aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
7(2) 118–134

DOI: 10.1177/1745691611434211
http://pps.sagepub.com
SAGE

Reversal Without Remapping: What We Can (and Cannot) Conclude About Learned Associations From Training-Induced Behavior Changes

Marc N. Coutanche and Sharon L. Thompson-Schill
Department of Psychology, University of Pennsylvania, Philadelphia

Abstract

The "cognitive revolution" in psychology is often framed as a departure from associationist principles rooted in animal learning research, yet it is clear that these principles have immediate relevance for contemporary questions in cognitive and social psychology. Intuitions about the consequences of learning procedures can easily be misleading, making these principles particularly important. To illustrate this point, we identified recent examples of studies applying a particular learning paradigm—response-reversal training—to the study of three different psychological problems (e.g., why objects in the right side of space are preferred to those in the left in right-handed people). The strategy of each study was to alter a typically encountered contingency once in the laboratory, in order to reverse a hypothesized learned response. Yet, contrary to intuitions, we demonstrate that behavior changes can be observed without the reversal of a prior association. Further, many different associative changes can underlie response reversals. We focus on these examples of response-reversal training, but our broader aim is to help connect the animal learning literature to problems in cognitive and social psychology in an effort to strengthen the inferences that might be drawn about learned associations in these contexts.

Keywords

cognition, learning (associative), methodology, behavioral

Consider the following scenario: You, a psychologist, are interested in some aspect of human behavior that you suspect might have its origins in an association that is typically learned from a relationship between stimuli in the environment, and you want to find a way to test your hypothesis. For example, perhaps you hypothesize that people like chocolate because they associate it with happy occasions and not because of a natural taste preference. One approach to test such a hypothesis could be to artificially manipulate the environment to create a different set of contingencies and see whether a different behavior emerges. At first glance, this would seem improbable, given that you cannot control someone's environment for much longer, in practice, than a single lab hour required for an introductory psychology course. But what if after exposing participants to a temporarily reversed contingency during one such brief training session, the behavior of interest changed? A provocative outcome like this could be interpreted as evidence that the behavior of interest was originally the result of the now-reversed association. But should it be interpreted in this way?

This is the question we asked about three recent representative cognitive and social psychology investigations that adopted exactly this approach—which we will call "response-reversal training"—to test hypotheses about associations that the investigators proposed are gradually learned from stimulus and response relationships encountered in our daily environment. And the answer we arrived at, informed by a careful study of associationist principles derived from decades of animal learning research, was "not necessarily." In the pages that follow, we explain why.

To begin, we introduce three example studies that aim to clarify the role of acquired associations on behavior. We will refer back to these investigations repeatedly. Next, we briefly review some key learning principles that must be understood in order to fully appreciate the behavioral effects in these three case studies and that will later be used to suggest new ways to interpret the data. With this foundation established, we then can illustrate the difficulties in interpreting data from response-reversal training studies, given the many and varied

Corresponding Author:

Marc N. Coutanche, Department of Psychology, University of Pennsylvania, Solomon Labs, 3720 Walnut Street, Philadelphia, PA 19104.
E-mail: coumarc@psych.upenn.edu

learning principles that could all produce the same pattern of results. A major goal of these studies is to speak to the basis of a relationship involving a particular stimulus, but this becomes difficult to achieve when other nontarget changes can also produce response reversals. Last, we propose experimental methods that could be used to elucidate existing associations more clearly or to prolong the behavioral changes produced in the lab, returning to our three case studies by way of example.

Reviewing Three Example Training Studies

We begin with a brief description of three studies that illustrate the response-reversal training paradigm—both its implementation and its interpretation—to motivate the theoretical discussion that follows. These three investigations from *Psychological Science* have different hypotheses and backgrounds, but all reported interesting behavioral changes by using the response-reversal training approach. We will refer back to these examples later as illustrations of how theory and evidence from the associative learning field can illuminate and inform such findings. Following descriptions of the studies, Table 1 gives an overview of their relevant contingencies for future reference.

Example 1: Reversing racial bias

In order to understand whether racial prejudice can be modulated, Ito and colleagues reported a reversal of implicit negative racial biases after training subjects to associate black faces and smiling (Ito, Chiao, Devine, Lorig, & Cacioppo, 2006). Participants viewed a set of images of different faces, the majority of which were black, while holding a pencil between their teeth in order to engage muscles involved in smiling (a procedure that has been found to induce or increase feelings of happiness; Flack, 2006; Soussignan, 2002; Strack, Martin, & Stepper, 1988; for reviews and discussions, see McIntosh, 1996; Niedenthal, 2007). The participants, who were of various races except African American, then completed the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), with black and white faces and positive and negative words, with the aim of assessing their racial biases.[1] Inducing participants to smile during the training session appeared to reduce their subsequent IAT-measured bias, compared with having participants maintain a nonsmiling position or smiling while viewing faces, the majority of which were white. The authors concluded that an existing implicit racial bias can be diminished through a smiling induction.

Example 2: Reversing a relationship between hand dominance and valence

Casasanto and Chrysikou (2011) examined a surprising correlation between affective evaluations and motoric "fluency" or competence: Building on previous studies suggesting that people have a tendency to allocate positive interpretations to stimuli appearing on the same side of space as their dominant hand (from here on referred to as their "dominant side"), Casasanto and Chrysikou examined how this bias might be formed. They hypothesized that because a person's dominant hand is more adept at interacting with the world than is the nondominant hand, a learned association may develop between the dominant side and positive valence, with the nondominant side being more closely linked with negative valence. One nonlearning alternative is that genetically determined neurological factors influence both the development of hand dominance and the valences linked with each side of space.

Table 1. A Summary of the Three Example Response-Reversal Training Studies

Study	Possible lifetime associations	Laboratory training	Immediate posttraining response patterns
Ito, Chiao, Devine, Lorig, and Cacioppo (2006)	(a) Black faces → negative or (b) White faces → positive or (c) Both of the above	(a) Black faces → positive	Smaller differences between response latencies for key-response pairings that are "compatible" (white faces–positive words / black faces–negative words) versus "incompatible" (white faces–negative words / black faces–positive words) with existing associations, compared with procedures pairing smiling and white faces or nonsmiling control procedures
Casasanto and Chrysikou (2011)	(a) Dominant side → positive or (b) Nondominant side → negative or (c) Both of the above	(a) Dominant side → not positive or (b) Dominant side → negative or (c) Nondominant side → relatively positive	Increased probability of assigning preferred stimulus to nondominant side & nonpreferred stimulus to dominant side (forced choice)
Unkelbach (2006)	(a) High cognitive fluency → familiar or (b) Low cognitive fluency → not familiar or (c) Both of the above	(a) Low cognitive fluency → familiar and (b) High cognitive fluency → not familiar	Bias to consider low cognitive fluency stimuli as familiar

Note. The columns "Possible lifetime associations" and "Laboratory training" give ways to conceptualize these contingencies.

To examine this question, the authors conducted training to reverse this hypothesized typical contingency in a group of right-handed participants. In order to alter motor fluency, the investigators encumbered one hand of each participant with a bulky ski glove (with its paired glove attached and dangling from the wrist), as the participant performed a fine motor task involving both hands. In a test phase, the participants were asked which of two side-by-side boxes a character would choose to place animals that the character liked or disliked. The majority of those who had experienced the encumbering glove on their left hand (the "consistent" group) indicated that the liked animal should be placed in the right box, while a majority of those with experience of the right-hand glove (the "reversal" group) placed the liked animal in the left box. In another experiment, the authors gave this animal-box assignment task to a group of naturally right-handed unilateral stroke patients who had experienced interrupted functioning of one hand. Patients with an acquired right-hand impairment responded in the same way as the trained right-glove participants. The investigators suggested that the trained participants and stroke patients had experienced alterations to their side–valence associations through their new motor fluency experiences.

Example 3: Reversing a relationship between cognitive fluency and familiarity

Many cognitive investigations of memory familiarity have revealed the tendency for subjects to equate ease of processing of a stimulus (i.e., fluency) with familiarity. In a series of response-reversal training experiments, Unkelbach (2006) attempted to discover the origins of this association. In one experiment, participants viewed a series of names that were either familiar (from a prior exposure) or unfamiliar; critically, the ease with which subjects processed the names was altered through a perceptual manipulation (i.e., low vs. high color contrast) in order to manipulate contingencies between fluency and familiarity. Manipulating the correlation between color contrast and name familiarity created differing contingencies in two groups ("classic," where familiar names had high contrast, and "reversed," where they had low contrast). In a test phase, the participants performed a recognition test with another set of old and new names that were randomly assigned to have a high or low color contrast. Whether participants had been in the classic or reversal group was found to influence whether they now considered items with a difficult-to-process contrast as old or new.

A second experiment employed a different fluency manipulation: the degree of mental rotation required to make same versus different judgments for pairs of three-dimensional shapes. In training, shapes requiring large mental rotations (low fluency) were more likely to require a same response than those needing smaller rotations. In a test phase familiarity task, participants were shown familiar and unfamiliar words tilted at small or large angles (giving high and low cognitive fluency, respectively), with no correlation between angle and familiarity. The second experiment's results were in line with the first: Training affected familiarity responses in the word task. This was suggested as evidence that the typical human bias for considering fluent items as familiar is formed from learning that fluency is a cue for familiarity.

What Can We Conclude From These Findings?

Response-reversal training studies, such as the three examples reviewed above, have varied goals. Some are motivated by an interest in a hypothesized "lifetime" association, which people may develop gradually over time as they naturally encounter correlations between items in their daily environment, as between motor experience and valence (Casasanto & Chrysikou, 2011) or cognitive fluency and familiarity (Unkelbach, 2006). For others, the primary objective is not to comment on the possible existence of an original association per se but to investigate how it can be changed, such as influencing racial bias through facial muscle configurations (Ito et al., 2006). Such studies investigate effects central to topics in cognitive and social psychology, but the primary question of whether humans currently have a learned relationship between an item in the environment and a response, or whether such a relationship can be altered, is a line of inquiry that has motivated decades of associative learning work.

The three studies we have introduced recognize the applicability of associative learning to their respective research questions to varying degrees. For example, in their study of changing racial bias, Ito et al. (2006) suggested that associations learned for exemplar faces may transfer to other members of that racial group through an associative learning process. Others are less direct in linking their results to learning theory but allude to ideas reminiscent of learning theory concepts, such as the formation of "implicit associations in memory" (Casasanto & Chrysikou, 2011, p. 2) or cues that depend on "ecological correlation" (Unkelbach, 2006, p. 339).

A variety of learned associations can lead to the same pattern of data

Research using associative learning paradigms that closely mirror response-reversal studies has shown that the most intuitive interpretations of response reversals can often be unjustified. To anticipate our upcoming discussion, we note that a broad range of associative changes can all give rise to the same pattern of data. This suggests that the most we can say from a response-reversal result is that the behavior but not necessarily a particular association has changed. An induced reversal can thus be observed without giving an understanding of why or how the original response arose.

The investigation of how organisms adjust their internal representations to reflect relationships between events in the world has produced an impressive array of experimental

approaches and prediction-generating models. It is no exaggeration to say that associative learning is one of the most refined and sophisticated fields of psychology, and far from being irrelevant to today's psychological questions of interest, it has enormous potential for advancing our most important and current lines of inquiry. Although Pavlov's famous experiments focused on the salivary conditioning of dogs (Pavlov, 1927), conditioning is not "all spit and twitches" but instead is "intimately involved in the control of central psychological processes" (Rescorla, 1988, p. 157). Accordingly, conditioning phenomena have been repeatedly observed in humans with stimuli that are very relevant to ordinary human experience (e.g., Collins & Brandon, 2002; Hermans et al., 2005; Iberico et al., 2008; Lanzetta & Orr, 1981; Rosas & Callejas-Aguilera, 2006; Smith, 1979; Vansteenwegen et al., 2005; Vervliet, Vansteenwegen, Baeyens, Hermans, & Eelen, 2005; Walther, 2002), and several recent studies have shown by example that applying learning theory can yield valuable results for cognitive and social psychology areas, including sequential learning (Gureckis & Love, 2010), delusional beliefs in the context of psychosis (Corlett et al., 2007), and attitudes (e.g., Olson & Fazio, 2001).[2] In the pages ahead, we closely examine the implications of associative learning findings for the conclusions of response-reversal training studies.

A Brief Review of Relevant Associative Learning Principles

Before discussing the associative changes that may underlie response reversals, it is necessary to review some principles and findings from the associative learning field. Familiarity with these ideas will be valuable as we move forward in our analysis of the three response-reversal training studies that are the focus of our discussion. Although this review will introduce new concepts for readers who are not already familiar with learning theory, these foundations will give a powerful framework for considering response-reversal results. Table 2 presents examples for some of the concepts we introduce, for readers encountering the learning terms for the first time.

In order to predict and understand the relationships between items in our environment, we must calibrate our internal representations with occurrences of events in the world. This kind of internal representation adjustment to encountered correlations is frequently under examination by response-reversal studies. It is also the focus of the associative learning tradition. Certain stimuli frequently precede or accompany intrinsically rewarding or aversive stimuli. In the associative learning field, predictive cues are known as *conditioned stimuli* (CSs; often tones or lights), whereas intrinsically rewarding or aversive events are *unconditioned stimuli* (USs; often food or shocks).

Table 2. A Summary of Key Learning Concepts With Examples

Learning concept	Associative learning procedure (with examples)	Responses to stimuli	Response-reversal study analogy
Blocking	CS_1 (tone) → US (shock) repeatedly	CS_1 → *Fear response*	—
	CS_1 and CS_2 (light) → US repeatedly	CS_1 → *Fear response* CS_2 → *No fear response*	
Inhibition	CS_1 (tone) → US (shock) repeatedly CS_2 (light) made inhibitor	CS_1 → *Fear response* CS_1 and CS_2 together → *Reduced fear response compared to CS_1 alone*	—
Extinction	CS (tone) → US (shock) repeatedly CS alone repeatedly	CS → *Fear response* CS → *No fear response*	Prior to the study, participants' dominant side (CS) had been associated with strong motor fluency (US). During the study, the dominant side is present without strong motor fluency (Casasanto & Chrysikou, 2011).
Counterconditioning	CS (tone) → US_1 (shock) repeatedly CS → US_2 (food) repeatedly	CS → *Fear response* CS → *Approach response*	Prior to the study, black faces (CS) had been associated with negative affect (US_1). During the study, black faces are paired with smiling-induced positive affect (US_2; Ito et al., 2006).
Discrimination reversal learning	CS_1 (tone) → US (shock) repeatedly CS_2 (light) alone CS_1 alone repeatedly CS_2 → US repeatedly	CS_1 → *Fear response* CS_2 → *No response* CS_1 → *No response* CS_2 → *Fear response*	Prior to the study, high cognitive fluency (CS_1) had been associated with strong familiarity (US), while low cognitive fluency (CS_2) had not. During the study, low cognitive fluency is paired with strong familiarity, while high cognitive fluency is not (Unkelbach, 2006).

Note. CS = conditioned stimulus; US = unconditioned stimulus.

Pavlovian (or classical) conditioning concerns the learned associations between such stimuli. Successfully learning relationships between stimuli allows animals to subsequently predict the occurrences of stimuli—for example, learning that a certain taste or smell (CS) predicts nausea (US). Framing, for example, the association targeted by the discussed Casasanto and Chrysikou (2011) response-reversal study in these terms views a person's dominant side as a CS and the rewarding motor fluency as a US.

Key principles of learning

A central principle of Pavlovian conditioning is the importance of *contingency* between stimuli. An animal forms a Pavlovian association when a relationship of dependence seems to exist between stimuli, as if the animal is seeking out the causal structure of the world. This is very similar to the idea of learning an "ecological correlation," proposed for cognitive fluency and familiarity in one of the discussed response-reversal studies (Unkelbach, 2006). The existence of contingency can successfully predict when learning will occur where simple co-occurrence cannot (Rescorla, 1988), most clearly illustrated by the finding that intermixing additional US presentations into CS–US pairings can eliminate learning (Rescorla, 1968). These extra USs do not change the number of CS–US co-occurrences but greatly disrupt contingency.

Another important principle is the role of error correction during learning (Rescorla & Wagner, 1972). One of the most striking demonstrations of this principle was also one of the earliest: Kamin (1968) showed that although two neutral CSs presented together with a US will each form a CS–US association, if one of the CSs had already formed a strong US association, its presence will now block learning for the other CS (see Table 2 for a blocking example). This is because the presentation of two neutral CSs does not predict the encounter with the US, thereby generating a discrepancy with reality that then drives learning. When one CS already predicts the US, however, a discrepancy is not generated, and so learning does not occur for either CS. This reflects an informational requirement for learning stimuli relationships: associations are formed if there is a gap between internal predictions of stimulus events and the actual events experienced. In other words, surprise drives learning.

The Rescorla-Wagner model

Learning principles have been incorporated into formal associative learning models, one of which will be drawn upon when we discuss the associations that can underlie response-reversal results. This model, created by Rescorla and Wagner (1972), is the most well known and tested of the associative learning models and has "proved to be one of the most remarkable and influential models in psychology" (Lieberman, 1999, p. 140). As well as its enormous contribution to associative learning, the Rescorla-Wagner model has been applied to a variety of other areas in psychology (for a review, see Siegel & Allan, 1996). Understanding this model will aid our later considerations of response-reversal effects. The Rescorla-Wagner model operates by first assigning a total potential *associative strength* to each US. A newly introduced neutral CS starts with zero strength. The CS will gain associative strength as it comes to predict occurrences of a US. More positive associative strength for a CS reflects a stronger expectation that the US will accompany a CS presentation. This expectation is typically seen through a conditioned response to the CS, such as freezing upon hearing a tone that predicts a shock. If there is a mismatch between expectations and encountered events in the environment, the error-correction algorithm at the heart of the Rescorla-Wagner model adjusts associative strengths. Larger discrepancies between expectations and reality produce greater adjustments to the relevant associative strengths, giving greater learning.

A relevant strength of the Rescorla-Wagner model is its ability to model *inhibition*, a concept that will be important in our later discussion. An inhibitory stimulus predicts that a US is less likely to occur when it is present. In the case of aversive events, this can be considered a "safety signal." Inhibitory stimuli counteract typically conditioned CSs (*excitors*) when the two are presented concurrently (Rescorla, 1969; see Table 2 for an example). The Rescorla-Wagner model frames inhibitors as having negative associative strengths, reflecting the expectation that the US is unlikely to occur. When an excitor and inhibitor are presented together, the inhibitor's negative associative strength is said to sum with the excitor's positive associative strength, to lower the total strength. This is important because the total strength (calculated from all the CSs present) gives the overall level of expectation of a US at a given moment. The overall expectation is therefore the result of contributions from all present excitatory and inhibitory stimuli.

There are naturally limitations to the Rescorla-Wagner model (see Miller, Barnet, & Grahame, 1995, for a discussion), which have motivated modifications and alternative models (e.g., Mackintosh, 1975; Pearce & Hall, 1980), but the principles underlying the Rescorla-Wagner model serve as the core of many of the later conditioning models.

How Are Existing Associations Affected by Response-Reversal Training?

Having reviewed some basic concepts in learning theory, we now return to a closer examination of the specific case of response reversals. The conclusions that can and cannot be drawn from the general response-reversal training approach are rarely discussed, perhaps because its applications in cognitive and social psychology span a wide spectrum of areas. Yet, the learning induced through reversal training and its relation to the lifetime associations of interest are of great theoretical and practical importance. For example, the introduction to a *Journal of Personality and Social Psychology* special issue on

implicit prejudice and stereotyping noted: "If such spontaneous processes are so easily moderated, there are likely to be serious implications for models that have argued that automatic processes are well learned and are, as a result, difficult to overcome" (Devine, 2001, p. 758). Despite its importance, the associative basis for reversals in responses is rarely critically examined in training studies.

To fully understand these reversals, we must delve into the learning literature and examine the learning procedures that closely mirror response-reversal training studies. Several experimental procedures exist where a new contingency is introduced that conflicts with an existing association. These have been referred to as the *interference paradigms* (Bouton, 1993), and they allow us to address a question central to interpreting response-reversal results: How is the original association affected by the training? The most straightforward interpretation would be that the original association has been altered or even reversed. As we will discuss, despite the intuitive appeal of this conclusion, learning studies have suggested that a change in contingency between a CS and US will often not remove the original association (Bouton, 2002; Rescorla, 2004). Instead, a new CS–US association can develop, which then comes to influence behavior. We now examine findings from three interference paradigms—extinction, counterconditioning, and discrimination-reversal learning—all of which have analogous response-reversal training studies.

The extinction paradigm

The most basic and studied of the interference paradigms is extinction. In this procedure, a CS that has been associated with a US is now repeatedly presented alone ("extinction trials"). Following a suitable number of extinction trials, the animal's prior response to the CS is eliminated. For example, after a tone–shock association has been learned by a rat, repeatedly presenting the tone without a shock will eliminate the fear behavior that was previously displayed to the tone. To draw a parallel with one of the example training studies considered here (Casasanto & Chrysikou, 2011), this is analogous to extinguishing a learned association between a person's dominant side and positive valence by ensuring the participant's dominant side is not accompanied by strong motor fluency (extinction trials). These extinction examples are presented in Table 2. It could also be argued that more than extinguishing a positive association, this particular study's intervention (a cumbersome glove) generates its own aversive experience, which then becomes paired with the dominant side. This possibility (counterconditioning) and ways to empirically determine which process is occurring will be discussed in later sections of this article.

Although extinction produces a response change, an array of findings suggests that this does not reflect a reversal of the underlying association (Bouton, 2000; Delamater, 1996; Rescorla, 1996). This counterintuitive conclusion is perhaps most clearly demonstrated by observed resurgences of behaviors that have been extinguished. For instance, presenting trials of the US alone after an extinction procedure can produce a resurgence of the "extinguished" behavior when the CS is next presented (*reinstatement*; Rescorla & Heth, 1975), showing that the original association was not removed during extinction and that it can still influence behavior. This phenomenon has been reported in humans; for example, presenting participants whose fear responses have been extinguished with four trials of the US on its own produces reinstatement of the original conditioned fear response (Hermans et al., 2005). In other supporting findings, the mere passage of time between extinction and testing can lead to *spontaneous recovery* of behaviors that reflect an association, despite prior extinction (Pavlov, 1927; see Rescorla, 2004, for a discussion), and presenting an extinguished CS in a new context can renew an extinguished behavior (*renewal*; Bouton, 2004). Renewal effects have been repeatedly recorded in human participants (e.g., Van Gucht, Vansteenwegen, Beckers, & Van den Bergh, 2008; Vansteenwegen et al., 2005; Vila & Rosas, 2001).

Given these behavioral resurgences, what produces the suppression of responses during extinction? A variety of evidence suggests that exposure to the new set of contingencies in an extinction procedure generates a second association with inhibitory properties (Bouton & Nelson, 1994; Pavlov, 1927).[3] This inhibitory association combines with the original excitatory association, canceling it out, resulting in an overall expectation of "no US" and no response to the CS. In the framework of the Rescorla-Wagner model, the original association's positive associative strength sums with the newly formed negative associative strength, to give a total strength of zero, equivalent to having no expectation for the US (Rescorla & Wagner, 1972).

The counterconditioning paradigm

Extinction is one of the most widely studied of the interference paradigms, but some response-reversal training studies go one step further: more than removing a CS's correlation with a US, they introduce a relationship with a new US that has an opposing valence. This is analogous to a *counterconditioning* procedure. In this paradigm, a CS is first conditioned with a US from one motivational set (e.g., CS–food) and subsequently paired with a US from another (e.g., CS–shock; see Table 2 for an example). This could be considered akin to some of the training studies discussed here. For example, rather than viewing the Casasanto and Chrysikou (2011) experiments as extinction, the study could be framed as seeking to change a postulated lifetime dominant side–positive association by now pairing the dominant side with the negative affect produced from encumbering the dominant hand. The study by Ito et al. (2006) took this approach but with a negative-to-positive induction: They suggested an existing negative association with black faces and aimed to reduce this bias by pairing black faces with a smiling muscle configuration. Whether a particular training study is better considered

extinction or counterconditioning depends on how a new stimulus pairing interacts with existing associations. For example, in the Casasanto and Chrysikou (2011) study, introducing a glove during training may simply remove the dominant hand's normally rewarding motor dexterity (extinction) or may produce frustration or clumsiness that is itself aversive (counterconditioning). Approaches to disentangling these alternatives will be discussed later in this article.

As with extinction, there is evidence that counterconditioning does not remove original learning. Repeatedly pairing a CS with a rewarding US, after it had been paired with an aversive US in a prior context, at first leads the CS to signal the new rewarding US, but the originally learned aversive response can renew if the animal is returned to the prior context (Peck & Bouton, 1990). Similarly, presenting the animal with the original US on its own can lead to reinstatement of the pre-counterconditioning response (Brooks, Hale, Nelson, & Bouton, 1995), and spontaneous recovery can occur after a period of time (Bouton & Peck, 1992; Rescorla, 1997), all suggesting that the original learning was not overwritten.

The discrimination-reversal learning paradigm

A final relevant interference paradigm is *discrimination-reversal learning*, where a CS is first associated with a US (e.g., tone–shock), while another CS remains unpaired (e.g., light–no US; shown in Table 2). In a subsequent training session, the CS associations are swapped (e.g., tone–no US, light–shock), producing an equivalent change in the responses given to each CS. Unkelbach's (2006) cognitive fluency study may be more analogous to this paradigm than the other two example studies. The training in this study was intended to reverse a lifetime association between high cognitive fluency and familiarity by pairing low cognitive fluency with familiarity, and high cognitive fluency with no familiarity: a switch from one CS to another. In a similar manner to the previous two interference paradigms, a shift in context after "successful" discrimination-reversal learning leads to renewal of the previous responses to each CS, again suggesting that the original association is not replaced (Bouton & Brooks, 1993; Dekeyne & Deweer, 1990; McDonald, King, & Hong, 2001; Spear, 1971).

Summary

It is clear that for a variety of changes to the contingencies between stimuli in existing associations, the original associations remain intact, despite a dramatic shift in behavior after the procedure. This has repeatedly been shown for experimental paradigms that closely resemble human response-reversal training studies, including extinction, counterconditioning, and discrimination-reversal learning. Delamater (1996) succinctly described the resilience of original associations following his own attempts to alter them: "specific Pavlovian S-O [stimulus-outcome] associations, once acquired, are rather insensitive to a host of manipulations designed to undermine those associations" (p. 448).

The Results of Response-Reversal Studies Cannot Tell Us Which Associative Strengths Have Been Changed

We will now discuss why it is not possible to tell which associative strengths have changed from observing trained reversals in responses. In the previous section, we outlined evidence that shifts in responses from a contingency change may not reflect alterations to an original association. While a change to a hypothesized original association could theoretically account for induced response shifts, aside from being very difficult to produce (as described above), it is only one of a host of associative changes that can lead to the same result. We now describe three sets of associative changes that would lead to a reversal in responses: a context-moderated new association, a second CS forming an association with the US, and other CSs protecting the target CS from change. In each described case (summarized in Table 3), the same results would be recorded but due to very different underlying associations. Notably, the data from response-reversal training studies cannot distinguish between these various alternatives.

The context acts as an occasion setter

The context of learning is known to play a particularly important role in Pavlovian conditioning (Bouton, 1993, 2004). Less widely appreciated are the diverse forms that context can take. As well as its more typical forms, such as the background, room, or lighting, context can include interoceptive stimuli: Context-related effects have been recorded for the presence of drug states, hormonal states, moods, deprivation states, expectations of events, and the passage of time (see Table 1 in Bouton, 2000,). For example, Mystkowski, Mineka, Vernon, and Zinbarg (2003) demonstrated that a caffeine state can act as a context in humans, in a placebo-controlled study.

The influence of context is particularly important for this discussion. Context appears to play an important role in extinction and counterconditioning (for a review, see Bouton, 1993), two of the learning paradigms that are analogous to the response-reversal approach. One perspective on context's influence suggests that "an extinguished CS is ambiguous, with two available 'meanings,' either of which can be cued by the right context, exactly like an ambiguous word" (Bouton, 2000, p. 57). In this fashion, the context acts as a moderator that selects between two interpretations of a CS. In the associative learning literature, a stimulus selecting between relevant associations is known as *occasion setting* (Holland, 1983)—a role that can be taken by the learning context itself (Swartzentruber, 1991). This perspective suggests that a new association can form when a new contingency is presented.

Table 3. A Summary of the Three Discussed Alternative Sets of Associations

Concept	Learning processes	Learning literature example	Third training example (Unkelbach, 2006) analogously framed	Definitions
Occasion setting	Context moderates the development of a new inhibitory association between the target CS and US, with alternative associations selected by occasion setting.	Training: Context A: tone–shock; Context B: tone–no shock. Testing: greater shock-related responses to the tone in Context A than in B. The context moderates the actual relationships: Context B did not simply become an inhibitor for shock. (Bouton & Swartzentruber, 1986)	Training: Context A (lifetime): high cognitive fluency–familiar; Context B (lab): high cognitive fluency–not familiar. Testing: increased bias to consider highly cognitively fluent items as not familiar in lab	*Occasion setting:* When a CS or context "sets the occasion" for which one of two different CS–US associations apply.
Inhibition/ configural learning	A second CS develops an inhibitory association with the US, or a context-and-CS configural stimulus forms and develops an inhibitory association with the US.	Training: tone–food; [tone and noise]–no food. Testing: less food-related activity to the tone and noise together (in configuration) than to the tone or noise alone. (Wilson & Pearce, 1992)	Training: Lifetime: high cognitive fluency–familiar; Lab: [high cognitive fluency through rotation] configuration–not familiar. Testing: bias to consider [high cognitive fluency through rotation] configuration as not familiar.	*Configural stimulus*: A stimulus composed of multiple CSs, which acts as an independent CS. This contrasts with a compound, where simultaneously presented stimuli act separately.
Protection from extinction	(a) The target CS and another CS both form inhibitory US associations that overshadow each other, giving partial protection from extinction. (b) The partial reinforcement extinction effect further divorces response levels from the strength of the target association.	(a) Training: white block–shock; white block and tone–no shock. Testing: diminished extinction of the white stimulus after being presented with the (protecting) tone during the extinction procedure (Lovibond et al., 2000) (b) Training: noise–airpuff to eye (partial vs. continuous reinforcement); noise–no airpuff. Testing: resistance to extinction in the partial reinforcement group (Leonard, 1975)	Training: Lifetime: high cognitive fluency–familiar (partial reinforcement); Lab: high cognitive fluency–not familiar and large rotation–familiar (continuous reinforcement). Testing: bias to consider highly cognitively fluent and less rotated items as not familiar.	*Overshadowing:* CSs share the total change in associative strengths when simultaneously paired with a US, reducing the acquired associative strength of each. *Protection from extinction:* A stimulus presented with a target CS during extinction can become inhibitory and reduce a target CS's extinction-induced drop in associative strength. *Partial reinforcement extinction effect:* Learning an association through partial reinforcement can give its response resistance to later extinction.

Note. Two examples of each are provided: first, a simple example drawn from the animal learning literature from which these ideas were developed; and second, an analogous example derived by reframing the manipulations of Unkelbach (2006). CS = conditioned stimulus; US = unconditioned stimulus.

Each "meaning" of the CS (original or new) is then selected based on other cues or the context.

Of particular interest for our discussion, Bouton has suggested that the very occurrence of change to a cue's meaning (through a contingency shift, for example) produces ambiguity, which then drives an animal to direct attention to the context (Bouton, 2002), in turn making subsequent conditioning context specific (Bouton, 2002; Rosas & Callejas-Aguilera, 2006). Response-reversal training studies are therefore positioned to be particularly susceptible to occasion-setting effects: When participants are brought into the lab, they are presented with a new contingency for an existing CS. As soon as this CS–US contingency changes through training, ambiguity is introduced for the meaning of the cue, increasing attention to

the context. Any subsequent conditioning is then likely to become context dependent. This can explain why original associations (which do not develop from a change to an already established contingency) are generally robust to context changes, while associations generated in extinction (Bouton, 2004), or through any new overlaying learning (Nelson, 2002), are vulnerable to context shifts.

In response-reversal training studies, the occasion-setting context could take a number of forms. For example, in a study pairing a positive experience with images that normally trigger racial bias, a continually present stimulus such as the computer screen or format of the stimuli (two-dimensional images of faces) would be excellent cues for the participant that the previously learned contingency has now shifted (e.g., Gawronski, Rydell, Vervliet, & De Houwer, 2010). The particular state of the learner could also act as the context governing a relationship (Bouton, 2000). For example, Casasanto and Chrysikou (2011) showed that after experiencing a stroke handicapping their dominant right hand, patients were more likely to associate their left and right sides with positive and negative valences, respectively. A number of cues were available to act as a context at the time of their stroke (when a handedness–valence contingency may have changed), including altered motor control or some interoceptive change after this severe neurological event. This new context could now serve as a reliable indicator that a new contingency is now relevant (with the previous association remaining unchanged).

The evidence discussed suggests that a new context-controlled association can develop when a CS–US contingency is changed. This new association can then influence behavior specifically in the newly established context, while the original association remains intact and unaltered.

A new CS forms an association with the US and influences the same behavioral system

When a training session successfully affects a participant's responses, the possibility exists that a different (nontarget) CS has been learned as a cue, which then triggers the same response system hypothesized for the target CS. This possibility has been considered in some response-reversal studies. For example, in Unkelbach's (2006) investigation into cognitive fluency and familiarity, Experiment 1's discussion noted that the training-induced response change could have resulted from participants learning that color contrast (the property varied to manipulate cognitive fluency), rather than fluency per se, can predict familiarity. This motivated a second experiment. Here, the training manipulated a cognitive fluency–familiarity correlation by influencing the relationship between the amount of mental rotation required to judge two three-dimensional shapes as being the same or different and the shapes actually being the same. Experimenters then assessed a posttraining fluency–familiarity bias by examining whether (random) in-plane rotations of words would bias responding in a word familiarity test. This training-to-testing shift from shapes to words was greater than in Experiment 1, but rotation was a common factor: The reversal group was trained to judge highly rotated items as being the same and subsequently showed a bias toward considering highly rotated text as familiar. In this example, although the final behavioral result could indeed have been due to a relationship with cognitive fluency, rotation may also have acted as a cue, in the same manner that color contrast could have in Experiment 1. This possibility assumes that stimulus-independent rotation can be conditioned, but this seems likely given that associations between rotation and other visual properties can be learned (Backus & Haijiang, 2007). If another CS does become associated with the US during training, the conclusions that can be drawn about a hypothesized original CS–US pairing are limited, as another CS entirely is influencing the recorded responses.

A close examination of the results from another of the example training studies discussed here is consistent (although not conclusive) with an additional cue contributing associative strength. In Casasanto and Chrysikou's (2011) investigation, 77% of the right-handed participants who received the encumbering glove on their nondominant left hand (the consistent group) associated positive valence with their right side in the test phase. Compare this result with Casasanto's (2009) Experiment 3, where the same testing experiment was conducted but without any prior training. Here, 58% of right-handed individuals associated positive valence with their right side. The glove training appears to have increased the probability that participants would show a bias toward positive valence in their dominant side. The presence of an additional trained CS can account for this increase, as the extra CS's associative strength would summate with any existing bias to increase the typical response. It is noted that the training study's higher percentage can be no more than suggestive due to limitations in comparing numerical results across studies and because other explanations could also account for an increase, such as a switch from partial reinforcement during lifetime exposures to intense continuous reinforcement during the training; this factor is discussed later in this article.

The risk of another CS forming an association is compounded by the possibility that the context and target CS will form a *configural stimulus*. Configural theories argue that two or more CSs, or a CS and context (Darby & Pearce, 1995), can, as a configuration, form a single association with the US (Pearce, 1987). The context is most likely to be included in such a configuration when it is salient to the subject (Darby & Pearce, 1995). As subjects pay greater attention to the context as soon as new overlaid conditioning begins (discussed above), response-reversal training may be vulnerable to associations forming that involve a new CS-plus-context configural stimulus rather than the target CS itself.[4]

This second set of associations would produce a response reversal through a different CS. The resulting response reversal would not be drawing on the CS that motivated the study or at best would reflect an association involving a new configuration that includes the CS.

A lifetime CS–US relationship is protected from change

The previous suggestion illustrates our main argument, namely, that a behavioral change can be elicited (in this case driven by a new CS or configuration), with no change occurring to the original association. A less dramatic alternative is that an additional CS and the target CS share responsibility for predicting a US. A variety of factors influence the eventual strength of an association formed between a particular CS and US. One such factor is the concurrent presence of other predictive CSs. As reviewed above, some associative learning models suggest that discrepancies between an animal's predictions and reality drive changes to the associative strengths of relevant cues (Rescorla & Wagner, 1972). The total change in strength produced from a given trial or encounter is shared among the CSs that are present at that particular moment (Rescorla & Wagner, 1972).

Because strength changes are shared among CSs, intermixing presentations of another CS while a target CS is being extinguished can give the target CS a degree of protection from extinction (Chorazyna, 1962). If the intermixed CS starts out as neutral to an animal, it will gradually acquire inhibitory properties during the extinction procedure, as the animal learns that this CS can reliably predict that the US will not occur. Through sharing the overall drop in associative strength in this way, the added CS can leave an original CS with some positive strength even when the behavior is extinguished (Lovibond, Davis, & O'Flaherty, 2000). Introducing a CS that is already inhibitory to extinction can even offer complete protection for the target CS: The inhibitory CS can completely account for the absence of the US during the extinction phase, leaving the original CS's association unchanged (e.g., Rescorla, 2003).

Applying this idea to response-reversal training studies, a neutral CS that is presented alongside a target CS, such as the rotation example discussed above, could become inhibitory during the reversal training, thereby changing a person's responses without any substantial change to the (protected) target CS.

The discussed protection effect muddies the link between the recorded behavioral responses and a target CS's associative strength. This link may be further weakened by the original association's prior pattern of reinforcement. Response-reversal training paradigms frequently present stimuli in a continuous reinforcement schedule, where every CS is paired with a US. For associations acquired from the environment, however, the US is unlikely to have followed every CS presentation. For example, for the side–valence association proposed by Casasanto and Chrysikou (2011), the sides of space will not always be accompanied by a particular valence. Instead, lifetime associations are likely formed from a partial reinforcement schedule. Response patterns that have developed from partial reinforcement have been found to possess resistance to extinction, compared with those formed from continuous reinforcement (e.g., see Leonard, 1975, for a Pavlovian conditioning demonstration in humans). The influence of this *partial reinforcement extinction effect* (see Pipkin & Vollmer, 2009, for a relevant discussion) on an existing lifetime CS–US association may further blur the link between the original association and the behavioral responses recorded during response-reversal training.

Summary

The potential sets of associations reviewed here suggest that a shift in a participant's responses after training can reflect the influences of a variety of different underlying associations. Although not exhaustive, the possibilities discussed can all lead to changes in responses without a substantial change to the original CS–US association. It is worth noting that the sets of underlying associations we have discussed are not mutually exclusive and may interact. For example, a configural stimulus could develop and then take an occasion-setting role for other associations. The discussed alternatives are summarized with examples in Table 3.

In this section, we demonstrated a many-to-one mapping: A variety of processes and associations can produce the same response shift. This suggests uncertainty about the mechanisms that produce a response change in the response-reversal paradigm and limits the conclusions we can draw about an association of interest—either about its origins or its current state. After this kind of reversal training, the most we can conclude is that the behavior has changed from the training. For researchers who purely wish to modify participants' behaviors without making a theoretical point, this effect may be satisfactory (although knowledge of learning principles and alternative associations may still be invaluable, as they can be manipulated to an investigator's advantage, which we discuss in the next section). For response-reversal studies that are motivated to understand the origins of a typical response, however, the recorded behavioral change is intended to be a way to understand an existing naturally developed relationship. Given the preceding discussion in which we showed how such behavioral changes may not be informative about the original association, readers may wish to be cautious in drawing too strong conclusions. For example, several subsequent studies have interpreted these results as suggesting that response-reversal training can "reverse . . . implicit associations of good with right and bad with left" (De La Fuente, Casasanto, Román, & Santiago, 2011, p. 2620) and that a behavioral reversal means that an explanation (for how a response pattern typically develops) is now "validated" (Brookshire & Casasanto, 2011, p. 2611; Casasanto, 2011, p. 1259). Caution may be warranted given the potential for observing a response change through such a variety of mechanisms. The variety of possible mechanisms also speaks to the larger general point that a behavioral result is the outcome of a process or representation that may not be apparent from a recorded response alone. It can be easy to underappreciate the

uncertainty present in the translation from representation to behavior, but it is of particular importance for investigations that manipulate a behavior in the lab.

The important points of our argument have now been presented. For those interested in preserving behavioral changes or in understanding existing associations further, we next draw on associative learning ideas and findings to address these issues, requiring us to delve further into associative learning concepts. We now turn to strategies for prolonging any induced behavioral changes and then to the question of how to alternatively investigate existing associations, drawing on learning theory to inspire some novel approaches.

How Can You Preserve an Induced Change in Behavior?

Understanding the character of a naturally developed association is of theoretical importance to many cognitive and social psychology areas. For some associations, however, investigators may be more interested in finding ways to preserve any training-induced response reversals. In these cases, the resilience of a response reversal, rather than how it developed, is of primary importance. Associative learning theory can inform this search for effective training interventions. For example, the field of clinical psychology has successfully drawn on associative learning evidence to help change behaviors, such as in treating phobias (Laborda, McConnell, & Miller, 2011). How we measure "effectiveness" will naturally help determine the best approach to take. For example, approaches that induce stronger response reversals during training can actually conflict with the aim of maintaining an induced response shift over time.

One approach to implementing and prolonging behavioral changes is to accept that training sessions may create new associations (leaving the original relationship intact) and work toward maintaining the new learning as much as possible. Once we are aware of the different forms of new learning in response-reversal training paradigms, we can work at enhancing associations with potential for having prolonged effects while diminishing others. Some investigators in the clinical field have embraced this idea when they advocate a focus on the learning that takes place during exposure treatments (e.g., Lang, Craske, & Bjork, 1999). We now draw on learning theory to describe a selection of possible enhancement strategies for investigators interested in maintaining induced response shifts.

Emphasizing the target CS

The particular characteristics of a training session can be examined and adjusted to increase the longevity of the induced behaviors. One central component of a training session is the type of stimuli employed and the task used to present them. As Lang et al. (1999) have noted, varying a task gives the benefit of pairing more cues with a desired US, increasing the opportunities to retrieve the new associations in the future. An additional benefit of introducing variation into training is to promote learning involving the target CS: Ensuring the CS remains constant while the task and other parameters are varied will increase the probability that the target CS is the most reliable cue. Ito et al. (2006) followed this approach when they presented faces of different individuals during training to encourage generalization to the racial group as a whole. Nevertheless, this can be taken further. In this example, if the essence of the CS is a racial group, presenting video clips of individuals, typical names given to members of the group, and other varied stimuli in an intermixed fashion may encourage the common feature (the racial group) to form the strongest relationship with the US by outcompeting other cues.

Undermining other cues' predictive powers can further help the target CS in this competition. For example, pairings of a nontarget CS and the US can be intermixed with a target CS's extinction trials, to strengthen the inhibition developed by the target CS during extinction. In a counterconditioning paradigm, a nontarget CS can be repeatedly presented without any reinforcement, to reduce its power for predicting the opposing US. For example, to inhibit learned racial bias, pairings between black faces and an induced smile could be intermixed with other faces presented without an induced smile, to reduce the reliability of general face cues, the computer screen, and other simultaneously present CSs. To create an even stronger association, the nontarget CSs can be paired with an opposing (in this case, negative) US.

Introducing stimuli into extinction schedules

Another factor determining the longevity of new learning is the nature of the extinction schedule. Ricker and Bouton (1996) have suggested that the level of reacquisition (i.e., how rapidly extinguished responses reappear when an animal is reexposed to the original contingency) is influenced by the learned relationship between a given trial and the general presence or absence of reinforcement on the next trial. In a typical extinction schedule, a lack of reinforcement reliably signals subsequent nonreinforcement. In contrast, in a continuous reinforcement schedule, the presence of reinforcement always signals future reinforcement. Bouton, Woods, and Pineño (2004) reported that introducing reinforced trials or simply lone US presentations into an extinction procedure helped slow future reacquisition after extinction. They hypothesized that introducing reinforcement trials into the extinction process creates pairings between "reinforcement" and the "nonreinforcement" of extinction. When an animal then inevitably encounters reinforcement in the future, this signals nonreinforcement, and the effects of extinction are once more retrieved. Adding US trials in this manner would reduce the magnitude of any response drop during the extinction procedure, but the induced response change may last longer.

The idea that a trial can signal a subsequent trial's reinforcement presents another strategy for prolonging training

effects. A training schedule could be amended to include presentations of cues that are found in the posttraining environment, in order to later cue retrieval of the training session. For example, if one wished to prolong a new left side–positive valence association (in right-handed individuals), images of a left hand can be intermixed into training. Participants' left hand will frequently be seen when they make judgments in their left visual field, so including this stimulus during extinction may allow it to later cue the trained association, in much the same way that context can cue a particular CS–US contingency.[5] Consistent with this idea, Brooks and Bouton have shown that presenting a cue during testing that had been present during the extinction procedure can attenuate spontaneous recovery (Brooks & Bouton, 1993) and renewal (Brooks & Bouton, 1994). As this cue does not appear to acquire excitatory or inhibitory properties itself, it may be a signal for the nonreinforcement context of extinction. A similar result has been reported in humans, by using a distinctive pencil and clipboard during extinction (Collins & Brandon, 2002). Subsequently employing these cues in testing attenuated the typical renewal of the original responses. A related effect has been found from participants mentally reinstating the extinction context, suggesting that these cues need not be physical objects (Mystkowski, Craske, Echiverri, & Labus, 2006).

Partial reinforcement

Response-reversal training studies often present new contingencies in an intense continuous reinforcement schedule. As we have discussed above, however, a partial reinforcement schedule can give resistance to future extinction. Just as partial reinforcement can give resistance for an original association, it could also prolong new learning through the addition of unpaired CS trials into training. Survival of the partial reinforcement-extinction effect after context shifts (Boughner & Papini, 2006) suggests that this could continue to be effective once the participant has left the training context.

Increasing the quantity of trials

A more basic (although not always practical) strategy is to conduct a very large number of extinction trials. Exposing rats to numerous extinction trials can reduce the renewal of original responses that is typically found after a context shift (Denniston, Chang, & Miller, 2003), suggesting that this coarse approach may be effective. Unfortunately, a substantial number of trials may be required to see an effect (see Laborda et al., 2011, for a discussion), making this approach impractical for many human investigations.

Summary

We have suggested a selection of strategies for outcome-focused investigators seeking to prolong the life of a response reversal recorded in the lab.[6] Many of these approaches acknowledge that new associations may drive behavioral reversals and focus on giving this new learning longevity outside of the training context. The associative learning field contains many further tried-and-tested methods for the cognitive and social psychology investigator looking to maintain trained response patterns.

Alternative Ways to Test for an Association

Our previous discussion illustrated that changing a contingency through intense training can produce response changes without changing an existing association of interest. Our main conclusion from our review of the learning literature is that one cannot make inferences about an original association—its origin or its current status—based on response-reversal training. So where does this leave the investigator wishing to understand a relationship between stimuli? Fortunately, we can use associative learning principles to further understand different components of an association without attempting to reverse it. We now suggest some possible approaches that utilize learning principles such as summation, blocking, and transfer.

Summation

One approach to assessing an existing association is to measure the summation effect from presenting the hypothesized CS simultaneously with a CS that has previously been conditioned to the suspected US. If the target CS is indeed associated with a US, pairing it with a preconditioned CS should produce a greater US-relevant response than when either is presented alone. Orr and Lanzetta (1984) successfully applied this technique when they examined the possible associative properties of facial expressions. After repeatedly pairing a tone with a shock, they presented participants with a compound of the tone and a happy, fearful, or neutral facial expression, with control groups receiving nonface stimuli. Presenting the tone and fearful expression together produced a skin conductance response that was higher than any other tone-and-stimulus pairing, suggesting that the fearful expression was an excitatory CS with an associative strength that summated with the tone's.

Pairing a proposed CS with a CS that has been conditioned in an experimentally controlled procedure can further improve our understanding of the precise nature of a hypothesized CS–US relationship. The testing methods suggested here can help disentangle whether training procedures could be considered closer to extinction or to counterconditioning. For example, we could examine whether two USs have opposing valences, a question that could be asked of dominant motor fluency and glove-induced frustration for the second response-reversal study. A summation test could be performed in this instance, by simultaneously presenting a CS that is lab conditioned to predict one US, with the CS hypothesized to predict an opposing US. If these two USs are indeed opposing (as expected for

counterconditioning), the net behavioral response should be lower than when the lab-conditioned CS is presented alongside a neutral cue.

Blocking

An association may also be tested with the blocking paradigm. A conditioned CS can block (or overshadow) learning for a new CS if the two CSs are presented simultaneously with the US (Kamin, 1968; see Table 2 for an example). If a hypothesized CS has developed an excitatory association with a US (through a natural contingency in the environment), it should block, or partially block, learning between a new CS and that same US. Lanzetta and Orr (1981) used this paradigm to examine possible excitatory associations with happy and fearful expressions. They found that fearful expressions blocked conditioning between a neutral cue (tone) and a shock. In contrast, the tone blocked learning between shocks and a happy expression.

Transfer

A further approach to examining an existing association is to conduct a Pavlovian-Instrumental Transfer Test. This test assesses whether a CS's prediction of a US can transfer to an instrumental response. Instrumental conditioning concerns associations between actions and USs, where performing an action actually influences whether a US will occur. For example, presenting a food reward after every press of a lever will condition a rat to rapidly press that lever. An excitatory association between a CS and US can transfer to a conditioned instrumental response if the CS and instrumental response both typically lead to the same US (Delamater, 1996). For example, a rat can be trained with a CS–reward pairing and independently trained to associate a lever press with receiving the same reward. Presenting the rat with the CS and lever together (nonreinforced during testing) produces a greater rate of lever pressing than does presenting the lever alone. Delamater (1996) has demonstrated that this transfer effect can reflect an originally paired US, even after extinction. Delamater first conditioned rats by pairing a noise with pellets and pairing a light with sucrose (counterbalanced across groups). The rats were also trained to associate responding on a lever and chain with one of the two rewards. Presenting the noise or light (while the animals had the opportunity to produce either instrumental response) led to increased responding on the particular action that had previously led to the same US as the presented CS. This transfer was US specific, even when both USs were rewarding. It also remained after the associations were extinguished.

The sensitivity of this paradigm to the existence of an original association suggests that it could be of interest to researchers looking to test for a lifetime association. To return to one of the training studies discussed here, we note that the dominant side–positive association posited by Casasanto and Chrysikou (2011) could be examined in a transfer test. An example of a possible approach would be to first establish an instrumental response with a reward. Aharon et al. (2001) have shown that participants will exert effort (measured through key presses) to see attractive faces, a stimulus that activates classic brain reward pathways. After training participants with this contingency, a study could present attractive faces in participants' dominant side versus nondominant side. The proposed existing association between participants' dominant side and positive valence should summate with the reward strength of an attractive face, causing participants to exert more effort to hold an attractive face on the display when it is presented on their dominant side. As this particular lifetime association is hypothesized to actually result from typical hand use, investigators would need to use an unconventional method of responding to avoid concurrent motor fluency interference.

It is interesting to note that the level of transfer recorded in this paradigm is sensitive to the strength of a Pavlovian CS–US relationship (Delamater, 1996), creating the potential to compare the relative strengths of alternative CSs for a given US. In the previous example, repeating the study for egocentric space (right vs. left visual field) and allocentric space (right vs. left side of a space, independent of visual field) could prove informative about the precise nature of the naturally learned CS.

Summary

The approaches to examining existing associations that we have described here do not simply avoid some of the pitfalls of the response-reversal training approach. They also have potential for elucidating features of associations that would not otherwise be accessible, such as the characteristics of a CS or the strength of a relationship. Creatively combining these and other methods from the associative learning field can give a powerful set of resources for investigators interested in naturally learned associations.

General Summary

The goal of this article was to provide a targeted application of animal learning principles to the relation between changes in behavior and changes in underlying associations following a training manipulation. The most important take-home point can be summarized simply: There are many possible associative changes than can underlie an induced behavioral change; therefore, caution is necessary when interpreting the cause of the change. If the reader takes only this point away, we will have succeeded. However, we hope we will inspire some readers to think about ways to harness the principles of learning theory in order to make stronger inferences about learned associations and also to develop more powerful paradigms for changing behavior.

Investigations into associations that develop over a person's lifetime have the opportunity to utilize learning

paradigms in novel ways to ask new and probing questions about particular associations of interest, which may otherwise remain a black box. Viewing trained associations with the perspective of learning theory also gives cognitive and social psychologists an array of possible techniques to prolong and strengthen any new learning. The associative learning field has a long and successful history of designing and interpreting studies that can ask interesting questions about learned associations. Investigators in the cognitive and social psychology fields may find that applying associative learning theory to their questions of interest yields valuable and unexpected rewards.

Authors' Note

We would like to thank Robert Rescorla, Russell Epstein, Joseph Kable, David Kraemer, Barry Schwartz, the anonymous reviewers, and the editor for their helpful comments on earlier versions of the article.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. The IAT is used to measure implicit racial bias by recording reaction times as participants respond to stimuli of interest (in this case, black and white faces) as well as positive and negative words. The stimuli and words are mapped onto the same set of response keys. A pattern of faster responses when "white" shares a key with positive (and "black" with negative) words, compared with the reverse, has been suggested to reflect implicit bias against the black group (Greenwald, McGhee, & Schwartz, 1998), although this interpretation is debated (e.g., Rothermund & Wentura, 2004).
2. It is worth noting that a current debate in this area centers on what has been referred to as *evaluative conditioning* (EC; De Houwer, 2007; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010), where a CS acquires the evaluative qualities of another stimulus. For example, the face of a neutral person may acquire a positive association if paired with the face of a liked person. Notably, the EC term is typically restricted to when the US in a pairing has an affective association without itself being rewarding or aversive in the classic concept of a US (Walther & Langer, 2008, p. 102). An active debate concerns whether EC is a unique form of conditioning (Walther & Langer, 2008) or is governed by established Pavlovian conditioning principles (Lipp & Purkis, 2005).
3. Overlaid inhibition can account for phenomena such as spontaneous recovery if the inhibitory relationship weakens over time (Rescorla, 2004), a tendency that has indeed been documented for conditioned inhibition (e.g., Hendersen, 1978). It should also be noted that although inhibition seems to be a key factor in extinction, as Rescorla has discussed, "it seems almost certain that the response decrement that is observed in extinction itself has multiple contributors" (Rescorla, 2004, p. 508). For example, Colwill (1991) and Rescorla (1993, 2001) have explored a similar factor, where new inhibitory learning forms between the stimulus and response, rather than CS and US per se.
4. Some solace may be found in suggestions that learning involving a configural stimulus is likely to generalize to some degree to its components (Darby & Pearce, 1995), leaving open the possibility of some indirect learning involving a target stimulus. The prior discussion on alternative CSs is still unfortunately relevant, however, so a context-and-CS configural stimulus could form with an unintentional CS.
5. This type of cue should not be presented before every extinction trial, as this would likely condition it to become an inhibitor, protecting the target CS from extinction in the process (Rescorla, 2003).
6. It should be noted that some approaches to prolonging extinction effects that could be predicted with learning theory have given mixed results. Spacing out extinction trials has sometimes shown positive effects for prolonging extinction (e.g., Barela, 1999; Urcelay, Wheeler, & Miller, 2009), but the advantages of spaced versus massed presentations appear to vary between paradigms (for discussions, see Laborda, McConnell, & Miller, 2011; Rescorla, 2004). It might also be predicted that conducting extinction training in multiple contexts would give resistance to context-driven relapses, but whereas some studies have given promising results (e.g., Gunther, Denniston, & Miller, 1998), others have not (e.g., Bouton, García-Gutiérrez, Zilski, & Moody, 2006). Bouton et al. (2006) have suggested that extinguishing in multiple contexts may be helpful only if the original association was learned across relatively fewer contexts. As lifetime CS–US relationships are likely encountered across many different environments, conducting training sessions across a larger number would be infeasible.

References

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: FMRI and behavioral evidence. *Neuron*, *32*, 537–551.

Backus, B. T., & Haijiang, Q. (2007). Competition between newly recruited and pre-existing visual cues during the construction of visual appearance. *Vision Research*, *47*, 919–924.

Barela, P. B. (1999). Theoretical mechanisms underlying the trial-spacing effect in Pavlovian fear conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*, 177–193.

Boughner, R. L., & Papini, M. R. (2006). Survival of the partial reinforcement extinction effect after contextual shifts. *Learning and Motivation*, *37*, 304–323.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, *114*, 80–99.

Bouton, M. E. (2000). A learning theory perspective on lapse, relapse, and the maintenance of behavior change. *Health Psychology*, *19*(Suppl. 1), 57–63.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*, 976–986.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory*, *11*, 485–494.

Bouton, M. E., & Brooks, D. C. (1993). Time and context effects on performance in a Pavlovian discrimination reversal. *Journal of Experimental Psychology: Animal Behavior Processes*, *19*, 165–179.

Bouton, M. E., García-Gutiérrez, A., Zilski, J., & Moody, E. W. (2006). Extinction in multiple contexts does not necessarily make extinction less vulnerable to relapse. *Behaviour Research and Therapy*, *44*, 983–994.

Bouton, M. E., & Nelson, J. B. (1994). Context-specificity of target versus feature inhibition in a feature-negative discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *20*, 51–65.

Bouton, M. E., & Peck, C. A. (1992). Spontaneous recovery in cross-motivational transfer (counterconditioning). *Animal Learning & Behavior*, *20*, 313–321.

Bouton, M. E., & Swartzentruber, D. (1986). Analysis of the associative and occasion-setting properties of contexts participating in a Pavlovian discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*, 333–350.

Bouton, M. E., Woods, A. M., & Pineño, O. (2004). Occasional reinforced trials during extinction can slow the rate of rapid reacquisition. *Learning and Motivation*, *35*, 371–390.

Brooks, D. C., & Bouton, M. E. (1993). A retrieval cue for extinction attenuates spontaneous recovery. *Journal of Experimental Psychology: Animal Behavior Processes*, *19*, 77–89.

Brooks, D. C., & Bouton, M. E. (1994). A retrieval cue for extinction attenuates response recovery (renewal) caused by a return to the conditioning context. *Journal of Experimental Psychology: Animal Behavior Processes*, *20*, 366-379.

Brooks, D. C., Hale, B., Nelson, J. B., & Bouton, M. E. (1995). Reinstatement after counterconditioning. *Animal Learning & Behavior*, *23*, 383–390.

Brookshire, G., & Casasanto, D. (2011). Motivation and motor action: Hemispheric specialization for motivation reverses with handedness. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2610–2615). Austin, TX: Cognitive Science Society.

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*, 351–367.

Casasanto, D. (2011). Bodily relativity: The body-specificity of language and thought. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1258–1259). Austin, TX: Cognitive Science Society.

Casasanto, D., & Chrysikou, E. G. (2011). When left is "right." *Psychological Science*, *22*, 419–422.

Chorazyna, H. (1962). Some properties of conditioned inhibition. *Acta Biologiae Experimentalis*, *22*, 5–13.

Collins, B. N., & Brandon, T. H. (2002). Effects of extinction context and retrieval cues on alcohol cue reactivity among nonalcoholic drinkers. *Journal of Consulting and Clinical Psychology*, *70*, 390–397.

Colwill, R. M. (1991). Negative discriminative stimuli provide information about the identity of omitted response-contingent outcomes. *Learning & Behavior*, *19*, 326–336.

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R. F., Shanks, D. R., Robbins, T. W., . . . Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: Evidence for an associative account of delusions. *Brain*, *130*, 2387–2400.

Darby, R. J., & Pearce, J. M. (1995). Effects of context on responding during a compound stimulus. *Journal of Experimental Psychology*, *21*, 143–154.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *Spanish Journal of Psychology*, *10*, 230–241.

De La Fuente, J., Casasanto, D., Román, A., & Santiago, J. (2011). Searching for cultural influences on the body-specific association of preferred hand and emotional valence. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2616–2620). Austin, TX: Cognitive Science Society.

Dekeyne, A., & Deweer, B. (1990). Interaction between conflicting memories in the rat: Contextual pretest cuing reverses control of behavior by testing context. *Animal Learning & Behavior*, *18*, 1–12.

Delamater, A. R. (1996). Effects of several extinction treatments upon the integrity of Pavlovian stimulus-outcome associations. *Animal Learning & Behavior*, *24*, 437–449.

Denniston, J. C., Chang, R. C., & Miller, R. R. (2003). Massive extinction treatment attenuates the renewal effect. *Learning and Motivation*, *34*, 68–86.

Devine, P. G. (2001). Implicit prejudice and stereotyping: How automatic are they? Introduction to the special section. *Journal of Personality and Social Psychology*, *81*, 757–759.

Flack, W. (2006). Peripheral feedback effects of facial expressions, bodily postures, and vocal expressions on emotional feelings. *Cognition & Emotion*, *20*, 177–195.

Gawronski, B., Rydell, R. J., Vervliet, B., & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of Experimental Psychology: General*, *139*, 683–701.

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

Gunther, L. M., Denniston, J. C., & Miller, R. R. (1998). Conducting exposure treatment in multiple contexts can prevent relapse. *Behaviour Research and Therapy*, *36*, 75–91.

Gureckis, T. M., & Love, B. C. (2010). Direct associations or internal transformations? Exploring the mechanisms underlying sequential learning behavior. *Cognitive Science*, *34*, 10–50.

Hendersen, R. W. (1978). Forgetting of conditioned fear inhibition. *Learning and Motivation*, *9*, 16–30.

Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van den Bergh, O., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy*, *43*, 533–551.

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

Holland, P. C. (1983). Occasion-setting in Pavlovian feature positive discriminations. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Vol. 4. Discrimination processes* (pp. 183–206). New York, NY: Ballinger.

Iberico, C., Vansteenwegen, D., Vervliet, B., Dirikx, T., Marescau, V., & Hermans, D. (2008). The development of cued versus contextual conditioning in a predictable and an unpredictable human fear conditioning preparation. *Acta Psychologica*, *127*, 593–600.

Ito, T. A., Chiao, K. W., Devine, P. G., Lorig, T. S., & Cacioppo, J. T. (2006). The influence of facial feedback on race bias. *Psychological Science*, *17*, 256–261.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimulation* (pp. 9–33). Miami, FL: University of Miami Press.

Laborda, M. A., McConnell, B. L., & Miller, R. R. (2011). Behavioral techniques to reduce relapse after exposure therapy: Applications of studies of experimental extinction. In T. Schachtman & S. Reily (Eds.), *Associative learning and conditioning theory: Human and non-human applications* (pp. 79–103). New York, NY: Oxford University Press.

Lang, A. J., Craske, M. G., & Bjork, R. A. (1999). Implications of a new theory of disuse for the treatment of emotional disorders. *Clinical Psychology: Science and Practice*, *6*, 80–94.

Lanzetta, J. T., & Orr, S. P. (1981). Stimulus properties of facial expressions and their influence on the classical conditioning of fear. *Motivation and Emotion*, *5*, 225–234.

Leonard, D. W. (1975). Partial reinforcement effects in classical aversive conditioning in rabbits and human beings. *Journal of Comparative and Physiological Psychology*, *88*, 596–608.

Lieberman, D. (1999). *Learning: Behavior and cognition* (3rd ed.). Belmont, CA: Wadsworth Publishing.

Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cognition & Emotion*, *19*, 269–282.

Lovibond, P. F., Davis, N. R., & O'Flaherty, A. S. (2000). Protection from extinction in human fear conditioning. *Behaviour Research and Therapy*, *38*, 967–983.

Mackintosh, N. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.

McDonald, R. J., King, A. L., & Hong, N. S. (2001). Context-specific interference on reversal learning of a stimulus-response habit. *Behavioural Brain Research, 121*, 149–165.

McIntosh, D. N. (1996). Facial feedback hypotheses: Evidence, implications, and directions. *Motivation and Emotion*, *20*, 121–147.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin, 117*, 363–386.

Mystkowski, J. L., Craske, M. G., Echiverri, A. M., & Labus, J. S. (2006). Mental reinstatement of context and return of fear in spider-fearful participants. *Behavior Therapy*, *37*(1), 49–60.

Mystkowski, J. L., Mineka, S., Vernon, L. L., & Zinbarg, R. E. (2003). Changes in caffeine states enhance return of fear in spider phobia. *Journal of Consulting and Clinical Psychology*, *71*, 243–250.

Nelson, J. B. (2002). Context specificity of excitation and inhibition in ambiguous stimuli. *Learning and Motivation*, *33*, 284–310.

Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316*, 1002–1005.

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

Orr, S. P., & Lanzetta, J. T. (1984). Extinction of an emotional response in the presence of facial expressions of emotion. *Motivation and Emotion*, *8*, 55–66.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. London, England: Oxford University Press.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, *94*, 61–73.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*, 532–552.

Peck, C. A., & Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. *Learning and Motivation*, *21*, 1–31.

Pipkin, C. S., & Vollmer, T. R. (2009). Applied implications of reinforcement history effects. *Journal of Applied Behavior Analysis*, *42*, 83–103.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, *66*, 1–5.

Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, *72*, 77–94.

Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *American Psychologist*, *43*, 151–160.

Rescorla, R. A. (1993). Inhibitory associations between S and R in extinction. *Learning & Behavior*, *21*, 327–336.

Rescorla, R. A. (1996). Spontaneous recovery after training with multiple outcomes. *Animal Learning & Behavior*, *24*, 11–18.

Rescorla, R. A. (1997). Spontaneous recovery after Pavlovian conditioning with multiple outcomes. *Animal Learning & Behavior*, *25*, 99–107.

Rescorla, R. A. (2001). Retraining of extinguished Pavlovian stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*, 115–124.

Rescorla, R. A. (2003). Protection from extinction. *Learning & Behavior, 31*, 124–132.

Rescorla, R. A. (2004). Spontaneous recovery. *Learning & Memory*, *11*, 501–509.

Rescorla, R. A., & Heth, C. D. (1975). Reinstatement of fear to an extinguished conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, *1*, 88–96.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Ricker, S. T., & Bouton, M. E. (1996). Reacquisition following extinction in appetitive conditioning. *Learning & Behavior*, *24*, 423–436.

Rosas, J. M., & Callejas-Aguilera, J. E. (2006). Context switch effects on acquisition and extinction in human predictive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 461–474.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the implicit association test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*, 139–165.

Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin & Review*, *3*, 314–321.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 460–471.

Soussignan, R. (2002). Duchenne smile, emotional experience, and autonomic reactivity: A test of the facial feedback hypothesis. *Emotion*, *2*, 52–74.

Spear, N. E. (1971). Forgetting as retrieval failure. In W. K. Honig & P. H. R. James (Eds.), *Animal memory* (pp. 45–109). New York: Academic Press.

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, *54*, 768–777.

Swartzentruber, D. (1991). Blocking between occasion setters and contextual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*, 163–173.

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, *17*, 339–345.

Urcelay, G. P., Wheeler, D. S., & Miller, R. R. (2009). Spacing extinction trials alleviates renewal and spontaneous recovery. *Learning & Behavior*, *37*, 60–73.

Van Gucht, D., Vansteenwegen, D., Beckers, T., & Van den Bergh, O. (2008). Return of experimentally induced chocolate craving after extinction in a different context: Divergence between craving for and expecting to eat chocolate. *Behaviour Research and Therapy*, *46*, 375–391.

Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquisition context. *Behaviour Research and Therapy*, *43*, 323–336.

Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behaviour Research and Therapy*, *43*, 357–371.

Vila, N. J., & Rosas, J. M. (2001). Reinstatement of acquisition performance by the presentation of the outcome after extinction in causality judgments. *Behavioural Processes*, *56*, 147–154.

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*, 919–934.

Walther, E., & Langer, T. (2008). Attitude formation and change through association: An evaluative conditioning account. In R. Prislin & W. B. Crano (Eds.), *Attitudes and attitude change* (pp. 87–109). New York, NY: Psychology Press.

Wilson, P. N., & Pearce, J. M. (1992). A configural analysis for feature-negative discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *18*, 265–272.