et al., 2009). At other times, however, people will not be highly aware of the origin of the associations that they are retrieving. Finally, our ideas about a loss of access control can be tested on non-valenced stimuli. This can be advantageous because at times the use of non-valenced stimuli may serve to reduce awareness of the conceptual basis of the pairings that are being learned.

The hybrid evaluative conditioning source-monitoring paradigm

Our first priority is to create a new paradigm that would produce good learning and where there is a reasonable chance that participants would not understand the categorical nature of the stimuli during the learning phase. Our starting point in designing an alternative paradigm is to note that evaluative conditioning and discrimination learning approaches are similar to the source-monitoring paradigm (Johnson, Hashtroudi, & Lindsay, 1993; Johnson, Taylor, & Raye, 1977). However, in source-monitoring, participants study a list of items that have been arbitrarily paired with different contexts, and are then asked to discriminate the source of the items (e.g., words presented in two different colors or styles of font, spoken in a male or female voice, presented auditorally or visually). As far as we know, a source-monitoring paradigm has never paired sets of stimuli that have a family resemblance with cues, even though such pairings are common in evaluative conditioning and discrimination learning. In addition, it seems possible to reduce awareness of the conceptual basis for the pairings without reducing the learning of the individual pairings by making classification an incidental part of another task, such as an explicit memory task (e.g., Brooks, Squire-Graydon, & Wood, 2007).

In Experiment 1, we pair fictitious brand names with attractive and unattractive faces. To encourage learning of the brand–face pairings (while at least partially concealing the fact that we were manipulating attractiveness), we presented short lists of brand–face pairs followed by a associative-recognition test. That is, participants were told to remember the brand that was paired with each face, so attractiveness was not the focus of the task. In order to make concept formation more difficult, each brand was paired with a small number of attractive faces or a small number of unattractive faces. However, each face only occurred once.[1] After learning and being tested on many such lists, half of the participants were asked to rate how much they liked each brand name ("evaluative conditioning"), and half were asked to decide whether the brand name had been paired with one or more attractive faces ("source monitoring"). We used a likeability task to show that we could replicate the standard findings in the literature, though we made no attempt to see if the participants making these ratings were aware that brands had been paired with attractive or unattractive faces.

The source monitoring task was used to assess the likelihood that participants could recover categorical information at test. No attempt was made, however, to determine whether this information had been explicitly stored at study (e.g., a verbal label or response such as "this brand was paired with an attractive face" is stored with the brand at study) or was based on the retrieval of some other form of information (e.g., an image of a face, sub-symbolic information).

The following two experiments were used to determine whether the intrusion of recently learned information into a personal judgment task (likeability ratings) or a semantic memory task could occur in an unbidden manner. In an effort to show that retrievals of the recently acquired information in our paradigm could occur in an unbidden fashion in Experiments 2 and 3, we made it progressively less likely that participants would see the relevance of this earlier information. That is, in Experiment 1 there was no a priori basis for the likeability ratings, so it is possible that participants would deliberately use information from the experimental session to make their ratings. In Experiment 2, we used a speeded classification task to reduce the probability that participants would deliberately try to retrieve the recently acquired information. We paired words with male and female faces, and at test, we asked participants whether the words were stereotypically masculine or stereotypically feminine. In normative data, the words that we used differed along this dimension, so there was a real basis in semantic memory for making the decision. In addition, during study, half of the normatively masculine words had been paired with male faces, and half had been paired with female faces (and vice versa for the normatively feminine words). This meant that the pairing was not a valid cue for the masculine–feminine decision. Note that although the distinction between male and female faces is very obvious, it is unlikely to attract much attention. That is, participants would not be surprised by the use of both male and female faces in a face recognition experiment whilst they might be surprised at the use of pleasant and unpleasant photographs.

The associative recognition phase for Experiment 3 was the same as Experiment 2, but instead of masculine and feminine words, we presented active and inactive words. Then, rather than presenting a speeded classification task (as we did in Experiment 2), we presented participants with a modification of the IAT procedure in which they classified words as active or inactive and faces as male or female (Greenwald et al., 1998). "Congruence" in Experiment 3 was determined by the relationship between the word–face pairings during the pair recognition task and the word–face pairings during the IAT. We assumed that in making active and inactive judgments to words, participants would not deliberately retrieve information that was irrelevant to this judgment and that they perceived as being irrelevant to the judgment.

Experiment 1

Method

Participants and design

Forty students from the University of Queensland were randomly assigned evenly to Conditions 1 and 2 where

[1] In this experiment, the targets that are paired with any one cue have a family resemblance. That is, they are similar in some aspects, although there may not be a defining feature or features shared by all members of the category. The concept that participants might or might not acquire is that a particular stimulus is consistently paired with attractive faces or with unattractive faces. In this paradigm, it would also be possible to instantiate a concept at the level of the list by pairing cues that had a family resemblance (active or inactive words) with targets that had a family resemblance (e.g., male or female faces). In such a design, active words might be paired with female faces and inactive words with male faces or vice versa.

brand name repetition (1, 3, or 6) and facial attractiveness (attractive, unattractive) were within-subject variables.

Stimuli

The stimuli consisted of 48 fictitious brand names and 160 (80 attractive and 80 unattractive) faces. For each participant, 24 brands were randomly presented one, three, or six times with photos of attractive faces and 24 brands were randomly presented one, three, or six times with photos of unattractive faces. Each photograph was presented once at study and again at test. The list of 48 fictitious brand names was compiled by searching published papers that had used fictitious brand names as part of their experiment. The final list consisted of 24 words (e.g., Star) and 24 non-words (e.g., Tiddip) and included product relevant brand names (e.g., X-Rust) and product irrelevant brand names (e.g., Circle). The photographs were selected on the basis of pilot work involving four separate groups of 20 participants each providing an attractiveness rating of 116 photographs (90 female and 26 male) on a 9-point scale. The 464 greyscale photos were obtained from a casting database http://www.interfaces.nl by searching for individuals ranging from 18 to 25 years. Each photo was then normalized according to each participant's ratings and averaged across the 20 participants in each of the four groups. The 80 female photographs that received the highest attractiveness ratings were selected for the attractive condition and the 80 female photos that received the lowest attractiveness ratings were selected for the unattractive condition. The chosen faces differed across several dimensions (e.g., ethnicity, hair color and length, make up, pose, etc.) typical of a photo shoot for a talent agency.

Procedure

For each participant, the study consisted of two tasks: (1) a pair recognition task and (2) a brand likeability rating (Condition 1) or a source monitoring task (Condition 2). The pair recognition task, completed by all participants, involved 20 trials. The study phase consisted of the presentation of eight brand–face pairs for 3 s each. There were always four attractive and four unattractive faces in each list. The test phase consisted of four intact and four rearranged pairs. In creating rearranged pairs, if a brand had been paired with an attractive face at study, then it was paired with another attractive face from the same study list at test and similarly for brands paired with unattractive faces. Consequently, in each test list, there were two intact and two rearranged pairs with an attractive face and two of each with an unattractive face. Following the test phase of a trial, participants pressed the space bar to begin the study phase of the next trial. In all, there were 20 trials, divided into three blocks (Block 1: 6 trials; Block 2: 6 trials; Block 3: 8 trials). The brand names that appeared six times in total were presented twice in each block. Those that appeared three times were presented once in each block and the brand names appearing only once were presented in the final block only (hence the extra number of trials in this block). The list that they appeared in was randomized within blocks, consistent with the overall restrictions described above. There was no repetition of a brand name within an 8-pair list. The instructions described the pair recognition task as if it were the sole task and this emphasis on explicit memorizing of the brand–face pair was maintained by the repeated study and test phases.

After completing all 20 trials, participants were given one of two tasks: a brand likeability rating task (Condition 1) or a source monitoring task (Condition 2). Those in Condition 1 were asked to rate the likeability of each brand name by clicking on the appropriate button (very likeable; likeable; neutral; unlikeable; very unlikeable). Those in Condition 2 were asked to indicate whether each brand name had been presented with one or more attractive faces or one or more unattractive faces by clicking on the appropriate button (attractive; unattractive). After responding, the next brand name immediately appeared on the screen.

Results

Pair recognition

Pair recognition was generally good. Collapsing across blocks and conditions, the hit rate (i.e., correctly classifying an intact brand–image pairing as "intact") was .84 and the false alarm rate (i.e., incorrectly classifying an rearranged brand–image pairing as "intact") was .21, with a d' value of 2.01. Thus, participants were learning the individual brand–face pairings. Similar results were obtained in the subsequent experiments and will not be reported.

Likeability ratings

The mean brand likeability ratings as a function of the number of presentations and the attractiveness of the paired faces are presented in Table 1. A 2 × 3 within-subjects ANOVA was conducted with facial attractiveness (attractive, unattractive) and number of pairings (1, 3, 6) as independent variables, and brand likeability ratings (1, very unlikeable to 5, very likeable) as the dependent variable. Results revealed a significant main effect of attractiveness, $F(1, 19) = 29.594$, $MSE = .262$, $p < .001$, partial $\eta^2 = .609$, such that participants rated brand names that were previously paired with attractive faces ($M = 3.30$) as more likeable than those previously paired with unattractive faces ($M = 2.79$). Likewise, there was a significant main effect of number of pairings, $F(2, 38) = 4.468$, $MSE = .165$, $p = .018$, partial $\eta^2 = .190$ (Ms = 2.922, 3.025 and 3.191 for 1, 3 and 6 repetitions, respectively). Finally, there was a marginal attractiveness by number of pairings interaction, $F(2, 38) = 2.646$, $MSE = .102$, $p = .084$, partial $\eta^2 = .122$. A simple effects analysis revealed that repeated presentations of the brand names increased brand name likeability when they were paired with attractive faces [$F(2, 38) = 7.334$, $MSE = .117$, $p = .002$, partial $\eta^2 = .278$] but not when they were paired with unattractive faces [$F(2, 38) = .971$, $MSE = .149$, $p = .388$, partial $\eta^2 = .049$].

Table 1
Brand likeability ratings (1, very unlikeable to 5, very likeable) as a function of attractiveness and number of pairings.

	Number of pairings		
	1	3	6
Attractive	3.08	3.33	3.49
Unattractive	2.76	2.73	2.89

Source monitoring

The hit and false alarm rates were calculated for the results from the source monitoring condition. Responding "attractive" to a brand that was originally paired with attractive faces is a hit, and responding "attractive" to a brand that was originally paired with unattractive faces is a false alarm. Presenting the results in terms of hits and false alarms instead of probability correct is arbitrary though the different methods of presentation are mathematically equivalent. A presentation in terms of hits and false alarms has the advantage of putting the results into the same format as the likeability ratings in Table 1, and it focuses attention on the central concept of discriminability. The mean hit and false alarm rates for Experiment 1 are presented in Table 2.

A 2 × 3 within-subjects ANOVA was conducted on the hit and false alarm rates. The variables were number of pairings and attractiveness. The effect of attractiveness was significant, $F(1, 19) = 45.158$, MSE = .071, $p < .001$, partial $\eta^2 = .704$, indicating that the probability of responding "attractive" was higher for brands paired with attractive faces ($M = .621$) than for brands paired with unattractive faces ($M = .294$). There was no significant main effect of the number of repetitions, $F(2, 38) = 2.342$, MSE = .031, $p = .110$, partial $\eta^2 = .110$, but the interaction between the number of repetitions and attractiveness was significant, $F(2, 38) = 3.809$, MSE = .039, $p = .031$, partial $\eta^2 = .167$. A simple effects analysis showed that the probability of responding "attractive" to the brands paired with unattractive faces (i.e., false alarm rate) was not influenced by the number of pairings, $F(2, 38) = .580$, MSE = .022, $p = .565$, partial $\eta^2 = .030$. However, there was a significant effect of number of pairings on the probability of responding "attractive" to the brands paired with attractive faces (i.e., hit rate), $F(2, 38) = 4.321$, MSE = .048, $p = .020$, partial $\eta^2 = .185$. Specifically, the hit rate increased from 1 to 3 repetitions [$t(19) = 2.843$, $p = .01$, $SEM = .068$] but not from 3 to 6 repetitions [$t(19) = -.702$, $p = .491$, $SEM = .062$]. In addition, the probability of responding "attractive" was higher for brands paired once with attractive faces ($M = 0.506$) than for brands paired once with unattractive faces ($M = 0.319$), $t(19) = 2.334$, $p = .031$, $SEM = .080$.

Discussion

Brands paired with attractive faces were rated as more likeable than brands paired with unattractive faces. In addition, the likeability ratings increased with number of pairings for brands paired with attractive faces but not for brands paired with unattractive faces. A similar pattern occurred with source monitoring judgments about whether a brand had been paired with one or more attractive faces. In particular, the hit rate increased from 1 to 3 pairings.

There is no contention that we have eliminated awareness of the fact that we were pairing some brands with attractive faces and some with unattractive faces. Furthermore, if we asked the same participants the source monitoring question (in addition to the likeability rating) and eliminated those who appeared to be aware, we do not think that it would be helpful. That is, the search for a situation where attitudes change without there being some level of performance on an explicit task is predicated on the assumption that different types of memory are involved. However, we think that the likeability ratings are predominantly a semantic memory process (Kashima & Kerekes, 1994), which happens to be affected by the episodic memories just acquired (Humphreys, Murray, and Maguire, 2009). In contrast, the source monitoring judgments are a more direct test of those same episodic memories. The result is that performance on the two tasks will likely be positively correlated making it difficult to find an effect on one task when there is no effect on the other task. This is especially true given the good performance on the source monitoring task. We used an extremely long list of 180 face-brand pairings and our participants could discriminate between once presented brands that had been paired with an attractive face and once presented brands that had been paired with an unattractive face. We cannot rule out the possibility that this discrimination is based on the retrieval of a label (attractive or unattractive) that was supplied by the study participant and associated with the brand at study. We also cannot rule out the possibility that this discrimination is based on the retrieval of some other form of information such as sub-symbolic information or an image. The point here is that we cannot assume that awareness on an explicit test is necessarily an indication of awareness at study.

Table 2
Performance (hit and false alarm rate) on the source monitoring task by attractiveness and number of pairings.

	Number of pairings		
	1	3	6
Attractive (hit rate)	.51	.70	.66
Unattractive (false alarm rate)	.32	.29	.27

Experiment 2

We cannot rule out the possibility that the likeability ratings in Experiment 1 are the result of demand characteristics. The problem with the likeability rating task as we instantiated it, and as it is generally instantiated in the evaluative conditioning literature, is that participants have no real basis for their ratings. That is, most of the brands are not rated as particularly likeable or unlikeable, and in the absence of any real difference in affect, it is possible that another basis for the ratings might be found, such as what the participant thinks the experimenter wants. To overcome this problem in Experiment 2, we examine the effect of recent learning on semantic judgments where there is a real basis for making the judgment.

Using the same pair recognition procedure as Experiment 1, we presented masculine and feminine words taken from Heise (1965) and Jenkins, Russell, and Suci (1958) alongside male and female faces. Instead of requesting likeability ratings afterward, participants were asked to classify each of the words as masculine or feminine as quickly as possible. Again, their goal in the pair recognition phase was to remember the word that was paired with each face, so its gender was not the primary focus during encoding. Each of the words was paired with three different faces during pair recognition. Half of the to-be-classified words are "congruent", as they are paired with the same gender during training, and half are "incongruent", as they are paired with the opposite gender during training. For example, the word "army" (a masculine word) may have been paired with three different male faces during the pair recognition phase, and "motor" (a masculine word) may have been paired with three different female faces during the pair recognition phase. The question is whether participants would classify "army" (a congruent word) faster and more accurately as masculine than "motor" (an incongruent word) in this latter phase given their prior exposure to the "army"–male faces and "motor"–female faces pairings.

Method

Participants and design

Forty-one students from the University of Queensland participated for course credit. Word type (masculine or feminine) and congruency (congruent or incongruent) were within-subject variables, and we measured both reaction time and accuracy. The words were randomly assigned to the congruent and incongruent conditions.

Stimuli

The stimuli consisted of 40 normatively masculine words (e.g., rugged, control, machine) and 40 normatively feminine words (e.g., smooth, heart, family) taken from Heise (1965) and Jenkins et al. (1958). Words were selected on the basis that they rated highly as either feminine or masculine relative to the other scales and also on the basis that they did not share a strong orthographic, morphemic or semantic similarity to other words in the list. 240 photographs (120 male and 120 female) were selected from the same casting database as Experiment 1.

Procedure

The study consisted of two tasks: (1) a pair recognition task and (2) a speeded classification task. During the pair recognition task, 20 of the masculine words were paired with three different male faces (masculine-congruent) and 20 were paired with three different female faces (masculine-incongruent). Similarly, 20 of the feminine words were paired with three different female faces (feminine-congruent) and 20 were paired with three different male faces (feminine-incongruent).

From the participants' perspective, their goal in the pair recognition task was to remember the word that was paired with each face, so they were asked to study a series of eight word–face pairings presented sequentially (counterbalancing the gender of the words and faces). After studying these eight word–face pairs, we tested their memory by presenting the eight pairs again. But this time, four of the pairs were intact (i.e., the words were paired with the same faces as they studied) and four of the pairs were rearranged (i.e., the words were paired with different faces than they studied). In order to form a rearranged pair, each word was paired with the same type of face with which it had been studied. We repeated this process 30 times (40 words and 240 faces = 30 blocks of eight pairs). The instructions described the pair recognition task as if it were the sole task. This process ensures excellent learning of the word–face pairings, while establishing the (in)congruence between the gender of the word and the gender of the face described previously.

After completing all 30 blocks, participants were given a speeded classification task. We presented each of the 80 words from the pair recognition phase one at a time in random order, and asked participants to determine whether the word presented on the screen was stereotypically masculine or stereotypically feminine by pressing the appropriate keys on the keyboard. They were asked to respond quickly, but accurately. After making their response, the next word immediately appeared on the screen.

Results

Reaction times

The mean reaction times for the speeded classification task (as a function of word type and congruency) are presented in Table 3. A 2 (word type: masculine, feminine) × 2 (congruency: congruent, incongruent) within-subjects ANOVA on the reaction time data revealed a significant main effect of word type, $F(1, 40) = 11.62$, MSE = 247654.418, $p = .002$, partial $\eta^2 = .225$, indicating that participants were generally faster to respond to feminine words than masculine words. Congruency was not significant, $F(1, 40) = 1.812$, MSE = 41206.409, $p = .186$, partial $\eta^2 = .043$, but resulted in a marginal interaction with word type, $F(1, 40) = 3.951$, MSE = 60137.823, $p = .054$, partial $\eta^2 = .090$. A simple effects analysis revealed that participants were faster to respond to masculine-congruent words than masculine-incongruent words, $F(1, 40) = 6.599$, MSE = 100452.27, $p = .014$, partial $\eta^2 = .142$, but not for feminine (in)congruent words, $F(1, 40) = .059$, MSE = 891.962, $p = .81$, partial $\eta^2 = .001$.

Accuracy

The mean proportion of words that were correctly identified as masculine or feminine as a function of congruency are also presented in Table 3. A 2 (word type: masculine, feminine) × 2 (congruency: congruent, incongruent) within-subjects ANOVA on these accuracy data revealed a nonsignificant main effect of word type, $F(1, 40) = 3.151$, MSE = 0.018, $p = .084$, partial $\eta^2 = .073$, but a significant main effect of congruency, $F(1, 40) = 34.598$, MSE = 0.024, $p < .001$, partial $\eta^2 = .464$, which interacted with word type, $F(1, 40) = 38.416$, MSE = .006, $p < .001$, partial $\eta^2 = .49$,

Table 3
Mean reaction time and accuracy for the speeded classification task in Experiment 2.

	Masculine		Feminine	
	Congruent	Incongruent	Congruent	Incongruent
Reaction time (ms)	946	1016	906	900
Proportion correct	.79	.72	.83	.61

where there was a larger effect of congruence for feminine words than masculine words.

Discussion

Participants were told that they were going to be shown a series of short (8-pair) lists. Each pair consisted of a masculine or feminine word paired with a male or female face. Across the experiment, each word was paired with three different male faces or three different female faces. Half of the masculine words were presented with male faces, and half were presented with female faces. Similarly, half of the feminine words were presented with male faces, and half were presented with female faces. Participants were told that they were to learn the pairings as they would receive an associative-recognition test following each list. These instructions were reinforced by the long series of study and test trials. Participants took the pair recognition task seriously and performed well. We then gave them a speeded classification task where they had to classify the words they had studied as masculine or feminine. Even though they had not prepared for this task, the congruence of the pairings clearly influenced the accuracy of their classification. That is, they were significantly more accurate at classifying the congruent words than the incongruent words (particularly the feminine words). They were also faster to classify the congruent masculine words than the incongruent masculine words (though the congruency by word type interaction did not reach significance). Thus, there is not a speed accuracy tradeoff. Note that the greater reactivity of the feminine words may reflect real changes in the perception of these words since the norms were collected in the 1950s and 1960s. Alternatively, the societal changes since then may have provided these words with multiple, partially competing associations.

Because there was a real basis in semantic memory to make these decisions, it seems unlikely that participants deliberately retrieved an image of one or more of the faces that had recently been paired with that word. There are two reasons for making this assertion. First, it seems unlikely that participants would have thought that retrieving this information was relevant to the decision that they were being asked to make. That is, they would have known that masculine words had been paired with both male and female faces (and similarly for female words). Second, memory retrieval of this kind is quite difficult (long list and retention interval, single learning trial on each pair), so it seems unlikely that participants would have imposed this extra burden on themselves.

Experiment 3

Although there was no relationship between masculine and feminine words and the gender of the face with which it was paired, it is possible that some participants formed the opinion that there was a relationship. In turn, this might have induced them to recall the image of a paired face when they were making their judgments about the words. Experiment 3 was designed to make it even less likely that participants would see any relationship between the paired associate task and the judgment task. The study phase of Experiment 3 was basically similar to Experiment 2, except that active and inactive words were paired with male and female faces. Half of the active words were paired with male faces and half were paired with female faces. Similarly, half of the inactive words were paired with male faces and half were paired with female faces. In Experiment 2, we think that the recent learning was intruding into the judgment task because the two tasks were conducted in similar contexts. In addition, the retrieval operations involved in retrieving aspects of masculinity and femininity may partially overlap with the retrieval operations involved in using facial features to retrieve the appropriate gender (we elaborate on this point in the final discussion). In order to enhance the possibility of retrieving the recent learning about pairings with male and female faces while making active–inactive judgments, we presented participants with a version of the IAT in which they classified words as active or inactive and faces as male or female. Our thinking was that recent experience with classifying pictures of faces as male or female would prime the tendency to retrieve aspects of masculinity and femininity when classifying the words as active or inactive. Half of the participants were instructed to press, say, the left key for male faces and active words and the right key for female faces and inactive words. The other half of the participants were instructed to press the left key for male faces and inactive words and the right key for female faces and active words. Therefore, "congruence" in Experiment 3 was determined by the relationship between the word–face pairings during the pair recognition task and the word–face pairings during the IAT. For example, the word "solve" (an active word) may have been paired with three different male faces during the pair recognition phase, and "produce" (an active word) may have been paired with three different female faces during the pair recognition phase. The question is whether participants would be faster and more accurate in classifying "solve" during the IAT when active words and male faces share the same response key (a congruent word) compared to the word "produce", which was originally paired with female faces and does not share the same response key (an incongruent word).

Method

Participants and design

Forty-one students from the University of Queensland participated for course credit. Word type (active or inactive) and congruency (congruent or incongruent) were within-subject variables, and we measured both reaction time and accuracy.

Stimuli

The stimuli consisted of 40 active words (e.g., carry, build, arrive) and 40 inactive words (e.g., limit, contain, belong) taken from the same source as Experiment 2, and 240 photographs (120 male and 120 female) were obtained from the same casting database as Experiment 1.

Procedure

The study consisted of two tasks: (1) a pair recognition task and (2) the IAT. During the pair recognition task, 20 of the active words were paired with male faces and 20 were paired with female faces. Similarly, 20 of the inactive words were paired with female faces and 20 were paired with male faces. Again, from the participants' perspective, their goal was to remember the word that was paired with each face, so they were asked to study 30 lists of eight word–face pairings presented sequentially (counterbalancing the word type and gender of the faces). The procedure for the pair recognition phase was identical to Experiment 2, but we used active and inactive words rather than masculine and feminine words.

After completing all 30 lists, participants were given the IAT. We presented each of the 80 words from the pair recognition phase along with 80 previously unseen faces (40 male and 40 female) one at a time in random order. On each of the 160 trials of the IAT, participants were instructed to press one of two response keys. For example, they were asked to press the "Q" key as quickly as possible when presented with male faces or active words, and the "P" key as quickly as possible when presented with female faces or inactive words. In fact, the response keys and assignment of word type and gender of the face were counterbalanced. They were asked to respond quickly, but accurately. After making their response, the next word (or face) immediately appeared on the screen.

Results

Reaction times

The mean reaction times for the IAT (as a function of word type and congruency) are presented in Table 4. A 2 (word type: active, inactive) × 2 (congruency: congruent, incongruent) within-subjects ANOVA on the reaction time data revealed a non-significant main effect of word type, $F(1, 40) = 1.933$, MSE = 46278.943, $p = .172$, partial η^2 = .046, but a significant main effect of congruency, $F(1, 40)$ = 5.706, MSE = 16835.728, $p = .022$, partial $\eta^2 = .125$, which did not interact with word type, $F(1, 40) = .166$, MSE = 24818.902, $p = .686$, partial $\eta^2 = .004$. That is, participants were faster to classify the congruent words ($M = 997$, $SD = 352$) than the incongruent words ($M = 1045$, SD = 442). A separate ANOVA on the reaction time for faces (rather than words) revealed no significant main effects or interactions. This was expected as the faces were new and had not been paired with active and inactive words. In addition there was no possibility of transfer from the original learning as male and female faces had occurred equally often with active and inactive words.

Accuracy

The mean proportion of words that were correctly identified as active or inactive as a function of congruency is also presented in Table 4. A 2 (word type: active, inactive) × 2 (congruency: congruent, incongruent) within-subjects ANOVA on these accuracy data revealed a significant main effect of word type, $F(1, 40) = 67.965$, MSE = 0.066, $p <$ .001, partial $\eta^2 = .63$, indicating that participants were less accurate in classifying active words than inactive words. Congruency was also significant, $F(1, 40) = 4.241$, MSE = 0.01, $p = .046$, partial $\eta^2 = .096$, indicating that congruent words were better classified than incongruent words. Word type and congruency did not interact significantly, $F(1, 40)$ = 1.792, MSE = .014, $p = .188$, partial $\eta^2 = .043$. A separate ANOVA on the accuracy in classifying faces (rather than words) revealed no significant main effects or interactions. Again this was expected because the faces were new and there was no possibility of differential transfer from the learning trials.

Table 4
Mean reaction time and accuracy for the IAT in Experiment 3.

	Active word		Inactive word	
	Congruent	Incongruent	Congruent	Incongruent
Reaction time (ms)	1015	1073	978	1017
Proportion correct	.50	.44	.80	.80

Discussion

As in Experiment 2, the participants would have focussed on learning the word–face pairings rather than figuring out what kind of face was being repeatedly paired (three pairings) with a particular word. Of course, some spontaneous abstraction of the type of face being paired might have occurred. In spite of this lack of focus on the type of face being paired with a particular word, there was a clear effect of congruency on both speed and accuracy on the IAT. Our interpretation is that when classifying words there was a tendency to retrieve information that had recently been paired with that word. This information would include the retrieval of aspects of the faces paired with that word and possibly verbal labels or expressions (e.g., "another male face") that were spontaneously produced at study. Two factors would have contributed to this tendency to retrieve recently learned information. First, the test was taking place in the same general context as the study experience. Second, the task of classifying faces as male or female would have primed the tendency to retrieve male or female features or characteristics. The retrieved information about maleness or femaleness would have then speeded up the active–inactive decision when the retrieved information was mapped onto the same response as was required by the word classification task (e.g., an active word retrieves information about maleness and the same response is used for male faces and active words). In addition, it seems very likely that this retrieval of the information about maleness and femaleness was not deliberate. That is, the participants would not have seen a connection between pairing words with male and female faces and making active–inactive judgments about those words. Furthermore, it is even less likely that they would have realized that the experimenter expected them to make faster responses to the words in the congruent than in the incongruent conditions. At times, participants may be aware that they are retrieving recently acquired information while making their speeded classifications. We doubt that this is common, but the current experiment does not address this issue.

General discussion

The traditional approach to evaluative conditioning sought to show that memory on an implicit test was above chance while memory on an explicit test was at chance. We were concerned about this approach for two reasons. First, we felt that attempts to reduce awareness by "complicating" the study conditions were seriously reducing learning. Second, we felt that above-chance performance on an explicit test need not indicate that the participant was fully aware of the contingencies involved at either study or on the implicit test.

We were able to show that a small number of pairings (1, 3, or 6) of fictitious brands with attractive or unattractive faces changed likeability ratings on brands, although in this study and many previous evaluative conditioning studies, this may have been a demand characteristic. The small number of pairings also produced a reasonable memory for whether a brand name had been paired with an attractive or an unattractive face. Labels that may have been applied at study ("this brand is paired with an attractive face") could have played a role in the ability to make these source monitoring judgments. Nevertheless, the good source monitoring performance and the nature of the paradigm that diverts attention from the categorical nature of the stimuli raises the possibility that categorization was occurring at test, not at study. This is a question that needs to be explored further.

A small number of pairings of masculine and feminine words with male and female faces also changed the accuracy of masculine and feminine judgments about the words. This is less likely to be due to demand characteristics because there was a real basis for categorizing the words, which there had not been for the likeability ratings. Furthermore, the speeded nature of the decision making, the single learning trial for each face-word pairing, and the very long study list would reduce the probability that participants would undertake the difficult task of episodic retrieval. Finally, pairing active and inactive words with male and female faces resulted in IAT-type congruency effects in a subsequent test where words were classified as active or inactive, and faces were classified as male or female. This appears to be a clear example of unbidden recall where the participants are not trying to retrieve faces that had been previously paired with the test word or the thoughts they may have had while studying that pairing.

There are of course other examples in the evaluative conditioning literature that support the conclusion that the prior learning can occur in an unbidden manner. For example, Mitchell, Anderson, and Lovibond (2003) and Olson and Fazio (2001) both used the IAT. Of particular relevance is the Meersmans, De Houwer, Baeyens, Randell, and Eelen (2005) study, which like ours, used non-valenced USs. However, their results using a priming measure were not very consistent.

The primary difference with our procedures is that we used a much longer list than has been used in previous research. This has several advantages. First of all, it permitted us to present stimuli that varied almost continuously along a dimension (masculine–feminine, or active–inactive). Thus, there was a real basis for making a decision, making it less likely that participants arrive at a decision about what they think the experimenter would want. In addition, it means that episodic retrieval would have been a resource demanding task making it less likely that participants would have adopted the strategy of attempting to retrieve information about a words pairings. The long list and the more or less continuous variation in the cues also helps to conceal the categorical way in which we assigned targets to particular stimuli. Finally, the large number of pairs that we use means that we do not have to reuse pairs or items during the test and it contributes to the stability of the results.

Note also that the use of a priming procedure does not eliminate the possibility that participants will "intentionally" use the prime to retrieve recently learned information. That is, it appears that retrieval operations can be set up in advance of the presentation of the cue so that once the cue is presented, the retrieval runs to completion without further input. For example, the production of an antonym or a synonym to a cue can be as fast (if not faster) than the production of a free association (Davidson & Cofer, 1968; Woodworth & Schlosberg, 1954). The very rapid, and apparently effortless, retrieval of a target that was paired with the cue in the immediately preceding list also appears to be an example of a retrieval operation that has been set up in advance of the presentation of the cue (Humphreys et al., 2009). A possible conclusion from this line of thinking is that if a participant believes that their recent learning history of the primes presented is relevant to the task they are performing, then this may have an impact on the amount of priming observed.

Note that our procedures differ from the traditional IAT procedures in that we did not use two blocks of trials in which the contingencies between the response keys used changes from Blocks 1 to 2 (Greenwald et al., 1998). That is, if the key used for active words and male faces is the same in Block 1, the key used for active words and female faces would be the same in Block 2. The two-block procedure is used to provide an estimate of the congruency effect at the level of the individual. The two-block procedure is not necessary if the focus is on a experiment wide level (the congruency effect that obtains after collapsing over the counter balancing conditions). In addition, the two-block procedure cannot help but draw attention to the relationship between the two judgments that are being made, which would reduce our ability to conclude that the recent learning was intruding in an unbidden manner.

An explanation for Experiments 2 and 3

In order to explain the results of Experiment 3 where we used an IAT paradigm, we assume that when attempting to classify words as active or inactive, there is a tendency to retrieve images of the photographs paired with those words and/or to retrieve thoughts about maleness or femaleness that occurred while studying the word-name pairs. In our situation, this tendency occurs for two reasons. First, the IAT is taking place in substantially the same context as the study phase. Second, we assume that there is some crosstalk when switching between the two retrieval tasks. In this case, the requirement to classify faces as male or female will promote the use of a cue to retrieve aspects of maleness or femaleness. Because there will be some carryover to the active–inactive classification task, there will also be some retrieval of aspects of maleness or femaleness during this task. If the retrieved image is of a male face, or if the retrieved thoughts are about maleness, the response of pressing the "male" button will be partially activated. Likewise, the response of pressing the "female" button will be partially activated if the retrieved image is of a female face, or if the retrieved thoughts are about femaleness.

The explanation for Experiment 2, where participants made speeded classifications of male and female words that had been paired with male and female faces, is similar. Again, we assume that there is a tendency to retrieve a paired image or thoughts that occurred while studying the word–face pair. This occurs because participants are trying to use the word to retrieve aspects of maleness and femaleness in order to perform the masculine–feminine classification task. Thus there will be a tendency to retrieve images of the faces that had been paired with the words or the thoughts which occurred when the pair was studied. There may also be a contextual component.

Relationships with evaluative conditioning theories and phenomena

Pavlovian conditioning

One of the frequently cited differences between evaluative and Pavlovian conditioning is that the former is resistant to extinction (Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Jones, Fazio, & Olson, 2009). That is, in Pavlovian conditioning, repeated presentations of the CS in the absence of the US leads to extinction, a failure to respond to the CS. This did not appear to be true in evaluative conditioning, although Davey (1994) had questioned whether there is enough data on this issue. More recently, a meta analysis by Hofmann, De Houwer, Perugini, Baeyens, and Crombez (2010) has shown that repeated presentations of the CS by itself in evaluative conditioning are associated with smaller effect sizes.

Similar findings do not, however, necessarily imply similar causes. The existence of extinction in Pavlovian conditioning is compatible with the idea that in some sense, the CS is serving to predict the US. That is, if one is learning that the CS predicts the US, then presentations of the CS alone would serve as evidence that the CS is not a predictor of the US, or that circumstances have changed so that it is no longer a predictor. Something similar to extinction occurs in the memory for a pair of words. That is, after learning a pair of words (AB) the presentation of the cue term (A) by itself can reduce the recall of B given A as a cue for B (Overton & Adolphson, 1979). Furthermore, this may only occur when an orienting task is used so that participants do not attempt to retrieve B while they are processing A. That is, an attempt to retrieve B when A is studied by itself should improve recall for B given A as a cue (Roediger & Karpicke, 2006). However, Hart, Harris, and Westbrook (2009) have argued that in fear conditioning, the CS needs to elicit fear in order for extinction to work. It may therefore be possible to distinguish between the reduction in performance produced by the presentation of the CS by itself in Pavlovian conditioning and the reduction in performance produced in evaluative conditioning and associative memory by determining whether the isolated presentation of the CS/cue is likely to lead to the reestablishment of the original memory for the pairing.

Our discussion of the possible differences between the reduction in performance produced by unpaired presentations of the CS/cue in Pavlovian conditioning and associative memory is potentially compatible with the referential account proposed by Baeyens, Eelen, Crombez, and Van den Bergh (1992). That is, they proposed that there were two types of Pavlovian conditioning. The first type is the learning of contingencies, which we have already noted. Here, the paired presentation of the CS and US leads to an expectation that when the CS occurs, the US will follow. The second type is referential learning where the paired presentation of the CS and the US leads to the presentation of the CS activating a mental representation of the US without there being an expectation that the US will actually occur. From the description provided by Baeyens et al. (1992), referential learning does not sound all that different to the episodic learning that is assumed to underlie performance on the source monitoring and associative recognition tasks. In this respect, it seems likely that a source monitoring task would show that there is an "episodic" component in many of the evaluative-conditioning paradigms. Note that the performance on the source monitoring task in Experiment 1 undoubtedly underestimates the memory ability of our participants. That is, there would have been genuine disagreement about whether some faces should be assigned to the attractive or unattractive category. We conducted a pilot study, which was identical to Experiment 1, except medium frequency words (11–50 occurrences per million) were paired with male and female faces. With these materials, performance on a source monitoring task was almost perfect after pairing the word with three different male faces or three different female faces.

The learning that underlies associative recognition can also occur under very low levels of attention. Humphreys and Magurie et al. (2010) looked at associative recognition following a maintenance rehearsal task. Participants were told that their task was to recall four digits after a few seconds spent rehearsing a pair of words. The participants were told that the rehearsal task was designed to prevent them from rehearsing the digits. A final unexpected associative recognition task on the word pairs showed that some learning was occurring. This learning was not enhanced by additional massed rehearsals of the word pairs. That is, rehearsing a pair twice during the retention interval for digit recall was no different than rehearsing it four or six times. However, distributed rehearsals (i.e., rehearsing the word pair with two or three different digit strings) produced better associative recognition than rehearsing the pair with a single digit string (also see Nairne, 1983). Although it hasn't been tested, it seems likely that the overt rehearsal of the word pair is not necessary in order to have some associative learning. Instead, some level of focal attention to the word pair may be all that is needed. That is, isolating the pair either temporally, spatially, or through the use of contrasting materials from the other pairs and stimuli used in the experiment may be all that is necessary. The Humphreys and Magurie et al. (2010) results also indicate that an associative-recognition test presented in a yes/no format might well reveal some level of explicit memory in many of the evaluative-conditioning paradigms. Although we are not aware of any formal comparisons, in associative recognition the yes/no format does appear to be more sensitive than either a forced choice or an associative matching format. Note that this proposal assumes that multiple pairs are available for testing to ensure stable results and that individual pairs and/or items are not repeated at test.

Propositional Learning

Mitchell et al. (2009) have argued that human associative learning, including the learning involved in evaluative conditioning, is propositional. We agree with many aspects of their argument, but not with their fundamental premise that all associative learning is propositional. That is, we agree that focal attention is required for associative learning. We also agree that in general there will be a positive relationship between performance on a sufficiently sensitive explicit test and performance on an implicit test. We think the reason for the disagreement is that the paradigms they considered are very much simpler than the ones we have examined in our experiments. That is, when there are only 2 cues, and one signals the US and the other does not, then, if focal attention is required for learning, there is little or no difference between associative learning and propositional learning. In our situation, however, the proposition that a brand has been paired with an attractive face is unlikely to occur on a single trial. That is, there is so much variation in the faces (hair length, amount of makeup, direction of gaze, etc.) that participants are unlikely to come up with the experimenter defined concept (attractive or unattractive face) after a single presentation. Instead, it appears that in this situation, propositional learning would ordinarily occur when the current pairing can be compared to a memory for a previous pairing. This kind of propositional learning undoubtedly occurs in our paradigm, although we doubt that it is entirely responsible for source monitoring performance. The point, however, is that we need to postulate a memory for a pair ("associative learning") that is separate from propositional learning.

Stimulus encoding

Field and Davey (1999) showed that conditioning-like effects could occur when the CS was perceptually similar to the US. This finding is irrelevant to our experiments because we have randomly paired stimuli for each participant. Field and Davey, however, went onto suggest a mechanism that may be relevant. They proposed that when encoding a CS–US pair, there would be a tendency for the encoded features to be shared by the CS and the US. This means that if a neutral face is paired with an especially attractive face, the more attractive aspects of the neutral face may be emphasized. Such an effect could result in increased likeability ratings for the neutral face. Something like this occurs with word pairs. Nelson and his associates (Nelson, Canas, & Bajo, 1987; Nelson, Fisher, & Akirmak, 2007) have shown that there is a cue set size effect in extralist cued recall. In this paradigm, individual words are studied and are then cued with a pre-existing (as determined by free association norms) associate. The cue set size effect is a reduction in recall with increases in the number of pre-existing associates of the cue. However, the cue set size effect disappears when the cue and the target are studied together. It therefore looks like all of the possible interpretations of the cue, except for the one determined by the target, are suppressed when the cue and the target are studied together. Again, this could provide a mechanism whereby an interpretation favored by the US or target comes to dominate the interpretation of the cue. However, at this time it is not known whether cue encoding and association formation are separable processes or whether, what appears to be cue encoding, is a direct consequence of retrieving the target or aspects of the target.

Implicit misattribution

Jones et al. (2009) have proposed that an evaluative response produced by a valenced stimulus (US) is misattributed to the paired stimulus (CS). One could extend this idea to our results by proposing that other characteristics evoked by stimuli such as maleness or femaleness could also be misattributed to the paired stimulus. However, if the context of misattribution is extended in this fashion, it is hard to see how it would differ from our memory theory where some of the properties of a picture, or of how one reacted to the picture at study, become associated with the paired word. In addition, most of the studies reported by Jones et al. (2009) also support an associative explanation. For example, in the first study, they monitored eye movements and reported that evaluative conditioning was stronger when the participant had repeatedly shifted their gaze between the two stimuli. In Experiment 2, attention was drawn to the two stimuli by flashing them in a rapidly alternating sequence. In Experiment 3, they manipulated whether the two members of a pair occurred close together on the screen or far apart. All of these manipulations should produce better associative learning, which should be apparent on a final source monitoring or associative recognition task though the retention interval may have to be reduced before the explicit task will show an effect.. However, at this time, it is not known whether the manipulation that Jones et al. used in Experiment 4 will enhance associative learning and/or alter the relative strength of the CS to US and US to CS associations. In Experiment 4, they manipulated the relative size of the CS and the US. Higher conditioning scores were observed when the CS was larger than the US. In Experiment 5, Jones et al. compared conditioning scores for highly and less highly evocative USs. With the less evocative stimuli, participants who were apparently unaware showed conditioning. However, with the highly evocative stimuli, more subjects were aware of the contingencies, and only those who were aware showed conditioning. Jones et al. argued that the emotion produced by the highly evocative stimuli would not be misattributed to the paired word. However, just what is learned when two stimuli are paired and one of them attracts attention is not known. For example, Diana and Reder (2006) argued that low frequency stimuli attracted more attention than high frequency stimuli. They then showed that words were learned less well when they were paired with a low frequency word than when they were paired with a high frequency word. Humphreys and

Magurie et al. (2010) had presented participants with four digits and then had them rehearse word pairs for a few seconds before recalling the digits. Under these conditions, low frequency and novel word pairs appeared to interfere with memory for the digits more than did high frequency and non-novel pairs. Again, it appears that some materials attract attention and that this can interfere with other items that are presented at the same time or which are being maintained while the rehearsal occurs. Thus, the effects of presenting highly evocative USs are likely to be complex and cannot be used at this time to support the misattribution theory.

Contextual dependency

We have suggested that one factor that may lead to the intrusion of recent learning into a semantic or personal judgment task is the fact that the semantic or personal judgment task is typically in the same context as the recent learning. Olson and Fazio (2006) attempted to separate the two situations by telling their participants that they were participating in two different experiments. Because there was still evidence for evaluative conditioning, it is possible that evaluative conditioning is not sensitive to contextual change. As we have indicated, the priming procedure employed by Olson and Fazio (2006) encourages participants to apply the retrieval operations that are needed to perform the target task to the prime. This may overcome (or partially overcome) a change in context. It is not clear, however, whether the Olson and Fazio (2006) contextual manipulation really separated the two occasions in the minds of their participants. After all, they were still participating in psychological experimentation in the same general setting.

Bain and Humphreys (1989; also see Humphreys, Bain, & Pike, 1989) showed that you can separate two occasions in the minds of your participants. At the end of a lecture, students were told that their regularly scheduled tutorial would not be held. Instead, they were then given an opportunity to participate in a data collection process for a well known Australian dictionary. The lecturer introduced the researcher and then excused himself. The students then spent the next hour generating 60 synonyms for what they thought was a projected children's dictionary. The following week, a new lecturer explained that the tutorial session would be spent collecting data for a class project that would require a report from each student. In this session, they read a narrative passage and answered questions about it. They then turned a page in a response booklet and received one of three sets of instructions. One group was asked to recognize the words they had generated synonyms to last week. A second group was asked to recognize the words from the passage they had just read. A third group was asked to rate words for their frequency of occurrence in the language. Words high and low in their frequency in the language were used and words could appear in the synonym generation task only, the passage only, in both, or in neither.

There was no sign that occurrence in either the synonym task or the passage affected ratings for frequency of occurrence in the language. Likewise, there was no sign that presence in the synonym task influenced ratings for the likelihood that a word had occurred in the passage. The results were mixed for the students who were asked to recognize the words from the synonym task. With the high frequency words, there was no indication that occurrence in the passage influenced the ratings. The same effect occurred with one of the two sets of low frequency words that were used. However, with one set of low frequency words, there was a slight tendency to say that words which had occurred in the passage were more likely to have occurred in the synonym generation task. Thus, it is possible to arrange conditions so that there is little or no cross talk between the two conditions. Procedures as strong or stronger than those employed by Bain and Humphreys (1989) will be required to determine whether there is a context effect within evaluative-conditioning paradigms.

Does something emerge with the use of valenced stimuli?

Our theory about the breakdown in access control applies as readily to valenced stimuli as it does to the male and female faces used in Experiments 2 and 3, the masculine and feminine words used in Experiment 2, and the active and inactive words used in Experiment 3. We cannot rule out, however, the possibility that something different is introduced by the use of positively and negatively valenced stimuli. The only way to resolve this issue will be to systematically compare the results obtained using valenced and non-valenced stimuli.

Conclusions

To produce good learning under conditions where there was a reasonable probability that the participant would not become aware of the categorical nature of the stimuli at study, we introduced a hybrid evaluative-conditioning/source-monitoring paradigm. In this paradigm, we paired words (brand names and words) with photographs of faces. The faces paired with a given brand name or word all share a family resemblance (e.g., they are all attractive female faces, they are all male faces, etc.). The object has been to see whether this pairing will transfer to a judgment task (and whether this transfer is deliberate) or whether the recently acquired information emerges in an unbidden fashion.

Across three experiments, we made it progressively less likely that participants would see a connection between the judgment task and the recently acquired memories. There was still a strong effect in Experiment 3 where the recent learning (words were paired with male and female faces) would appear to be irrelevant to the judgment (words were judged as being active or inactive). In addition, the critical conditions (i.e., the congruence between the word–face pairs at study and the assignment of the response buttons in an IAT paradigm) is so complex that participants would not have been able to anticipate what the experimenters were expecting.

More work needs to be done with this paradigm. Specifically, we need to test our hypothesis that awareness sometimes emerges at test—not during study—and that the level of awareness will depend on the test trial instructions. We also need to understand the nature of the information that is emerging in this unbidden fashion. In particular, whether it is symbolic or sub-symbolic.

We have also characterized the unbidden retrieval of recently acquired information in our situation, and many other situations, as a breakdown in access control. This hypothesis is testable as we can use task switching paradigms to investigate the control operations involved in gaining memory access (Humphreys et al., 2009; Logan & Delheimer, 2001; Nelson, et al., 1982). Our explanation also emphasizes the cues that are being used (including contextual cues). These cues can be manipulated, which should change the extent that recently acquired memories emerge in a judgment task.

Our review of alternative theories for evaluative conditioning reveals that there are many similarities with memory theories that have gone unnoticed in the evaluative conditioning literature. Future research will have to explore these similarities.

Acknowledgment

This research was supported by an Australian Research Council Grant DP0985830 to the first three authors.

References

Allen, C. T., & Janiszewski, C. A. (1989). Assessing the role of contingency awareness in attitudinal conditioning with implications for advertising research. *Journal of Marketing Research, 26*, 30–43.

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact, always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy, 10*, 179–199.

Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy, 30*, 133–142.

Baeyens, F., Eelen, P., & Van den bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective–evaluative learning. *Cognition & Emotion, 4*, 3–18.

Bain, J. D., & Humphreys, M. S. (1989). Instructional reinstatement of context: The forgotten prerequisite. In K. McConkey & A. Bennett (Eds.). *Proceedings of the XXIV international congress of psychology* (Vol. 3). Elsevier.

Brooks, L. R., Squire-Graydon, R., & Wood, T. J. (2007). Diversion of attention in everyday concept learning: Identification in the service of use. *Memory & Cognition, 35*, 1–14.

Cacioppo, J. T., Marshall-Goodell, B. S., Tassinary, L. G., & Petty, R. E. (1992). Rudimentary determinants of attitudes: Classical-conditioning is more effective when prior knowledge about the attitude stimulus is low than high. *Journal of Experimental Social Psychology, 28*, 207–233.

Davey, G. C. L. (1994). Is evaluative conditioning a qualitatively distinct form of classical conditioning? *Behaviour Research and Therapy, 32*, 291–299.

Davidson, E. H., & Cofer, C. N. (1968). Some determinants of controlled-association times. *Journal of Experimental Psychology, 78*, 200–207.

Diana, R. A., & Reder, L. M. (2006). The low-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 805–815.

Field, A. P. (2000). I like it, but I'm not sure why: Can evaluative conditioning occur without conscious awareness? *Consciousness and Cognition, 9*, 13–36.

Field, A. P., & Davey, G. C. L. (1999). Reevaluating evaluative conditioning: A nonassociative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 211–224.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Hart, G., Harris, J. A., & Westbrook, R. F. (2009). Systemic or intra-amygdala injection of a benzodiazepine (midazolam) impairs extinction but spares re-extinction of conditioned fear responses. *Learning and Memory, 16*, 53–61.

Heise, D. R. (1965). Semantic differential profiles for 1000 most frequent English words. *Psychological Monographs: General and Applied, 79*, 1–12 (whole no. 601).

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*, 390–421.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review, 96*, 208–233.

Humphreys, M. S., Cornwell, T. B., McAlister, A. R., Kelly, S. J., Quinn, E. A., & Murray, K. L. (2010). Sponsorship, ambushing and counter-strategy: Effects upon memory for sponsor and event. *Journal of Experimental Psychology: Applied, 16*, 96–108.

Humphreys, M. S., Dennis, S., Maguire, A. M., Reynolds, K., Bolland, S. W., & Hughes, J. D. (2003). What you get out of memory depends on the question you ask. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 797–812.

Humphreys, M. S., Maguire, A. M., McFarlane, K. A., Burt, J. S., Bolland, S. W., Murray, K. L., et al. (2010). Using maintenance rehearsal to explore recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 147–159.

Humphreys, M. S., Murray, K. L., & Maguire, A. M. (2009). Contexts and control operations used in accessing list-specific, generalized, and semantic memories. *Cognitive Psychology, 58*, 311–337.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12*, 852–857.

Jenkins, J. J., Russell, W. A., & Suci, G. J. (1958). An atlas of semantic profiles for 360 words. *The American Journal of Psychology, 71*, 688–699.

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

Johnson, M. K., Taylor, T. H., & Raye, C. L. (1977). Fact and fantasy: Effects of internally generated events on apparent frequency of externally generated events. *Memory & Cognition, 5*, 116–122.

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology, 96*, 933–948.

Kashima, Y., & Kerekes, A. R. Z. (1994). A distributed-memory model of averaging phenomena in person impression–formation. *Journal of Experimental Social Psychology, 30*, 407–455.

Lee, A. Y. (2002). Effects of implicit memory on memory-based versus stimulus-based brand choice. *Journal of Marketing Research, 39*, 440–454.

Levey, A. B., & Martin, I. (1975). Classical-conditioning of human evaluative responses. *Behaviour Research and Therapy, 13*, 221–226.

Logan, G. D., & Delheimer, J. A. (2001). Parallel memory retrieval in dual-task situations: II. Episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 668–685.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 3–26.

Meersmans, T., De Houwer, J., Baeyens, F., Randell, T., & Eelen, P. (2005). Beyond evaluative conditioning? Searching for associative transfer of nonevaluative stimulus properties. *Cognition and Emotion, 19*, 283–306.

Mitchell, C. J., Anderson, N. E., & Lovibond, P. F. (2003). Measuring evaluative conditioning using the implicit association test. *Learning and Motivation, 34*, 203–217.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*, 183–198.

Nairne, J. S. (1983). Associative processing during rote rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 3–20.

Nelson, D. L., Canas, J. J., & Bajo, M. T. (1987). The effects of natural category size on memory for episodic encodings. *Memory & Cognition, 15*, 133–140.

Nelson, D. L., Fisher, S., & Akirmak, U. (2007). How implicitly activated and explicitly acquired knowledge contribute to the effectiveness of retrieval cues? *Memory & Cognition, 35*, 1892–1904.

Nelson, D. L., McEvoy, C. L., & Friedrich, M. A. (1982). Extralist cuing and retrieval inhibition. *Journal of Experimental Psychology: Human Learning and Memory, 8*, 89–105.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413–417.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32*, 421–433.

Overton, R. C., & Adolphson, C. J. (1979). Multiple trace retrieval: A trace-to-trace retrieval process? *Journal of Experimental Psychology: Human Learning and Memory, 5*, 485–495.

Page, M. M. (1974). Demand characteristics and classical-conditioning of attitudes experiment. *Journal of Personality and Social Psychology, 30*, 468–476.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Shanks, D. R., & St.John, M. F. (1994). Characteristics of dissociable human learning-systems. *Behavioral and Brain Sciences, 17*, 367–395.

Shimp, T. A., Stuart, E. W., & Engle, R. W. (1991). A program of classical-conditioning experiments testing variations in the conditioned-stimulus and context. *Journal of Consumer Research, 18*, 1–12.

Speed, R., & Thompson, P. (2000). Determinants of sports sponsorship response. *Journal of the Academy of Marketing Science, 28*, 226–238.

Staats, A. W., & Staats, C. K. (1958). Attitudes established by classical-conditioning. *Journal of Abnormal and Social Psychology, 57*, 37–40.

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology* (revised ed.). New York: Holt, Rinehart & Winston.

Yang, M., & Roskos-Ewoldsen, D. R. (2007). The effectiveness of brand placements in the movies: Levels of placements, explicit and implicit memory, and brand-choice behavior. *Journal of Communication, 57*, 469–489.