The Quarterly Journal of Experimental Psychology

ISSN: 1747-0218 (Print) 1747-0226 (Online) Journal homepage: http://www.tandfonline.com/loi/pqje20

Mediated conditioning versus retrospective revaluation in humans: The influence of physical and functional similarity of cues

Mimi Liljeholm & Bernard W. Balleine

To cite this article: Mimi Liljeholm & Bernard W. Balleine (2009) Mediated conditioning versus retrospective revaluation in humans: The influence of physical and functional similarity of cues, The Quarterly Journal of Experimental Psychology, 62:3, 470-482, DOI: 10.1080/17470210802008805

To link to this article: http://dx.doi.org/10.1080/17470210802008805

Published online: 27 Feb 2009.

Submit your article to this journal

Article views: 97

View related articles

Citing articles: 5 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=pqje20

Download by: [The UC Irvine Libraries] **Date:** 18 December 2015, At: 15:51

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2009, 62 (3), 470–482

Psychology Press
Taylor & Francis Group

Mediated conditioning versus retrospective revaluation in humans: The influence of physical and functional similarity of cues

Mimi Liljeholm and Bernard W. Balleine

Two experiments assessed whether similarity between the two elements of a compound would influence the degree of mediated extinction versus recovery from overshadowing in human causal judgements. In both Experiments 1 and 2, we assessed the influence of extinguishing one element of a two-element compound on judgements about the other element. In Experiment 1 we manipulated the physical similarity of the two elements of the compound; in Experiment 2, we used equivalence and distinctiveness pretraining in order to vary their functional similarity. We found that these procedures influenced mediated extinction and recovery from overshadowing as a function of both physical and acquired similarity and distinctiveness, respectively. The implications of these results for previously reported differences between humans and nonprimate animals are discussed.

Keywords: Mediated extinction; Recovery from overshadowing; Cue selection; Acquired equivalence.

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

It has been well established that human causal judgements show a range of cue selection effects originally observed in nonprimate animals—that is, when two cues are presented in compound, what is learned about one of those cues is influenced by individual training with the other cue (e.g., Chapman, 1991; Shanks, 1985). One example of an effect of this kind is *prevention of overshadowing*; if Cue A is initially presented without any outcome, A−, and is then subsequently compounded with Cue B and paired with an outcome, AB+, the influence of B is judged to be greater than if the initial A− trials were not given (e.g., Carr, 1974; Navarro, Hallam, Matzel, & Miller, 1989).

There are, however, cue selection effects that have been considered unique to humans—for example, selection effects are often observed in human judgements even when the sequence of trial types is reversed, a phenomenon referred to as retrospective revaluation (Dickinson & Burke, 1996). If, in prevention of overshadowing, presentation of the single cue follows rather than precedes the compound trials (AB+, A−) the estimated causal strength of B is increased in similar manner to the forward trial order. Although retrospective revaluation effects in humans are often weaker than the analogous cue selection effects observed with forward designs (e.g., Chapman, 1991), they stand in sharp contrast to evidence from Pavlovian conditioning in animals indicating that, after AB+ training, the extinction of A reduces rather than increases conditioned responding to B, a phenomenon referred

Correspondence should be addressed to Mimi Liljeholm, Department of Psychology, UCLA, Box 951563, Los Angeles, CA 90095–1563, USA. E-mail: mlil@ucla.edu

This research was supported by Grant 56446 from the National Institute of Mental Health (NIMH) to B.W.B.

© 2008 The Experimental Psychology Society

http://www.psypress.com/qjep DOI:10.1080/17470210802008805

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

to as mediated extinction (Holland, 1999; Holland & Forbes, 1982; Rescorla & Cunningham, 1978).

Although the bulk of the evidence for retrospective revaluation comes from studies with humans, it has occasionally been demonstrated in rats (e.g., Balleine, Espinet, & Gonzales, 2005; Kaufman & Bolles, 1981; Liljeholm & Balleine, 2006). Likewise, whereas mediated conditioning and extinction effects have primarily been demonstrated in nonprimate animals, there is some evidence for such judgements from human studies as well (e.g., Hall, Mitchell, Graham, & Lavis, 2003). It appears, therefore, that both species can, under appropriate conditions, engage in both of these, apparently opposing, forms of learning. The question remains, however, why retrospective revaluation is more commonly observed in humans and mediated conditioning more frequently demonstrated in nonprimate animals.

One salient difference between studies on causality judgements in humans and Pavlovian conditioning experiments with nonprimate animals that may explain why seemingly similar experimental procedures should yield such disparate results is the nature of the stimuli used as cues. For example, a common scenario presented to human participants is that involving the influence of different foods on allergic reactions (e.g., Melchers, Lachnit, & Shanks, 2004), where the pairing of certain foods (e.g., banana and cheese) results in an allergic reaction whereas one of those foods alone (e.g., banana) does not. It is likely that human participants have extensive experience with these food stimuli; they can in fact be considered experts on the distinct tastes, tactile sensations, smells, and visual properties of each cue. Consequently, they might be expected to generalize less across the two stimuli on the basis of any shared features and to be less likely to configure the two stimuli into a single, unique, cue on compound trials. In other words, one may assume that participants are aware of the potentially independent causal influences of the two cues.

In contrast, in Pavlovian conditioning experiments with nonprimate animals, naïve subjects are generally presented with novel stimuli that often share common features. For example, Balleine et al. (2005, Exp. 1) first exposed thirsty rats to a two-element flavour compound in a sucrose solution (AB+) and subsequently to one of the flavour elements of that compound in either water (A−) or sucrose solution (A+). The rats were then food deprived, and the other flavour element (B) was presented in water. Balleine et al. found that rats given A + consumed more of B than did rats given A− (i.e., mediated conditioning). One potential explanation for these results is that some elements common to the two flavours (orange and lemon−lime) encouraged generalization between them—that is, the representation of any feature shared by A and B would be active during the extinction or conditioning of A in Phase 2, as well as during the test of B. The assumption that the amount of generalization between to events is a function (linear or not) of the number of shared features is ubiquitous to theories of associative learning (e.g., McLaren & Mackintosh, 2002).

Another factor that might have contributed to the absence of any retrospective revaluation effects is the mixing of the two flavours into a single solution during the compound phase, which could potentially have encouraged the formation of a unique configuration of the two cues (e.g., Pearce, 1994). This configuration may then have been retrieved both by Cue A in Phase 2 and by Cue B during test, thus providing a basis for generalization. Some evidence for this notion comes from a study on humans by Livesey and Boakes (2004), in which forward cue selection effects were completely abolished when the spatial separation between compound elements was reduced to zero (i.e., when the two compound elements were completely merged into a single cue).

Of course, in order to perceive two elements as distinct causes, one must first perceive them as distinct events (Anderson, 1960). Thus, whether the mediated conditioning effects observed by Balleine et al. (2005) occurred because of generalization due to common elements, or because of generalization due to the formation of a unique configuration, one would expect a treatment that enhanced

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

discrimination between the two flavours to increase the likelihood of observing retrospective revaluation. Consistent with this interpretation, Balleine et al. (2005; Exp. 2) found that giving rats the opportunity for perceptual learning of this kind, by giving them intermixed preexposure to the two flavours (Mackintosh, Kaye, & Bennett, 1991; Symonds & Hall, 1995), resulted in retrospective revaluation rather than mediated conditioning (i.e., in their second experiment, rats given A+ consumed less of B than did rats given A−).

If a treatment that encourages discrimination between two cues making up a compound increases the likelihood of observing retrospective revaluation in rats, one might expect, conversely, that a treatment that encourages generalization between cues making up a compound would increase the likelihood of observing mediated conditioning in humans. This suggestion was evaluated in the current series. In both Experiments 1 and 2, we assessed the influence of extinguishing one element of a two-element compound on causal judgements about the other element. In Experiments 1a and 1b we manipulated the physical similarity and spatial proximity of the two elements of the compound whereas, in Experiment 2, we employed a pretraining procedure in order to vary their functional similarity.

As mentioned, previous studies that have successfully demonstrated retrospective revaluation effects in humans have often employed compound elements with rich, preexisting, semantic content, and these compound elements were usually spatially separated and labelled as distinct entities even on compound trials. In a general attempt to reduce the influence of these factors on discrimination, throughout all conditions of the current experiments we presented human participants with compounds made up of two abstract shapes that were, at least partly, joined by a common contour and that were verbally referred to as a single cue.

EXPERIMENT 1

Livesey and Boakes (2004) abolished cue selection effects in a forward blocking design by reducing the spatial separation between compound elements to zero. They suggested that the complete lack of spatial separation had encouraged configural processing and, as a result, increased the amount of generalization between the compound and its elements. In Experiment 1a, we attempt to increase generalization between compound elements in two ways: (a) by increasing the number of shared features, and (b) following Livesey and Boakes (2004), by reducing spatial separation. In Experiment 1b, we assessed whether the complete lack of spatial separation was a necessary condition for any mediated extinction effects to emerge.

The stimuli and design of Experiment 1a are presented in Figure 1a. In each of four groups, participants were presented with a cover story stating that their task was to evaluate whether certain proteins activate a newly discovered receptor: the α receptor. The two target proteins (i.e., those making up the compound) each consisted of a

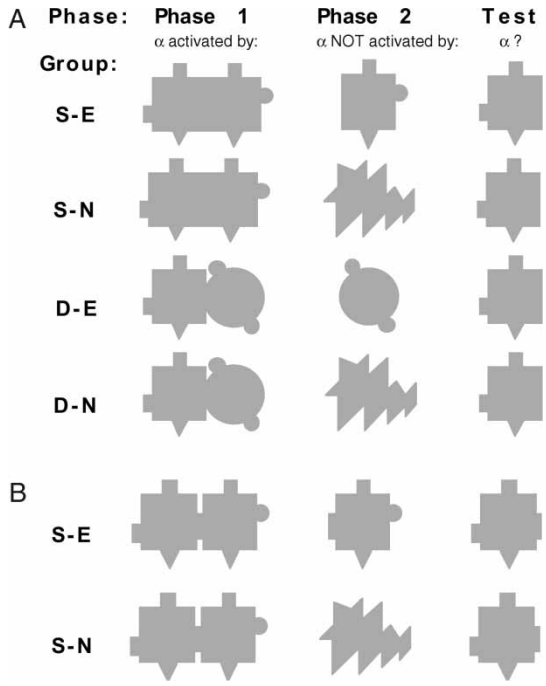

Figure 1. *(a) Design and graphics of Experiment 1a. (b) Design and graphics of Experiment 1b.*

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

large central feature in the shape of either a square or a circle and two or three smaller external features, each shaped as a square, a triangle, or a circle. The proteins were of approximately the same size and uniformly grey in colour.

In Phase 1 of the experiment, a protein compound activated the α receptor across groups. For two groups of participants, groups "similar", the two elements making up the compound were completely merged on compound trials and differed only with respect to the shape and location of one external feature, which occurred either on the left or on the right side and had the shape of either a square or a circle—see top two rows in Figure 1a. In one of these groups, group "similar extinction" (S-E), one element of the protein compound subsequently failed to activate the α receptor in a second phase, whereas, in the other group, group "similar no extinction" (S-N) this compound element was replaced with a novel protein shape. For two analogous groups, groups "distinct extinction" (D-E) and "distinct no extinction" (D-N), the elements making up the protein compound, although identical in colour and similar in overall size, did not share any feature in terms of shape and were partially separated by a spatial gap, so that their individual contours were clearly visible on compound trials—see the bottom two rows in Figure 1a.

On test, participants in all groups rated the likelihood that the compound element that had not been presented alone (i.e., the element that was absent in Phase 2 across groups) would activate the α receptor on a scale from 0 "not at all likely" to 9 "extremely likely".

If the recovery from overshadowing effect is influenced by generalization based on physical similarity, proximity or both, then this effect should emerge when the elements of the compound have distinct features and are separated by a spatial gap, but not when they are physically more similar, and the spatial separation is zero. Indeed, in the latter case, we predicted that, rather than increasing the predictive status of the unpresented cue, extinction of one element would generalize to the other element and reduce its predictive status. Specifically, mean ratings would be lower in group S-E than in group S-N but higher in group D-E than in group D-N.

Experiment 1b was conducted to explore whether the complete lack of spatial separation in Experiment 1a was necessary to observe a mediated extinction effect in group S-E. Only the "similar" groups were included in Experiment 1b, and these groups were identical to those in Experiment 1a except that the two compound elements were only partly joined by a common contour—that is, a spatial gap was inserted that made the individual contours of compound elements clearly visible on compound trials—see Figure 1b. If a high degree of feature similarity is sufficient to generate a mediated extinction effect, then the results of Experiment 1b should be comparable to those observed for the "similar" groups in Experiment 1a. If, however, generalization in Experiment 1a was largely due to the lack of spatial separation, then increasing this separation should reduce generalization between the elements of the compound and, hence, reduce mediated conditioning.

Method

Participants

A total of 110 undergraduates at the University of California, Los Angeles (UCLA), participated to obtain course credit in an introductory psychology course. In Experiment 1a, 80 participants were randomly assigned to four groups. In Experiment 1b, 30 participants were randomly assigned to two groups.

Procedure

All aspects of the procedure were identical across Experiments 1a and 1b. The materials were presented on a computer, and responses were given on the keyboard. In addition to the compounds and individual proteins shown in Figure 1, in order to make the task less monotonous, participants were presented with a distractor receptor (labelled β) and two distractor proteins that both activated this β receptor. One of the distractor proteins also activated the α receptor, while none

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

of the target proteins activated the β receptor. At the beginning of the experiment, participants in all groups were presented with the following cover story and instructions:

In this experiment, you will play the role of a research assistant working in a molecular neuroscience laboratory. Your task is to determine if certain proteins are capable of activating two newly discovered receptors; the α and β receptors. On each trial, you will be shown a particular protein together with a receptor and you will be asked if you think that protein will activate that receptor or not. If you think that it will activate the receptor, press the Y key for "yes" and if you don't think that it will activate the receptor, press the N key for "no". Immediately after answering you will be given feedback about if you were right or wrong and after a few trials with each protein you should be able to make accurate predictions.

On the feedback screen, participants were informed about whether they were right or wrong, and the state of the receptor (i.e., activated or not) was also indicated graphically.

In Phase 1, the compound protein was presented 24 times, and each distractor protein was presented 16 times. In Phase 2, the protein element (see Figure 1) was shown 16 times, and each distractor protein was shown 12 times. Half of these trials involved the α (target) receptor, and the other half involved the β receptor, for both target and distractor proteins. All trials within a phase were randomly presented. The participants were not informed about the separate phases and were presented with the trials in a single continuous session.

At the end of the second phase participants were presented with the following instructions:

In this part of the experiment, your supervisor wants to know what you learned about the proteins. He will show you a protein together with a receptor, and he wants you to rate how likely it is that the particular protein will activate that receptor, on a scale that ranges from 0 (Not at all likely) to 9 (Extremely likely). Even if you feel like you don't know the answer, your supervisor wants you to give him your best guess based on the training you just had.

On the next screen, participants were shown the relevant protein element (see Figure 1) together with the scale described above. They were asked how likely it is that the protein will activate the receptor and were told to type in a number between 0 and 9 to indicate their rating on the scale. No other ratings were collected.

Results

Experiment 1a

To ensure that training had proceeded as planned, statistical analyses were performed on the number of errors committed on trials involving the target proteins (the compound in Phase 1 and the compound element in Phase 2) and the target (α) receptor. The mean number of errors on trials with the protein compound was 1.2, 1.3, 0.95, and 1.5 in groups S-E, S-N, D-E, and D-N, respectively. The mean number of errors on trials with the compound element (or with the control protein in groups S-N and D-N) was 2.0, 1.9, 2.4, and 1.4 in groups S-E, S-N, D-E, and D-N, respectively. There were no significant differences between groups for either the compound or the compound element, Fs < 0.90.

The mean likelihood ratings from the test are presented in Figure 2 (first four bars). There was clear evidence for mediated extinction such that mean ratings were lower in group S-E than in group S-N. Conversely, as predicted, we observed retrospective revaluation in group D-E; the mean ratings in this group were higher than those in group D-N. As would be expected based on the mediated extinction and retrospective revaluation effects, the two extinction groups also differed, such that mean ratings were lower in group S-E

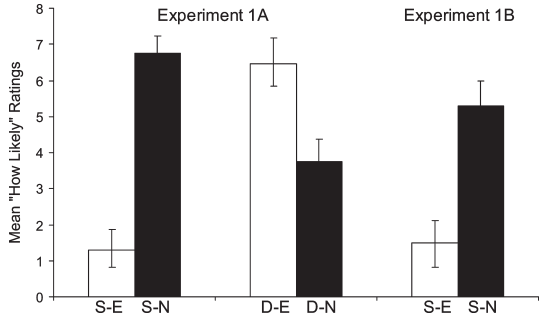

Figure 2. *Results from Experiment 1a and Experiment 1b. Mean likelihood ratings on test in each group. Error bars represent* ± *standard error of the mean.*

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

than in group D-E. Finally, there was a clear difference between groups S-N and D-N such that mean ratings were higher in the former group than in the latter, which is probably indicative of differential generalization from the training compound to the test element in the two groups.

This description was confirmed by the statistical analysis. A Similarity (2) × Extinction (2) analysis of variance was performed on the ratings from groups S-E, S-N, D-E, and D-N, with both factors as between-subject variables. There was no main effect of similarity, $F(1, 76) = 3.09$, $p = .08$; however, there was a main effect of extinction, such that mean ratings were significantly lower in the extinction groups than in the groups that did not receive extinction, $F(1, 76) = 5.05$, $p < .05$. Moreover, there was a highly significant interaction between similarity and extinction, $F(1, 76) = 44.39$, $p < .001$.

Simple effect analyses revealed that mean ratings were significantly lower in group S-E than in group S-N, $F(1, 38) = 53.76$, $p < .001$, indicative of mediated extinction. Furthermore, mean ratings were significantly higher in group D-E than in group D-N, $F(1, 38) = 7.72$, $p < .01$, indicative of a recovery from overshadowing effect. Finally, the difference between groups S-E and D-E was significant, $F(1, 38) = 31.78$, $p < .001$, as was the difference between groups S-N and D-N, $F(1, 38) = 13.60$, $p < .005$.

Experiment 1b

With respect to the training data, the mean number of errors on trials with the protein compound was low (1.1 and 1.4 for groups S-E and S-N, respectively) as was the mean number of errors on trials with the compound element or control protein: 2.0 and 1.8 for groups S-E and S-N, respectively. There were no significant differences between groups for either the compound or the compound element, Fs < 0.70.

The mean likelihood ratings from the test are presented in Figure 2 (last two bars). Again, there was clear evidence for mediated extinction: A one-way, between-subjects analysis of variance performed on the ratings revealed that mean ratings were significantly lower in group S-E than in group S-N, $F(1, 38) = 17.0$, $p < .001$. Adding the ratings from the similar groups of Experiment 1a, an Experiment (2) × Extinction (2) between-subjects analysis of variance was performed to directly assess the influence of spatial separation. There was a main effect of extinction, $F(1, 76) = 61.0$, $p < .001$, but no effect of the experiment variable, $F(1, 76) = 1.1$, $p = .3$, nor an interaction between experiment and extinction, $F(1, 76) = 1.9$, $p = .17$.

Discussion

The results of Experiment 1 support the argument that mediated extinction is based, primarily, on generalization due to the similarity of the elements presented in the compound phase. In Experiment 1a, when the elements of the compound were highly similar, and their spatial separation was reduced to zero on compound trials, mediated extinction emerged rather than retrospective revaluation. In Experiment 1b, the complete lack of spatial separation was ruled out as a necessary condition for this mediated extinction effect. Indeed, the cross-experiment analysis revealed no reliable difference between the similar groups of Experiment 1a and those in Experiment 1b.

It is unlikely, however, that the proximity of compound elements played no role in the mediated extinction effects observed in these experiments. Recall that, across all conditions, compound elements were at least partly joined by a common contour, and the compound was referred to as a single cue; this might have encouraged configural processing to a point where any additional reduction in spatial separation became negligible. A high degree of configural processing, due to the proximity of compound elements, might also explain why the retrospective revaluation effect in Experiment 1 was much less pronounced than the mediated extinction effects. More systematic parametric studies will be required to explore the respective roles of spatial proximity and physical similarity in the emergence of mediated conditioning versus retrospective revaluation effects.

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

Of course, many demonstrations of mediated conditioning in nonprimate animals have employed compound elements that are neither similar nor spatially proximal, such as auditory and visual stimuli (e.g., Shevill & Hall, 2004), making an account of mediated extinction based solely on the overlap of physical features less plausible. There are, however, sources of generalization available other than physical similarity. For example, it has been demonstrated that generalization between two events can be increased or reduced based on whether they predict a common consequence or distinct consequences. Functional similarity could, therefore, also influence the size of retrospective revaluation effects, and, as such, we assessed this possibility in Experiment 2.

EXPERIMENT 2

So far, we have considered two sources of generalization between compound elements: the number of shared features and the configuration of compound elements into a unique cue. Another potential source of generalization can be derived from the fact that, on compound trials, the two elements of the compound are paired with a common consequence, a treatment that has previously been found to increase generalization. For example, Honey and Hall (1989, Exp. 3) presented animals with three perceptual cues (A, N, and B). In one group A and N were each followed by a food pellet, whereas B was not (i.e., A+, N+, B−). In a second group B was reinforced whereas A and N were not (i.e., A−, N−, B+). Finally, Cue N was paired with foot-shock after which generalization of fear to Cues A and B was assessed. Animals in both groups were found to generalize fear conditioned to N more to Cue A than to Cue B, a phenomenon referred to as *acquired equivalence*. Likewise, in human participants, several researchers have demonstrated enhanced generalization between stimuli that had previously been treated the same way, as well as poor generalization between stimuli that had previously been treated differently—that is, *acquired distinctiveness* (e.g., Hall et al., 2003). In the current experiment, we explored whether equivalence and distinctiveness treatments would promote mediated extinction and recovery from overshadowing, respectively.

The design of Experiment 2 was the same as that for Experiment 1 with the following exceptions: (a) The same set of stimuli were used in all four groups, (b) a pretraining phase was added for all groups, (c) rather than replacing the extinction element with a novel protein, the extinction phase was simply eliminated for groups S-N and D-N, and (d) there was only one distractor protein. The experiment had three phases with each phase corresponding to training with one of three receptors (the α, β, and π receptors). In the first two phases, the influence of three different proteins (all black and shaped, respectively, as a square, circle, and triangle) on the activation of the α and α receptors (Phases 1 and 2, respectively) was evaluated. This constituted the pretraining. In the third phase, participants evaluated the influence of a protein compound, made up of the square and circle elements, on the activation of the π receptor. As before, the compound elements were partly joined by a common contour and were verbally referred to as a single cue on compound trials.

The overall design of Experiment 2 is presented in Table 1. In the "similar" groups, the circle (C) and square (S) both activate the α receptor but not the β receptor, whereas the triangle (T) activates the β receptor but not the α receptor. In other words, C and S were functionally similar to one another and functionally different from T (protein T was included to provide a frame of reference for the equivalence of S and C). In group S-E, the SC compound was presented as a cause of π receptor activation in a third phase, whereas, in a fourth phase (the extinction phase), protein C alone did not activate this receptor. Group S-N received identical treatment, except that there were no presentations of C alone in the fourth phase. Participants in the two analogous groups,

Table 1. *Activation and nonactivation of the α receptor, β receptor, and π receptor by the square, circle, triangle, and compound proteins, for all groups in Experiment 2*

	Pretraining				
Group	*Phase 1: α*	*Phase 2: β*	*Compound Phase 3: π*	*Extinction Phase 4: π*	*Test π*
S-E	S+, C+, T−	S−, C−, T+	SC+, T+	C−, T+	S?
S-N	S+, C+, T−	S−, C−, T+	SC+, T+	SC+, T+	S?
D-E	S−, C+, T−	S+, C− T+	SC+, T+	C−, T+	S?
D-N	S−, C+, T−	S+, C− T+	SC+, T+	SC+, T+	S?

Note: + denotes activation; − denotes nonactivation; α receptor = Phase 1; β receptor = Phase 2; π receptor = Phases 3–4. S = square; C = circle; T = triangle; SC = compound proteins. S-E = similar extinction. S-N = similar no extinction. D-E = distinct extinction. D-N = distinct no extinction.

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

groups D-E and D-N, receive a treatment identical to that in groups S-E and S-N, respectively, except that proteins C and T both activated the α receptor but not the β receptor, whereas protein S activated the β receptor but not the α receptor. We hoped, by this treatment, to encourage participants to treat the proteins S and C as functionally distinct; again, protein T was intended to highlight this distinctiveness. On test, participants in all groups rated the likelihood that protein S would activate the π receptor, on a scale from 0 "not at all likely" to 9 "extremely likely".

If our equivalence and distinctiveness manipulations are effective, and if the recovery from overshadowing effect is negatively related to the degree of generalization based on functional similarity, then this retrospective revaluation effect should emerge after distinctiveness pretraining but not after equivalence pretraining. Indeed, in the latter case, we predicted that mediated extinction would emerge. In other words, the mean likelihood ratings should be lower in group S-E than in group S-N but higher in group D-E than in group D-N.

Method

Participants

A total of 60 undergraduates at UCLA participated to obtain course credit in an introductory psychology course. They were randomly assigned to four groups.

Procedure

The procedure was identical to that of Experiment 1, except for modifications pertaining to the number of phases in the experiment and the number of receptors and proteins under evaluation. In Phases 1 and 2 each of the three proteins were presented 12 times in random order for all groups. For groups S-N and D-N, 16 SC-compound trials and 16 T trials were randomly presented throughout Phases 3 and 4. For groups S-E and D-E, 12 SC-compound trials and 8 T trials were randomly presented in Phase 3, while 12 C trials and 8 T trials were randomly presented in Phase 4. Participants were informed that the first three phases of the experiment were separate assessments, and each of these phases began with a screen announcing which of the receptors would be evaluated. However, no information was given about the shift from Phase 3 to Phase 4, and participants were simply presented with a continuous stream of trials across these phases.

The reason for not including nonreinforced presentations of a novel cue in groups S-N and D-N, to match the extinction phase in the other groups (as was done in Experiment 1), was that the cues in Experiment 2 only varied along the single dimension of basic shape. Any novel cue would therefore have had to involve either a drastic increase in complexity (i.e., sharing features with all other cues) or a substantial overlap with one, but not the other, cues. In order to reduce, albeit only slightly, the difference in the total number of trials presented to participants

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

as a result of eliminating the extinction phase, participants in groups S-N and D-N received four more compound trials than did participants in groups S-E and D-E.

On test, participants in all groups rated, in order (a) the likelihood that protein C will activate the α receptor, (b) the likelihood that protein S will activate the α receptor, (c) the likelihood that protein C will activate the π receptor, and (d) the likelihood that protein S will activate the π receptor (the target test trial). Except for these multiple test trials, the instructions and rating scale used for the test were identical to those in Experiment 1. Importantly, the influence of the different proteins on the α receptor was the opposite of that predicted for the π receptor in each group. Consequently, while reminding participants of the relationship between proteins S and C, the initial test trials do not encourage participants to provide ratings that correspond numerically to our predictions.

Results and discussion

As in Experiment 1, statistical analyses were performed on the number of errors committed on trials involving the target proteins. During pretraining with the α and β receptors, the mean number of errors was low (0.60). A three-way analysis of variance performed on these pretraining errors, with group, receptor, and protein as factors, revealed no significant differences between the four groups, nor between the two receptors or between the two proteins, and no interactions, all Fs < 1.5. With respect to the target receptor (π) the mean number of errors on trials with the protein compound (SC) was 0.27, 0.33, 0.33, and 0.4 in groups S-E, S-N, D-E, and D-N, respectively. The mean number of errors on trials with the extinguished compound element (i.e., protein C, which was presented alone together with the π receptor in groups S-E and D-E but not in groups S-N and D-N) was 0.60 and 0.47 in groups S-E and D-E, respectively. There were no significant differences between groups for either the compound or the compound element, Fs < 0.50.

To further ensure that pretraining had proceeded smoothly, an analysis of variance was performed on ratings of the likelihood that proteins S and C, respectively, would activate the α receptor. For groups similar, in which both proteins had activated the α receptor, mean ratings equalled 8.50 and 8.43 for the C and S protein, respectively. There was no significant difference between the two groups or between the two proteins and no interaction, all Fs < 0.12. For groups distinct, in which protein C but not protein S had activated the α receptor, mean ratings were significantly higher for protein C (8.03) than for protein S (0.60), $F(1, 28) = 172.76$, $MSE = 4.80$, $p < .001$, but there was no difference between the two groups, nor was there an interaction, Fs < 1.6.

To verify that the compound element had extinguished in groups S-E and D-E, a two-way analysis of variance, with similarity and extinction as factors, was performed on ratings of the likelihood that the extinguished compound element (C) would activate the π receptor. This analysis confirmed that the compound element had extinguished in groups S-E and D-E (mean rating = 0.81 and 1.3, respectively) but not in groups S-N and D-N (mean rating = 4.07 and 4.0, respectively) resulting in main effect of extinction, $F(1, 54) = 14.44$, $p < .001$. In contrast, there was no effect of similarity, nor any interaction between extinction and similarity, Fs < 0.8.

The mean likelihood ratings from the target test (i.e., the final rating) are presented in Figure 3. Evidence indicative of a mediated extinction effect as a consequence of the equivalence training was observed, such that mean ratings were lower in group S-E than in group S-N. Analogously, the distinctiveness training given to groups D-E and D-N appeared to encourage retrospective revaluation; the mean ratings in group D-E were higher than those in group D-N. A clear difference also emerged between the two extinction groups; mean likelihood ratings appeared to be lower in group S-E than in group D-E. In contrast, groups S-N and D-N did not appear to differ from one another.

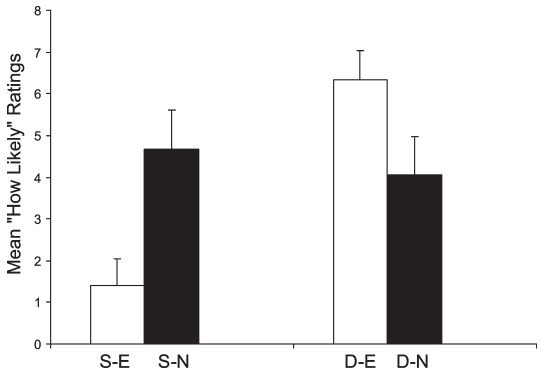

Figure 3. *Results from Experiment 2. Mean likelihood ratings on target test trial in each group. Error bars represent ± standard error of the mean.*

This description was confirmed by the statistical analysis. As in Experiment 1, a Similarity (2) × Extinction (2) analysis of variance was performed on the ratings from groups S-E, S-N, D-E, and D-N, with both factors as between-subject variables. There was a main effect of similarity, such that mean ratings were significantly lower in the acquired equivalence groups than in the acquired distinctiveness groups, $F(1, 56) = 5.05$, $p < .05$. Moreover, there was a highly significant interaction between similarity and extinction, $F(1, 56) = 11.71$, $p < .005$. There was no main effect of extinction, $F(1, 56) = 0.38$, $p = .54$.

Simple effects analyses revealed that mean ratings were significantly lower in group S-E than in group S-N, $F(1, 28) = 8.33$, $p < .01$, indicative of mediated extinction. However, while mean ratings were higher in group D-E than in group D-N, this recovery from overshadowing effect was only marginally significant, $F(1, 28) = 3.85$, $p = .06$. Finally, while the difference between groups S-E and D-E was significant, $F(1, 28) = 26.49$, $p < .001$, the difference between groups S-N and D-N, unlike in Experiment 1a, was not, $F(1, 28) = 0.21$, $p = .65$.

In order to assess whether the failure to obtain a significant recovery from overshadowing effect was a matter of statistical power, 10 additional participants were run on the two distinct conditions (5 each in Groups D-E and D-N, respectively). A Replication (2) × Extinction (2) analysis of variance performed on the target ratings found no main effect of the replication factor, nor an interaction, Fs < 0.2, indicating that these 10 additional participants did not differ as a group from those in the original experiment. There was, however, a main effect of extinction, such that mean ratings were significantly higher in group D-E than in group D-N, $F(1, 36) = 5.95$, $p < .05$. That is, with 10 additional participants the retrospective revaluation effect was indeed rendered reliable.

In summary, as predicted, acquired equivalence and distinctiveness treatments promoted mediated extinction and retrospective revaluation, respectively. It is important to consider the entire pattern of results when interpreting these effects. For example, recall that the two no extinction groups received a small number of additional trials with the SC+ compound relative to the extinction groups. Indeed, in the no extinction groups, the π receptor was always activated regardless of the protein (i.e., Phases 3 and 4). Thus, if focusing only on the similar groups, one might attribute the fact that the target test ratings were higher in group S-N than in group S-E to this difference in training. However, this alternative account fails to explain why a difference in the opposite direction emerged between the two distinct groups.

It is interesting to note that, as in Experiment 1a, the mediated extinction effect appears to be more pronounced than the retrospective revaluation effect. It is possible that, in spite of the pretraining, significant generalization between compound elements occurred in group D-E, and that more extensive pretraining would have resulted in greater recovery from overshadowing for this group. One interesting source of generalization that may have prevented a more robust recovery from overshadowing effect in group D-E is the acquired equivalence produced by the compound–outcome pairings in Phase 3. Indeed, the only difference between this compound training and acquired equivalence training is that the compound elements occur simultaneously rather than separately. Notably, although Pavlovian

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

conditioning experiments demonstrating mediated extinction in rats (e.g., Holland, 1999; Holland & Forbes, 1982) do not employ any equivalence pretraining per se, it is possible that such effects are encouraged by acquired equivalence induced during compound training.

GENERAL DISCUSSION

In two experiments we evaluated retrospective revaluation and mediated extinction in human participants, using procedures that encouraged generalization or discrimination between compound elements. In Experiment 1, we found that a high degree of physical similarity of compound elements resulted in mediated extinction. In contrast, when the compound elements were quite distinct, retrospective revaluation emerged. In Experiment 2, we manipulated the functional similarity of compound elements and found that this procedure was effective at promoting mediated extinction and retrospective revaluation as a function of acquired equivalence and distinctiveness, respectively. These results may shed some light on why retrospective revaluation effects are most frequently observed with human participants whereas mediated extinction and mediated conditioning are commonly demonstrated in nonprimate animals. Previous studies of human causality judgement have tended to use scenarios in which stimuli presented in compound were likely to be readily interpreted as distinct events with potentially distinct causal influences. It appears that a high degree of discriminability between compound elements, based on previously acquired semantic knowledge, distinctiveness pretraining, or a large number of unique physical features, is necessary for retrospective revaluation effects to emerge. Distinctiveness between compound elements may be particularly important because of the potential for enhanced generalization due to equivalence learning occurring on compound conditioning trials. Another apparent source of generalization that may interfere with retrospective revaluation effects is the merging of compound elements into a configural representation. Recall that, in all conditions of the current experiments, compound elements were partly joined by a common contour and verbally referred to as a single cue on compound trials. It is likely that this encouraged a configuration of compound elements (see Livesey & Boakes, 2004), which may account for why the meditated extinction effect was more pronounced than the recovery from overshadowing effect across experiments.

Both retrospective revaluation and mediated conditioning effects initially posed a problem for traditional associative accounts of causal learning. For example, in the *sometimes opponent process* (SOP) theory, Wagner (1981) proposed that stimuli are represented as collections of elements and that, at any given point in time, each element can be in one of three states: inactive (I), active 1 (A1), and active 2 (A2). The actual presentation of stimulus activates its elements into the A1 state, from which they decay into the A2 state. In addition, elements that are in the A1 state (i.e., present) can associatively retrieve absent elements into the A2 state. According to the SOP theory, different associations form between elements depending on what state they are in; however, no associations, inhibitory or excitatory, can form between elements that are simultaneously in the A2 state.

In order to account for mediated extinction and conditioning, Holland (1983) proposed a modification of Wagner's SOP theory, such that elements that are both in the A2 state form inhibitory associations, thus explaining why presenting one element of a previously reinforced compound without reinforcement would extinguish the associative strength of the other, nonpresented, element of that compound—that is, both the nonpresented element and the reinforcer would be associatively retrieved into A2 resulting in an inhibitory association between the two. In contrast, in order to account for retrospective revaluation effects, Dickinson and Burke (1996) proposed the exact opposite modification of Wagner's theory, arguing that elements that are both in the A2 state would form excitatory connections, thus explaining how presenting one element of a previously reinforced compound without reinforcement could increase

the associative strength of the other, nonpresented, element.[1] If, as early findings indicate, retrospective revaluation was a uniquely human phenomenon, and mediated extinction was unique to nonprimate animals, the contradictory nature of these two versions of the original SOP model might not pose a problem. However, the current results suggest that the source of the differences between human participants and nonprimate animals, with respect to retrospective revaluation and mediated extinction, does not lie in a difference between species.

This is not itself surprising; other researchers have found retrospective revaluation in nonprimate animals by manipulating different aspects of compound elements, such as increasing their discriminability (Balleine et al., 2005) or reducing their biological relevance (Miller & Matute, 1996). Here we show, conversely, that mediated extinction can emerge in humans and provide evidence that such effects are influenced by stimulus generalization based either on common elements or on common consequences. These findings imply, of course, that the species differences previously described were a product of the procedures and the particular stimuli that were used in those studies. We propose that the use of stimuli with which the human participants were highly familiar ensured that they were able to discriminate between them and were capable of ascribing individual causal influences to these cues. In contrast, nonprimate animals are typically exposed to novel cues during the compound phase and hence have little basis on which to discriminate between compound elements or represent them as discrete entities. Moreover, the common consequences with which compound elements are associated could result in increased generalization between them based on their apparent functional equivalence. Consequently, it seems that, rather than having one rule for nonprimate animals and another for humans, an associative theory could coherently account for both retrospective revaluation and mediated conditioning effects by considering the influence of factors affecting stimulus generalization.

Original manuscript received 13 June 2007
Accepted revision received 13 February 2008
First published online 12 May 2008

REFERENCES

Anderson, J. (1960). *Studies in empirical philosophy.* Sydney, Australia: Angus & Robinson.

Balleine, B. W., Espinet, A., & Gonzales, F. (2005). Perceptual learning enhances retrospective revaluation of conditioned flavor preferences in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*, 341–350.

Carr, A. F. (1974). Latent inhibition and overshadowing in conditioned emotional response conditioning in rats. *Journal of Comparative and Physiological Psychology*, *86*, 718–723.

Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 837–854.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgments. *Quarterly Journal of Experimental Psychology*, *49B*, 60–80.

Hall, G., Mitchell, C., Graham, S., & Lavis, Y. (2003). Acquired equivalence and distinctiveness in human discrimination learning: Evidence for associative mediation. *Journal of Experimental Psychology: General*, *132*, 266–276.

Holland, P. C. (1983). Representation mediated overshadowing and potentiation of conditioned aversions. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 1–13.

Holland, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or blocking cues. *Quarterly Journal of Experimental Psychology*, *52B*, 307–333.

[1]It should be noted that Dickinson and Burke's (1996) model would be able to predict mediated extinction and conditioning effects when the compound elements share a large number of physical features, as in Experiment 1, since those shared features will be actually present and therefore active in the A1 state when either of the two compound elements is presented. However, other factors relevant to our results (e.g., spatial proximity of compound elements and equivalence pretraining) clearly fall outside the scope of this model.

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015

Holland, P. C., & Forbes, D. T. (1982). Representation-mediated extinction of conditioned flavor aversion. *Learning and Motivation*, *13*, 454–471.

Honey, R. C., & Hall, G. (1989). The acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*, 338–346.

Kaufman, M. A., & Bolles, R. C. (1981). A nonassociative aspect of overshadowing. *Bulletin of the Psychonomic Society*, *18*, 318–320.

Liljeholm, M., & Balleine, B. W. (2006). Stimulus salience and retrospective revaluation. *Journal of Experimental Psychology: Animal Behavior Processes*, *29*, 97–106.

Livesey, E. J., & Boakes, R. A. (2004). Outcome additivity, elemental processing and blocking in human causality judgments. *Quarterly Journal of Experimental Psychology*, *57B*, 361–379.

Mackintosh, N. J., Kaye, H., & Bennett, C. H. (1991). Perceptual learning in flavour aversion conditioning. *Quarterly Journal of Experimental Psychology*, *43B*, 297–322.

McLaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior*, *30*, 177–200.

Melchers, K. G., Lachnit, H., & Shanks, D. R. (2004). Within-compound associations in retrospective revaluation and in direct learning: A challenge for comparator theory. *Quarterly Journal of Experimental Psychology*, *57B*, 25–53.

Miller, R. R., & Matute, H. (1996). Biological significance in forward and backward blocking: Resolution Pl of a discrepancy between animal conditioning and human causal judgment. *Journal of Experimental Psychology: General*, *125*(4), 370–386.

Navarro, J. I., Hallam, S. C., Matzel, L. D., & Miller, R. R. (1989). Superconditioning and unovershadowing. *Learning & Motivation*, *20*, 130–152.

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, *101*, 587–607.

Rescorla, R. A., & Cunningham, C. L. (1978). Within-compound flavor associations. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*, 45–67.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *Quarterly Journal of Experimental Psychology*, *37B*, 1–21.

Shevill, I., & Hall, G. (2004). Retrospective revaluation effects in the conditioned suppression procedure. *Quarterly Journal of Experimental Psychology*, *57B*, 331–347.

Symonds, M., & Hall, G. (1995). Perceptual learning in flavor aversion conditioning: Role of stimulus comparison and latent inhibition of common elements. *Learning and Motivation*, *26*, 203–219.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale, NJ: Lawrence Erlbaum Associates.

Downloaded by [The UC Irvine Libraries] at 15:51 18 December 2015