Journal of Experimental Psychology: Animal Learning and Cognition

Absence of Differential Protection From Extinction in Human Causal Learning

David N. George, Josephine E. Haddon, and Oren Griffiths

Online First Publication, April 22, 2024. https://dx.doi.org/10.1037/xan0000380

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Animal Learning and Cognition

ISSN: 2329-8456

https://doi.org/10.1037/xan0000380

Absence of Differential Protection From Extinction in Human Causal Learning

David N. George[1], Josephine E. Haddon[2], and Oren Griffiths[3]
[1] School of Psychology, University of Hull
[2] School of Psychology, Cardiff University
[3] School of Psychological Sciences, University of Newcastle

Elemental models of associative learning typically employ a common prediction-error term. Following a conditioning trial, they predict that the change in the strength of an association between a cue and an outcome is dependent upon how well the outcome was predicted. When multiple cues are present, they each contribute to that prediction. The same rule applies both to increases in associative strength during excitatory conditioning and the loss of associative strength during extinction. In five experiments using an allergy prediction task, we tested the involvement of a common error term in the extinction of causal learning. Two target cues were each paired with an outcome prior to undergoing extinction in compound either with a second excitatory cue or with a cue that had previously undergone extinction in isolation. At test, there was no difference in the causal ratings of the two target cues. Manipulations designed to bias participants toward elemental processing of cue compounds, to promote the acquisition of inhibitory associations, or to reduce generalization decrement between training and test were each without effect. These results are not consistent with common error term models of associative learning.

Keywords: extinction, causal learning, associative learning, prediction error

One of the key features of Rescorla and Wagner's (1972) model of Pavlovian conditioning is its use of a common prediction-error term. Following a conditioning trial, the change in the strength of an association between a cue and an outcome is dependent upon how well the outcome was predicted. When multiple cues are present, they each contribute to that prediction. The common error term allows the model to explain phenomena such as overshadowing (Pavlov, 1927), blocking (Kamin, 1968), conditioned inhibition (Pavlov, 1927), overexpectation (Kremer, 1978), and superconditioning (Rescorla, 1971). Several other models of associative learning have since incorporated some form of common error term (e.g., George, 2020; McLaren & Mackintosh, 2000; Pearce & Hall, 1980; Wagner, 2003).

According to the Rescorla–Wagner model, the same learning rules govern both increases in associative strength during excitatory conditioning and the loss of associative strength during extinction. Once an association between a cue and an outcome has been established, presenting the cue in the absence of the outcome will result in a negative prediction error. Consequently, the association will be weakened (or a new inhibitory cue–outcome association or a cue–no outcome association will develop; see Bouton et al., 2006). If two cues are extinguished in compound, the associative strength of each will contribute to the prediction error. Hence, when a target cue is extinguished in compound with a second excitatory cue, the prediction error should be greater than when it is extinguished in compound with a neutral cue, or by itself. As a result, the target cue should experience a greater loss of associative strength. Conversely, a cue that is extinguished in compound with a conditioned inhibitor should be protected from loss of associative strength because the prediction error would be lowered.

Studies with nonhuman animals have provided some evidence in support of these predictions. Using both Pavlovian and instrumental appetitive conditioning paradigms with rats, Rescorla (2000b) found that a target cue extinguished in compound with a second excitatory cue produced less responding at test than target cues extinguished alone or in compound with a neutral cue. In Pavlovian autoshaping experiments with rats and pigeons, Rescorla (2003) also showed that a cue extinguished in compound with a conditioned inhibitor was protected from extinction relative to a cue extinguished in compound with a neutral cue (see also McConnell & Miller, 2010; Thomas & Ayres, 2004).

There is, however, conflicting evidence concerning the role of prediction error in the extinction of cue compounds. Extinguishing a cue in compound with another excitor has been shown to confer protection from extinction in some situations in both pigeon autoshaping (Pearce & Wilson, 1991) and conditioned taste aversion in rats (Pineño, 2007; Pineño et al., 2007). Several explanations have been given for these failures of a concurrently presented excitatory cue to deepen extinction. Pearce and Wilson suggested that their results were consistent with a configural analysis of learning (e.g., Pearce, 1987, 1994, 2002) in which a cue compound is represented as a

Andrew R. Delamater served as action editor.

David N. George https://orcid.org/0000-0002-0038-2999

The authors would like to thank Dominic Dwyer, Mark Haselgrove, Robert Honey, and David Sanderson for their helpful comments on a version of this article presented at the Swansea, United Kingdom, meeting of the Experimental Psychology Society in July 2023. The experiments were not preregistered. All the experimental data and analysis code from this article are available from the Open Science Framework and can be accessed at https://osf.io/78mf5/.

Correspondence concerning this article should be addressed to David N. George, School of Psychology, University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom. Email: d.george@hull.ac.uk

pattern of stimulation that is distinct from the representations of the individual cues. These configural processes might result in a failure of extinction learning to generalize to the test trials (i.e., a generalization decrement; see also Urcelay et al., 2009). Alternatively, Witnauer and Miller (2012) concluded that the associative status of the training context is critical in determining whether a compound extinguished with a concurrent excitor will result in deepened or attenuated extinction, consistent with the predictions of the sometimes-competing retrieval model (Stout & Miller, 2007), an implementation of the extended comparator hypothesis (Denniston et al., 2001).

Human compound extinction experiments using fear conditioning preparations have produced similar, mixed results. Lovibond et al. (2000) extinguished target cues in compound with either a conditioned inhibitor or a second excitatory cue. At test, participants' shock expectancy ratings and skin conductance levels for each of the target cues were equivalent to those for a control cue that had been consistently paired with shock throughout the experiment; compound presentation protected the target cues from extinction regardless of the associative status of their partner cue. Vervliet et al. (2007) also found that extinction in compound with an excitatory cue protected a target cue from extinction. Responding at test was equivalent to preextinction levels. Two other studies, however, found that compound extinction led to deepened extinction. In both cases, the target cues were individually extinguished prior to compound extinction. Culver et al. (2015) paired two cues (A and B) with a startling auditory stimulus whereas a third cue (C) was paired with no outcome. Each cue was then extinguished individually. Different groups of participants were given further nonreinforced presentations of A alone or of Compound AB. At test, the group given compound extinction showed lower skin conductance levels in response to A, although there was no difference between the groups' expectancy ratings. Using a very similar experimental design, but with an electric shock as the outcome, Coelho et al. (2015) found that compound extinction resulted in both lower skin conductance levels and lower expectancy ratings for Cue A at test. Extinction to the individual cues might account for the difference in the results of these studies and that of Vervliet et al., by reducing generalization decrement between the extinguished Compound AB and Cue A at test. Krypotos and Engelhard (2019) employed a similar design to Culver et al. and Coelho et al., but with the difference that participants received avoidance conditioning between the fear conditioning and extinction phases. On avoidance trials, participants were able to cancel the shock outcome by pressing a key during the first 3 s of a stimulus presentation. At test, there was no difference in shock expectancy ratings or the frequency of avoidance responses between the groups given extinction to Compound AB, or extinction with Cue A alone. In this case, compound conditioning resulted in neither deepened extinction nor protection from extinction.

Finally, three studies have examined the effects of compound extinction in human causal learning. Each of these employed an allergist prediction task in which participants were presented with entries from the food diary of a fictious patient. On each trial, the participants were told that the patient had consumed one or two foods and were asked to predict whether or not the patient had suffered an allergic reaction before receiving feedback. In the first of these studies, Griffiths and Westbrook (2012) taught their participants about two patients simultaneously across three phases of learning. For Mr. Y, foods L, M, N, and O were initially paired with no outcome. In the second phase, L was paired with allergy, and in the third phase, Compound LM and NO were each paired with allergy. At test, causal ratings for Cue O were higher than those for Cue M. This blocking effect suggests that acquisition of causal learning was influenced by a common error term (although the effect can also be explained by an attentional theory of learning with no common error term such as Mackintosh, 1975). Mr. X experienced a similar sequence of events, but with the contingencies reversed. Initially, Foods A, B, C, and D were each individually paired with allergic reactions. A was then extinguished, followed by compound extinction of AB and CD. Participants' causal ratings for Cues B and D did not differ; D, which was extinguished in compound with a second excitor, did not experience deeper extinction than the B, which was extinguished in compound with an extinguished cue. Using a very similar task, Holmes et al. (2014) found a target cue extinguished in compound with a conditioned inhibitor was protected from extinction relative to a cue extinguished in compound with a neutral partner. Subsequent experiments, however, revealed that the presence of a partner cue was more important than its causal history. Protection from extinction was conferred equally by neutral cues, extinguished cues, and excitatory cues. Finally, Griffiths et al. (2017) replicated key features of the experimental designs of Vervliet et al. (2007) and Culver et al. (2015). In their first experiment, they found no difference in causal ratings for target cues extinguished in compound with a second excitatory cue or with an extinguished cue. In a second experiment, four cues were individually extinguished before receiving additional extinction either alone or in compound. At test, causal ratings were higher for the cues that had been extinguished in compound.

The results of these causal learning experiments are broadly, but not entirely, consistent with the predictions of Pearce's configural theory (Pearce, 1987, 1994, 2002). If a compound cue is represented as a distinct configuration rather than being treated simply as a collection of individual cues, extinction to the cue compound will generalize imperfectly to its constituent cues. As a result, any cue extinguished in compound should enjoy some protection from extinction dependent upon the degree of generalization between cues and compounds. The purpose of the experiments reported here was to further explore the effects of compound extinction in human causal learning. In Experiments 1 and 2, we sought to replicate key features of the results of previous studies (Griffiths et al., 2017; Holmes et al., 2014). In Experiments 3 and 4 we examined the effects of manipulations designed to bias participants toward treating cue compound as collections of individual elements rather than as distinct configurations. Finally, in Experiment 5, we combined these manipulations with a compound testing procedure intended to further reduce generalization decrement between compound extinction and testing phases.

Experiment 1

Experiment 1 was a near-exact replication of Griffiths et al.'s (2017) first experiment. The main differences were that our participants were students at a U.K. university and were tested individually in a laboratory, whereas Griffiths et al. tested Australian university students in classes of about 20 at a time. The experiment employed a facsimile of the same task in which participants were asked to review the food diary of a fictitious patient, Mrs. X, to try to learn the causes of allergic reactions suffered by her. The foods served as experimental cues, and the allergic reactions served as outcomes. The design of the experiment is shown in Table 1. Cues A–D were initially trained with a large outcome (++ = *a severe allergic reaction*) and then later

Table 1
The Design of Experiments 1 and 2

Study	Phase 1	Phase 2	Phase 3	Choice
Experiment 1	A++	A−	AB−	A versus C
	B++			B versus D
	C++		CD−	
	D++			
Experiment 2	A++	A−	AB−	A versus C
	B++	B++		B versus D
	C++	C−	CD−	
	D++	D−		
Experiments 1 and 2	E−	E++		
	F−	F−		
	G−	G−	GH++	
	H−	H−		
	I−	I+		
	J+	J+	J+	
		K+		

Note. Letters A–K represent different foods consumed by the fictitious patient Mrs. X. The severity of her allergic reactions to those foods are indicated by $-$ = *no reaction*, $+$ = *minor reaction*, and $++$ = *severe reaction*. Treatment of Cues E–K were the same in the two experiments. Following the choice test, participants were asked to rate the likelihood that each food would cause a reaction the next time the patient ate it.

extinguished in compound with each other (AB−, CD−). Cue A received extinction by itself (A−) between these phases, whereas none of the other three cues were presented in Phase 2. According to the Rescorla–Wagner model, this treatment should result in Cue A ending up with a lower causal value than Cues C and D, whereas Cue B should have been provided with some protection from extinction relative to those cues. Following training, participants were asked to select the safer food from the choices A versus C and B versus D. Since the design did not include a control stimulus that was extinguished alone and some previous studies have found that any cue extinguished in compound might enjoy some protection from extinction, the comparison between Cues B and D provided a measure of differential protection from extinction as a function of the causal status of their partner cues (A and C). The choice between the partner Cues A and C served as a manipulation check for the expected difference in causal value. Finally, they rated the likelihood that Mrs. X would suffer an allergic reaction the next time she ate each of the individual foods. Griffiths and colleagues (Griffiths et al., 2017; Griffiths & Westbrook, 2012; Holmes et al., 2014) noted that participants tended to aggregate their experience with cues across phases, despite explicit instructions to base their rating at test on the current value of the cues. Trials with filler cues were included to assess this tendency, and the task instructions stressed that Mrs. X's reactions to some food might change over time and that it was critical that test ratings reflected what participants believed at the end of the food diary.

Method

Participants

A total of 55 undergraduate psychology students at the University of Hull served as participants in exchange for course credit. The mean age of the participants was 20.7 years ($SD = 4.3$, range = 18–41) and 43 were female. Power sensitivity calculations using G*Power 3.1 (Faul et al., 2009) indicated that for a comparison between two dependent means with $\alpha = .05$, a sample size of 52 was sufficient to detect a small to medium effect size (Cohen's $d = 0.35$) with power $(1 - \beta) = .80$. This is the size of the protection from extinction effect reported by Holmes et al. (2014) using a related design.

Design

The design of the experiment is shown in Table 1. Each of four cues, A–D, was individually paired with a serious allergic reaction (++) in Phase 1, and then in Phase 3 was paired with no reaction (−) in compound with another of the four cues (in the pairings AB and CD). The treatment of Cue A differed from Cues B–D in that it was also presented during Phase 2, when it was individually paired with no reaction. At the start of Phase 3, participant's might have predicted a greater reaction to Compound CD, consisting of two cues both of which predicted an increase in antibody level, than to Compound AB, one of which's components had undergone extinction. This difference should have two effects. First, the greater prediction error should result in a greater reduction in the causal value of Cues C and D, than of Cues A and B during Phase 3. Second, because Cue A no longer a predicted an allergic reaction at the end of Phase 2, it might have gained negative causal value during Phase 3. That is, Cue B should have been protected from extinction, whereas Cue A should have experienced deeper extinction, relative to Cues C and D.

The remaining cues, E–K were included to ensure that participants learned the relationship between individual foods and Mrs. X's reactions. They balanced the number of trials on which the patient experienced a reaction or not during each phase and ensured that not all compounds were associated with no reaction in Phase 3. They exposed participants to a range of outcomes (−, +, ++) during training with the intention that participants might understand that different cues could cause greater or smaller allergic reactions. They also allowed us to evaluate whether participants' ratings during the test phase were biased toward the most recent outcome associated with a food, or whether participants aggregated their experience with a food across phases of training. Cues G and H were each paired with no outcome in Phases 1 and 2, but with a large outcome in Phase 3. If we assume that learning approached asymptote by the end of Phase 3, then the Rescorla–Wagner model would predict that each of the cues should have a moderate level of causal strength—approximately equivalent to the value of the minor allergic reaction. In contrast, Cue J was paired with a moderate outcome (+ = *minor allergic reaction*) across all three phases. Hence, the causal strengths of these three cues should be approximately equal. Cues I and K were both paired with a moderate outcome in Phase 2 but Cue I had previously been paired with no outcome in Phase 1. If we again assume that learning was approaching asymptote by the end of Phase 2, the two cues should have equivalent causal strength. There is, however, some evidence that people sometimes base their final test ratings on the aggregation of their experience of cues across an entire experiment in a way that they don't for online responses collected during learning (Collins & Shanks, 2002). In that case, their test ratings for G and H would be expected to be lower than ratings for J, and their ratings for I would be lower than for K.

Stimuli and Materials

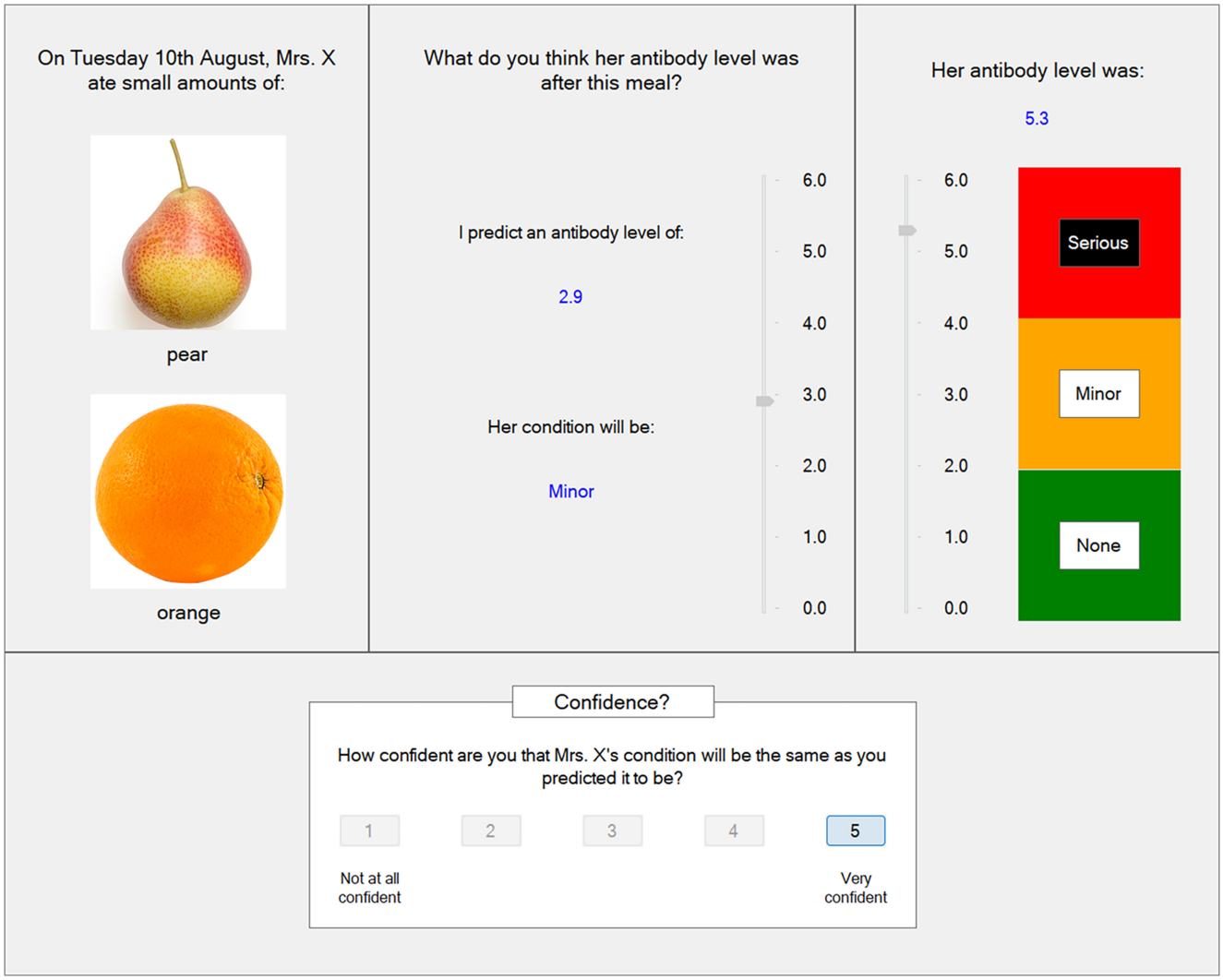
The experiment was conducted using personal computers running the Windows 10 operating system and programmed in Visual Basic

15 using Microsoft Visual Studio 2017 (Microsoft Corporation, Redmond, Washington, United States). Stimuli were presented on 5:4 aspect ratio iiyama ProLite 1906S 48-cm TN-TFT monitors (iiyama Corporation, Tokyo, Japan) with a native resolution of 1,280 × 1,024 pixels (width × height) and 60-Hz refresh rate connected to Nvidia GeForce GT 430 graphics cards (Nvidia Corporation, Santa Clara, California, United States). Responses were made using a standard optical mouse. Stimuli comprised the names of 40 common foods (see Appendix) and photographs of each food (original resolution of 400 × 400 pixels) taken against a white background. At the beginning of each experimental session, a different set of 11 of the foods were randomly assigned to Cues A–K by the computer for each participant.

For the training task, the computer screen was divided into four panels, as shown in Figure 1. Three of these were arranged in a row across the top two thirds of the screen, and the fourth entirely occupied the lower third. The top-left panel extended across 30% of the screen width and was used to display entries from Mrs. X's food diary. It contained the text "On [Sunday 16th October], Mrs. X ate (a) small amount(s) of:" where the date incremented by 1 day across successive trials and the date on the final training trial was the day before the experiment took place. The text was adjusted depending upon whether the meal consisted of a single food or two foods. Below this text, photographs (reduced to a resolution of 216 × 216 pixels) and names of one or two foods were displayed in a vertical arrangement. The top-middle panel took up 40% of the screen's width. Text at the top of this panel read "What do you think her antibody level was after this meal?." To the right of the panel there was a vertical slider marked with "0.0" at the bottom, "6.0" at the top, and five equal increments labeled with the numbers

Figure 1

A Screen Capture Showing a Trial From Phase 3 of Experiment 1



Note. Participants were shown one or two foods and their name(s) in the upper left panel of the display. They were then required to move a slider in the upper middle panel to make their outcome prediction before clicking on one of the five buttons in the lower panel to indicate the confidence they had in this prediction. Once they had done this, the correct outcome for that trial was displayed in the upper right panel. See the online article for the color version of this figure.

"1.0" to "5.0" between them. Participants could move this slider in steps of 0.1 between the lower and upper limits. To the left of the slider, its current value and an associated condition (0.0–2.0 = *none*, 2.1–4.0 = *minor*, and 4.1–6.0 = *serious*) were shown in blue (RGB: 0, 0, 255) text. A feedback panel occupied the right 30% of the screen. In this panel there was a second slider which the participant could not move, next to a scale showing the conditions associated with different antibody levels. A green (RGB: 0, 128, 0) rectangle with the text "None" extended from 0.0 to 1.95 along the scale, an orange (RGB: 255, 165, 0) rectangle with the text "Minor" was alongside values 1.95–4.05 on the scale, and a red (RGB: 255, 0, 0) rectangle with the text "Serious" was next to the values 4.05–6.0. At the end of each trial, the slider in the feedback panel moved to indicate Mrs. X's antibody level, the label in the appropriate rectangle flashed, and the text "Her antibody level was:" appeared at the top of the panel along with the outcome value. The lower panel, occupying the bottom third of the screen contained five buttons labeled "1," "2," "3," "4," and "5," below the text "How confident are you that Mrs. X's condition will be the same as you predicted it to be?." Below the button labeled "1" was the text "Not at all confident," and under the button labeled "5" was the text "Very confident."

On each choice test trial, the names and pictures of two foods were presented, one on either side of the screen. Above them was the text "Which of these two meals should Mrs. X consume? (i.e., which is least likely to cause her to suffer an allergic reaction?)." Next to the picture of each food was a button labeled "This meal."

The display for the rating test was split into three panels. The top-left panel was the same size as for the training task and was used to display the picture and name of a single food. Next to this was a rating panel which occupied the rest of the upper two thirds of the screen. At the top of this panel was the text "How likely is this food to produce an allergic reaction in Mrs. X right now?." The panel also contained a rating slider identical to the prediction slider from the training task with the same seven numerical labels (0.0–6.0) spaced equally along its length. The text "Very unlikely to produce a reaction next time she eats it" and "Very likely to produce a reaction next time she eats it" appeared at the ends of the slider next to the 0.0 and 6.0 points, respectively. Participants could move the slider position in increments of 0.1. A button labeled "Next food" was positioned in the bottom-right corner of the lower panel which was otherwise empty.

All text was displayed in the Microsoft Sans Serif font at a size of 12, 14, or 16 points and was black unless otherwise stated. The screen background was a light grey (RGB: 240, 240, 240). When the screen was divided into panels, black lines marked their boundaries.

Procedure

The procedure was closely modeled on that used by Griffiths et al. (2017), with the exception that participants were tested individually in a dedicated laboratory at the University of Hull. Experimental protocols for all five experiments were approved by the Faculty of Health Sciences Research Ethics Committee of the University of Hull.

At the start of the experiment, each participant was given the following instructions which were presented on the computer screen and were also read aloud by the experimenter:

> It is your job to assume the role of a doctor who specialises in food allergies. You have a new patient (Mrs. X) who is currently undergoing chemotherapy. She has come to you because the chemotherapy is affecting her immune system, resulting in a strange collection of food allergies. In fact, her allergies appear unstable, perhaps as a result of her constantly changing drug regimen. This means that she sometimes acquires food allergies spontaneously, and sometimes food allergies will suddenly disappear. She has come to you to try and learn which foods she should avoid. To help diagnose her specific allergies, you've provided Mrs. X with a blood analyser that measures her antibody levels (a measure of allergic reaction severity). You have asked Mrs. X to eat a simple diet, and record (i) the foods she eats and (ii) the antibody levels that these foods produce. Mrs. X has meticulously documented all foods and antibody measurements over a four-month period.
>
> You will review this information. To test yourself along the way, after each meal, you'll be asked to predict whether Mrs. X will have a reaction, and if so, how bad that reaction will be. Also, you'll be asked to rate how confident you are of that prediction (on a scale of 1–5).

Each training trial began with the presentation of a meal consisting of one or two foods in the top-left panel. Both the prediction slider and the feedback slider were set to 0.0. The participant was required to move the prediction slider in the top-center panel of the screen to predict the value of the outcome they believed was associated with that meal. They then had to rate their confidence in this prediction by clicking on one of five buttons labeled 1–5 in the lower panel. There was no limit on the amount of time that participants could take to make these responses. Once a button was clicked, the slider in the top-center panel was inactivated and feedback was presented in the top-right panel. The patient's actual antibody level following the meal was displayed in text, on a slider, and the label corresponding to the severity of her allergic reaction flashed. This antibody level was randomly jittered around a central value for each reaction. Hence, on no-reaction (−) trials, feedback was randomly selected from a uniform distribution in the range 0.4–1.3 ($M = 0.85$), for minor reaction (+) in the range 2.5–3.4 ($M = 2.95$), and for severe reaction (++) in the range 4.6–5.5 ($M = 5.05$). This feedback was shown for 1.5 s after which the screen went blank for 0.5 s before the next trial started automatically. Each phase of the experiment was divided into eight blocks of trials. In each block, each of the trial types was presented once in a random order, giving a total of 176 trials.

Following the completion of training, participants immediately underwent testing. The following instructions were presented on the screen:

> You've now studied all of Mrs. X's records. You should have a fairly good idea of which foods make her feel ill, and which do not. This is important because Mrs. X is about to undergo a particularly difficult treatment, during which time it is vitally important that she doesn't experience any allergies at all. That is, it is important to determine which foods are likely to be dangerous to her right now.
>
> First, the hospital has provided Mrs. X with a choice of two meals. You must choose which of these is least likely to make her feel ill. Once that decision is made, you will be asked to advise her regarding how dangerous the other foods are to her right now.

Participants then completed two choice trials. On the one hand, the choice was between Cues A and C, and on the other hand, the choice was between Cues B and D. The order in which these choices were presented was determined randomly for each participant, and so was the position of the two foods on the left and right of the screen. In a final rating test, the picture and name of each food was presented individually, and the participant was asked to rate how likely that food was to cause an allergic reaction in Mrs. X at

that point in time. There were 11 rating trials on which the Cues A–K were presented, once each, in random order.

Data Analysis

Outcome predictions and confidence ratings were collected on each trial of the three phases of training. A participant's data were excluded from further analysis if their mean outcome predictions for Cues A–D were lower than the midpoint of the minor reaction (+) outcome (2.95) averaged over the last four blocks of Phase 1. This was very similar to the exclusion criterion employed by Griffiths et al. (2017). Also in keeping with Griffiths et al. (2017), the only other analysis of the prediction and confidence ratings was to compare responses to Compounds AB and CD at the beginning of Phase 3. Because A had undergone extinction by itself during Phase 2, outcome predictions for Compound AB were expected to be lower than for Compound CD. These comparisons were made using paired Student's *t* tests.

During the choice test, each participant received one trial in which they were asked to choose between Cues A and C and one trial in which they were asked to choose between B and D. Data from these two trials were analyzed using binomial tests. Cue A was paired with no reaction (−) over Phases 2 and 3, whereas Cue C was extinguished only in Phase 3. It was the difference in the causal status of these cues at the start of Phase 3 that might have resulted in Cues B and D experiencing differential changes in their causal status. We therefore conducted a third binomial test on the choices between Cues B and D, only for those participants who believed that A was safer than C.

At the end of the experiment, participants rated each of the 11 foods. A series of planned contrasts were then carried out on the ratings for Cues A–D using paired Student's *t* tests. The first two of these comparisons were manipulation checks to ensure that the outcomes associated with the cues in Phases 2 and 3 affected rating in the expected way. First, the average rating for all four cues was compared with the ratings for Cue E, which had most recently been paired with a serious reaction (++), to determine whether extinction of those cues had been effective in reducing ratings. Second, ratings for Cue A were compared with the average ratings for Cues B–D to assess whether two phases of extinction resulted in lower causal ratings than a single phase of extinction. The third comparison was the critical test to determine whether Cue B, which was extinguished in compound with the already extinguished A, enjoyed greater protection from extinction than Cue D, which was extinguished in compound with Cue C than had not undergone prior extinction. Although Cues C and D were treated equivalently throughout the three phases of training, there was an asymmetry in their treatment in the forced choice test. Nevertheless, we also conducted an additional *t* test to compare ratings for Cue B with the average of the ratings for Cues C and D. This comparison was made for consistency with Griffiths et al. (2017).

Two further *t* tests were conducted to examine whether participants based their ratings on their most recent experience with a cue, or whether they aggregated their experience of the cue–outcome relationships over the three phases of training. In the first test, the average of the ratings for Cues G and H were compared with the rating for Cue J. G and H were presented in compound and paired with a severe reaction (++) during Phase 3, whereas J was paired with a minor reaction (+). If participants' ratings were based on this training alone, one might expect the cues to all be rated similarly. G and H were, however, paired individually with no reaction (−) during Phases 1 and 2, whereas J was paired with the same minor reaction in all three stages. Aggregation of this experience should have resulted in lower ratings for G and H than for J. The logic for the second test was similar. Prior to the rating test, Cues I and K were most recently paired with a minor reaction during Phase 2. Cue I had previously been paired with no reaction during Phase 1, when K was not presented at all. Hence, if participants based their ratings on their most recent experience with a cue, ratings for I and K should not have differed. If, however, their ratings were based on an aggregation of their experience with the cues, then ratings for I should have been lower than those for K.

Each of these planned comparisons between ratings was further analyzed by computing a Bayes factor (BF_{10}) evaluating the relative weight of evidence in favor of the alternative hypothesis over the null using the default Cauchy prior. Bayes factors were interpreted according to the scale provided by Kass and Raftery (1995) based on Jeffreys (1961). Values close to 1 provide no clear support in either direction. Those >3, >20, and >150 provide positive, strong, or very strong support for the alternative hypothesis, respectively. Values $<1/3$, $<1/20$, and $<1/150$ provide positive, strong, or very strong support for the null hypothesis.

Both *t* tests and binomial tests were two-tailed, and $\alpha = .05$ unless stated otherwise. Bayes factor analyses were also two-tailed. Data were analyzed using R 4.3.1 (R Core Team, 2023) and the packages afex (Singmann et al., 2023), apaTables (Stanley, 2021), BayesFactor (Morey & Rouder, 2022), broom (Robinson et al., 2023), and tidyverse (Wickham et al., 2019). Data figures were prepared using MATLAB (Version 9.14 [R2023a]; The Mathworks Inc., Natick, Massachusetts, United States) and the Statistics and Machine Learning Toolbox (Version 12.5 [R2023a]; The Mathworks Inc., Natick, Massachusetts, United States), and GIMP (Version 2.10.34; The GIMP Development Team).

Transparency and Openness

All the experimental data and analysis code from this article are available from the Open Science Framework and can be accessed at https://osf.io/78mf5/ (George et al., 2024). The experiments were not preregistered.

Results

Exclusion Criterion

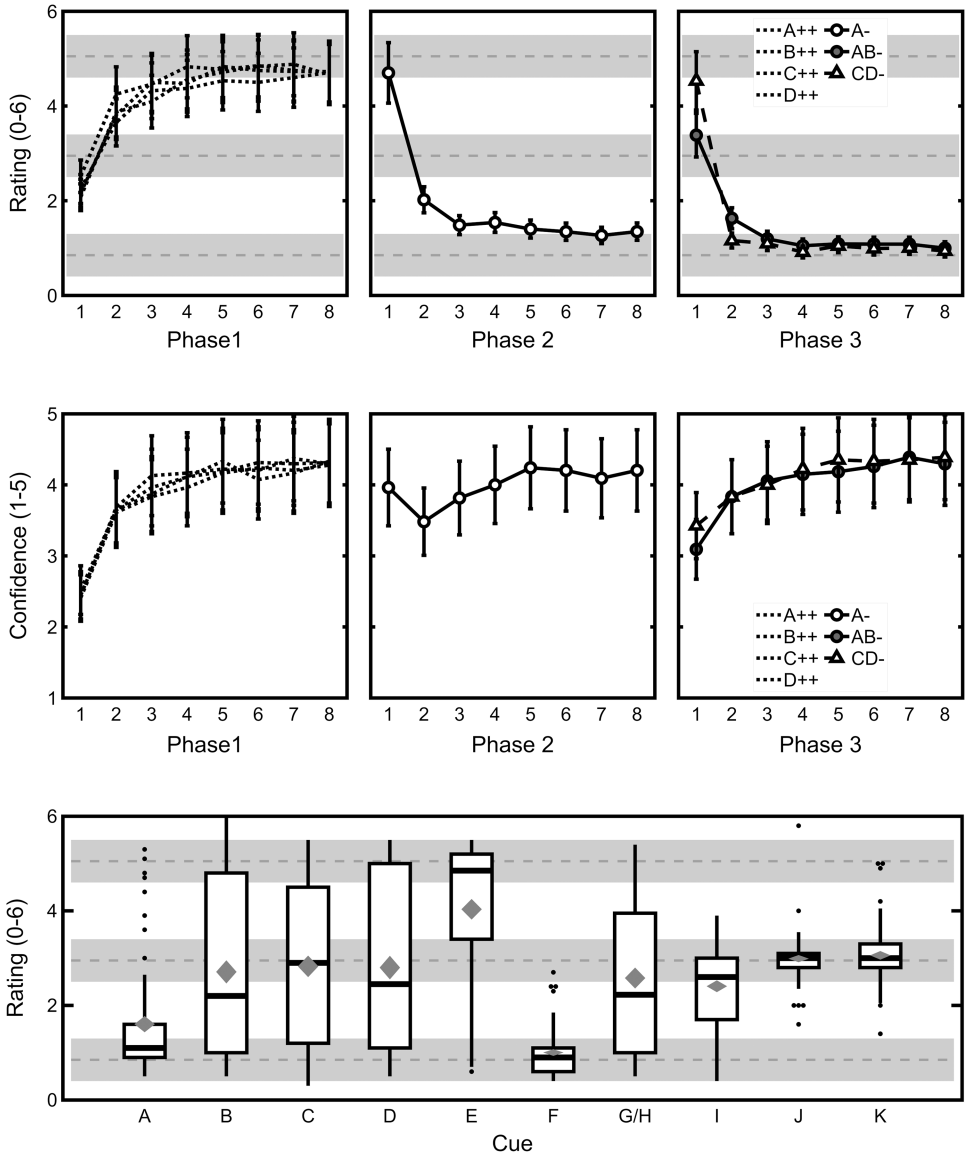
A single participant failed to meet the Phase 1 learning criterion. The reported analysis was performed on the data from the 54 remaining participants.

Training

Outcome predictions and confidence ratings for Cues A–D during each of the three phases of training are shown in the top two rows of Figure 2, respectively. Both outcome predictions and confidence ratings increased across Phase 1 where each cue was paired with a severe reaction and by the end of the phase the mean predictions were within the range of values associated with that outcome (4.6–5.5). During Phase 2, Cue A was paired with no reaction and ratings declined. Following the first pairing of A with this unexpected absence of a reaction, confidence ratings dipped but recovered soon after. Following extinction of Cue A, outcome predictions for Compound AB at the start of Phase 3 were lower than predictions

Figure 2

Results From Experiment 1



Note. The top panel shows participants' mean outcome predictions across the three phases of training for trials on which Cues A–D and their compounds (AB and CD) were presented. Error bars represent standard error of the mean. The gray shaded bars indicate the range of antibody values associated with the three outcomes, and the gray dotted lines show the midpoint for each outcome. Participants' mean confidence ratings across the three phases are shown for the same stimuli in the middle panel. Test ratings are presented in the bottom panel. The boxes range from the first to the third quartile of the distribution of scores for each cue, representing the interquartile range. Median prediction ratings are indicated by thick horizontal black lines, and the whiskers extend from the boxes to the most extreme ratings. Outliers, which differ from the first or third quartile by more than 1.5 times the interquartile range, are represented by dots. The gray diamonds are centered on the mean prediction rating for each cue, and their heights indicate standard error of the mean.

for Compound CD, $t(53) = 4.00$, $p < .001$, $M_{\text{difference}} = 1.15$, 95% confidence interval (CI) of the mean difference [0.57, 1.72], $\text{BF}_{10} = 120$. There was, however, no difference in participants' confidence in these ratings, $t(53) = 1.90$, $p = .063$, $M_{\text{difference}} = 0.33$, $\text{BF}_{10} = 0.79$. Outcome predictions for each compound declined over the course of Phase 3.

Forced Choice Test

Given a choice between Cues A and C, which differed in the number of phases across which they were extinguished, 43 out of 54 participants (80%) thought that A was the safer food. A binomial test found that this distribution of choices was significantly different

from chance ($p < .001$, two-tailed). Twenty-five of the 54 participants (46%) selected D as being safer than B on the other choice test trial. This proportion did not differ from chance ($p = .683$). Of the 43 participants who chose A as safer than C, 21 (49%) also chose D ($p = 1$). Hence, there was no evidence that Cue B was protected from extinction relative to Cue D.

Ratings Test

Summary statistics for the 11 cues are shown in the bottom row of Figure 2. Ratings are collapsed across Cues G and H, because they received identical treatment throughout the experiment. Four planned t tests were used to assess the effects of extinction of Cues A–D. The first confirmed that extinction was effective. Mean ratings for the four extinguished cues were lower than for Cue E which had most recently been paired with a severe reaction, $t(53) = 5.50$, $p < .001$, $M_{\text{difference}} = 1.55$, 95% CI [0.98, 2.11], $BF_{10} > 10^4$. The second t test found that two phases of extinction were more effective than one. Ratings for Cue A were lower than mean ratings for the other three cues extinguished in compounds B–D, $t(53) = 5.54$, $p < .001$, $M_{\text{difference}} = 1.17$, 95% CI [0.75, 1.60], $BF_{10} > 10^4$. The third test revealed no evidence that B was protected from extinction relative to Cue D. Ratings for B did not differ from ratings for Cue D, $t(53) = 0.34$, $p = .734$, $M_{\text{difference}} = 0.09$, $BF_{10} = 0.16$, or the mean ratings for Cues C and D, $t(53) = 0.44$, $p = .664$, $M_{\text{difference}} = 0.10$, $BF_{10} = 0.16$. Hence, the Bayes factor analyses indicated positive evidence in support of the null hypothesis that Cues B and D were not subject to differential protection from extinction.

The final two t tests provided mixed evidence that participants' ratings were influenced by the treatment of cues across all stages of the experiment. First, mean ratings for Cues G and H were found to not differ significantly from ratings for Cue J, $t(53) = 1.86$, $p = .068$, $M_{\text{difference}} = 0.41$, $BF_{10} = 0.74$. In a second test, however, ratings for K were found to be significantly greater than ratings for I, $t(53) = 4.11$, $p < .001$, $M_{\text{difference}} = 0.65$, 95% CI [0.33, 0.97], $BF_{10} > 150$.

Discussion

Our results very closely replicated of those of Griffiths et al. (2017). At the beginning of Phase 3, ratings for Compound CD, which consisted of two cues that had previous been paired with a large outcome, were higher than for Compound AB, an element of which had undergone extinction. In the forced choice test, participants tended to select Cue A as being safer than Cue C, but they showed no preference for Cue D over Cue B (or vice versa). The rating test revealed that the cue that had undergone two phases of extinction, first by itself and then in compound (A), had a lower causal value than the cues that had only undergone compound extinction (B, C, and D). The cue that underwent compound extinction in partnership with an already extinguished cue (B) was not protected from extinction relative to a cue that were extinguished in compound with another excitatory cue (D). We also found some evidence that participants based their test rating on aggregated experience of the cues across phases of training, although that evidence was not as strong as that found by Griffiths et al.

Experiment 1 failed to support one prediction of the Rescorla–Wagner model: that reducing prediction error by extinguishing one component of a cue compound should reduce the effectiveness of compound extinction. In Experiment 2, we tested a second, related, prediction of the model: that restoring prediction error during compound extinction should result in deeper extinction learning.

Experiment 2

Culver et al. (2015) separately conditioned three cues as signals for a startling noise (A+, B+, C+) before extinguishing each cue individually (A−, B−, C−). For different groups of participants, additional extinction trials were given with one of the cues either alone (A−) or in compound with a second (AB−). At test, skin conductance levels following the presentation of this target cue (A) were lower in the group given compound extinction. These results suggest that following the initial individual extinction trials, the cues maintained some association with the aversive outcome. Thus, by presenting the cues in compound, prediction error was restored, allowing additional extinction learning to take place. Indeed, there is evidence from animal experiments that even after extensive extinction training cues retain some associative strength (Hendry, 1982; Reberg, 1972; Rescorla, 2006). Nevertheless, Culver et al only found an effect of compound extinction on skin conductance levels but not participants' expectancy ratings. In their second experiment, Griffiths et al. (2017) used a within-subject version of Culver et al.'s design and also found no effect of compound extinction on participants' predictions in their allergist task.

Holmes et al. (2014) also found little evidence that restoring prediction error affected learning during compound extinction trials. In their third experiment, Cues A–E were initially each individually paired with an outcome whereas Cue F was not (A+, B+, C+, D+, E+, F−). In a second phase of the experiment, the three critical Cues A, C, and E were each extinguished, as was D (A−, C−, D−, E−). Finally, the six cues all underwent compound extinction (AB−, CD−, EF−). According to the Rescorla–Wagner model, prediction error at the start of Phase 3 should have been greatest for Compound AB (one of whose components, B, was still an excitor) than for either of the other compounds. This should have led to a greater loss of causal value for A than for either C or E. A subsequent rating test revealed no difference between causal ratings of the three cues.

Experiment 2 employed a design very similar to the one we used in Experiment 1, but manipulated the causal status of cues in a manner similar to Holmes et al. (2014). Two target cues (A and C) were individually extinguished. Each then underwent compound extinction either alongside a partner cue that had also been extinguished, or one that had not. The key difference between this experiment and that of Holmes et al, was that we continued to pair the nonextinguished cue with the outcome during the second phase of the experiment to maintain its causal value. Hence, the only difference between the designs of Experiment 1 and Experiment 2 was the treatment of Cues A–D during the second phase of training. Cues A, C, and D were individually extinguished whereas Cue B was paired with a severe allergic reaction (A−, B++, C−, D−). If a common-error term governs learning during compound extinction, we would expect ratings for Cue A to be lower than for Cue C at test.

Method

Participants

Fifty-three undergraduate psychology students at the University of Hull participated in Experiment 2 in exchange for course credit.

They had a mean age of 21.4 years (range = 18–42, $SD = 5.2$) and 42 were female.

Design

The design of Experiment 2 is shown in Table 1. The only differences between Experiments 1 and 2 concerned the treatment of Cues B–D during Phase 2. During this phase, Cues C and D underwent individual extinction in the same manner as Cue A. Cue B, however, continued to be paired with a serious reaction (++). Hence, at the start of Phase 3 Cue B should have had a higher causal value than Cue D, which should have resulted in a Cue A undergoing greater extinction than Cue C.

Stimuli and Materials

All stimuli and materials were the same as for Experiment 1.

Procedure

The procedure was the same as for Experiment 1, with the exception of some training contingencies in Phase 2, as detailed in Table 1. Each trial type was presented once during each of the eight blocks of trials in each phase of the experiment for a total of 200 trials.

Data Analysis

Data were analyzed in the same manner as for Experiment 1 with the exception that ratings for Cue A were compared with the ratings for Cues C. These cues had undergone individual extinction followed by compound extinction and differed only in the causal status of their compound extinction partners. We also compared the mean ratings for Cues A, C, and D which had each undergone two phases of extinction with those for Cue B which had been paired with a strong reaction during Phase 2 prior to a single phase of compound extinction. The purpose of this comparison was to check that participants' ratings were affected by phase two training and were not determined purely by their initial and/or most recent experience with a cue. Predictions for the nonreinforced Cues A, C, and D, at the end of Phase 2 were compared to predictions for the compounds containing these cues (AB and CD) at the start of Phase 3 using a two-way analysis of variance (ANOVA) to assess generalization decrement between individual cues and their compounds.

Results

Exclusion Criterion

The same exclusion criterion was applied as in Experiment 1. All participants met this criterion, and data from all 53 were analyzed.

Training

Outcome predictions and confidence ratings for Cues A–D over the three phases of Experiment 2 are shown in the top two rows of Figure 3. Both predictions and confidence ratings increased rapidly over the first half of Phase 1 before stabilizing. Outcome predictions for Cues A, C, and D declined across Phase 2, while those for Cue B suffered a small initial decline before recovering. At the start of Phase 3, outcome predictions for the Compound AB (consisting of one cue that had been extinguished in Phase 2, and one that had been paired with a large outcome) were significantly greater than for Compound CD (both of the components of which had been extinguished in Phase 2), $t(52) = 8.62$, $p < .001$, $M_{\text{difference}} = 1.99$, 95% CI [1.53, 2.45], $\text{BF}_{10} > 10^8$. There was no difference in the confidence that participants had in their ratings for the two compounds, $t(52) = 1.64$, $p = .107$, $M_{\text{difference}} = 0.25$, $\text{BF}_{10} = 0.52$.

Predictions for Compound AB at the beginning of Phase 3 were significantly lower than for Cue B at the end of Phase 2, $t(52) = 3.66$, $p < .001$, $M_{\text{difference}} = 0.87$, 95% CI [0.39, 1.34], $\text{BF}_{10} = 44.9$. Conversely, ratings for Compound CD at the start of Phase 3 were numerically higher than those for the greater of Cues C and D at the end of Phase 2, but this difference was not significant, $t(52) = 1.98$, $p = .054$, $M_{\text{difference}} = 0.45$, Šidák corrected $\alpha = .025$, $\text{BF}_{10} = 0.90$. Hence, there was evidence of incomplete generalization from Cue B to Compound AB, but not of summation for Cues C and D.

Forced Choice Test

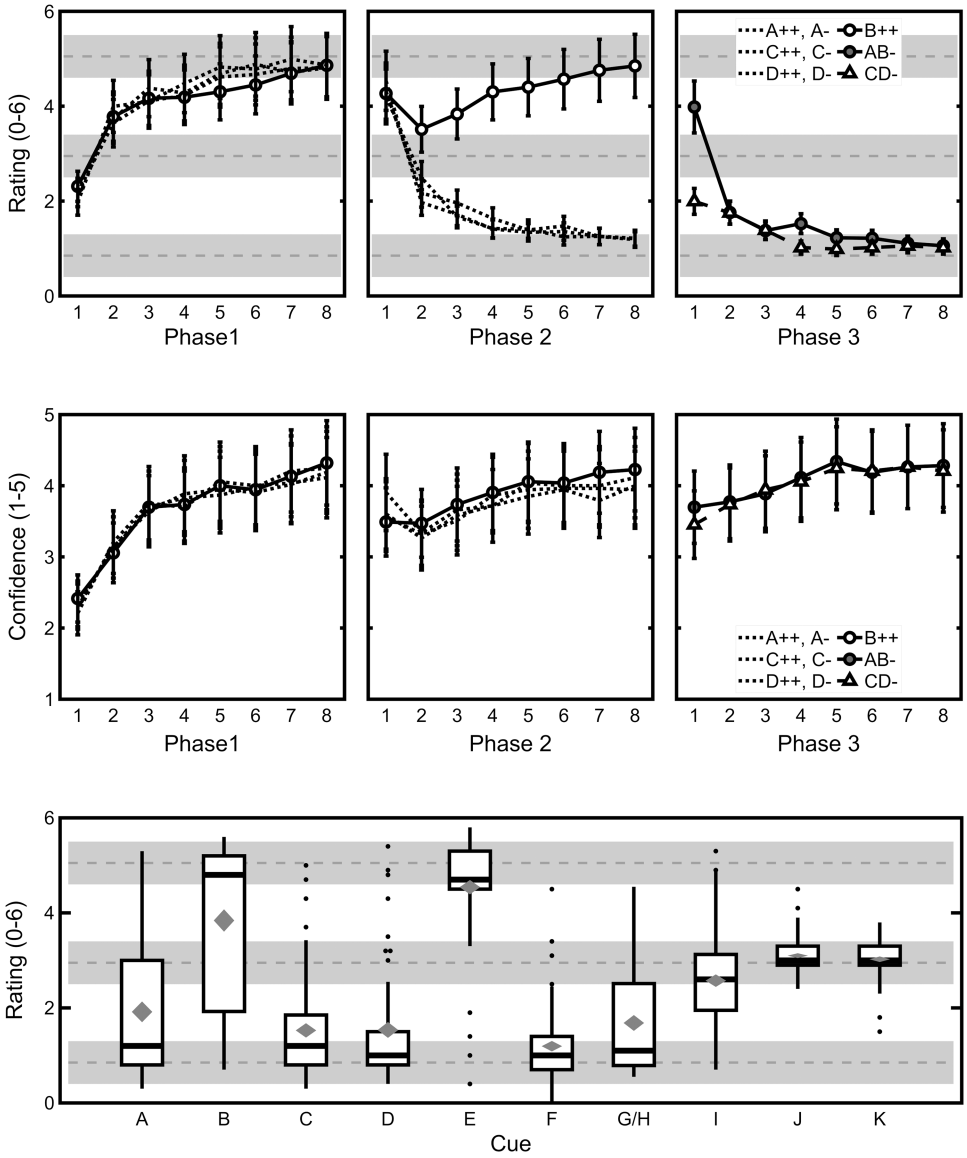
When given a choice between Cues A and C, 21 of the participants (40%) chose A as the safer alternative ($p = .169$). Forty-six (87%) chose D as safer than B ($p < .001$). The prediction that A should have experienced a greater loss in causal value than C was dependent upon A's partner (B) having a greater causal value than C's partner (D) at the start of Phase 3. Hence, a preference for A over C might have been expected only in those participants who identified D as being safer than B. In fact, the opposite was true; of the 46 participants who identified D as being safer than B, only 15 (33%) thought that A was safer than C ($p = .026$). Extinguishing Cue A in compound with an excitor (B) did not result in a greater loss of causal strength than that experienced by Cue C which was extinguished in compound with another extinguished cue (D).

Ratings Test

Summary statistics for each cue are shown in the bottom row of Figure 3. Extinction training was effective. Mean ratings for the four cues that underwent compound extinction (A–D) were lower than for Cue E which did not, $t(52) = 10.73$, $p < .001$, $M_{\text{difference}} = 2.34$, 95% CI [1.90, 2.77], $\text{BF}_{10} > 10^{11}$. Three of the cues, A, C, and D, were extinguished over two phases: individually in Phase 2 and in compound in Phase 3. Mean ratings of these cues were lower than that of B which was paired with a large outcome in Phase 2 before undergoing compound extinction in Phase 3 (in partnership with A), $t(52) = 8.61$, $p < .001$, $M_{\text{difference}} = 2.18$, 95% CI [1.67, 2.69], $\text{BF}_{10} > 10^8$. During Phase 3, Cues C and D were extinguished in compound with each other. That is, each was extinguished in partnership with an already extinguished cue. Cue A, however, was extinguished during Phase 3 in compound with B which was a signal for a serious allergic reaction. This difference in the reinforcement history of the cues' compound extinction partner had no effect on causal ratings at the end of Phase 3. Ratings for Cue A did not differ significantly from ratings for Cue C, $t(52) = 1.53$, $p = .133$, $M_{\text{difference}} = 0.39$, $\text{BF}_{10} = 0.44$, or from mean ratings for Cues C and D, $t(52) = 1.70$, $p = .095$, $M_{\text{difference}} = 0.39$, $\text{BF}_{10} = 0.58$.

As in Experiment 1, there was some evidence that participants' ratings of cues were based on an aggregation of their experience with those cues across the three phases of training. Ratings for Cue I were lower than for K, even though both cues had most

Figure 3
Results From Experiment 2



Note. The top panel shows participants' mean outcome predictions across the three phases of training for trials on which the Cues A–D and their compounds (AB and CD) were presented. Participants' mean confidence ratings across the three phases are shown for the same stimuli in the middle panel. Test ratings are presented in the bottom panel for each cue.

recently been paired with the same (minor reaction) outcome, $t(52) = 3.09$, $p = .003$, $M_{\text{difference}} = 0.45$, 95% CI [0.16, 0.74], $\text{BF}_{10} = 9.98$. Similarly, mean ratings for Cues G and H were lower than those for J, $t(52) = 8.02$, $p < .001$, $M_{\text{difference}} = 1.42$, 95% CI [1.06, 1.77], $\text{BF}_{10} > 10^7$.

Discussion

This experiment provides a conceptual replication of aspects of Holmes et al.'s (2014) results; there was no difference in the ratings for cues extinguished in compound with an excitor or with an already extinguished cue. According to the Rescorla–Wagner model, prediction error should be greater in the former case, leading to greater extinction and lower ratings. In fact, those differences we did observe were in the opposite direction; ratings for A were numerically higher than for C and D. The difference was small, but given that ratings were very low, it is possible that a significant difference was masked by a floor effect. In the choice test participants again showed a nonsignificant preference for C over A. When we considered only those participants who, at test, showed awareness that B had a stronger association with an allergic reaction than D, this preference for C was statistically significant. One way to explain these

results is in terms of within-compound associations. During compound cue presentations in Phase 3, participants may have formed associations between Cues A and B (and between Cues C and D). When each cue was subsequently presented alone, it might have evoked a memory of its compound partner which could have affected participants' expectation of an allergic reaction. At the end of the experiment, the causal value of Cue A might have been the same, or lower, than that of Cues C and D, but participants' responses to A were influenced by the higher causal value of Cue B.

There is evidence that within-compound associations may be acquired during human causal learning. Dickinson and Burke (1996) suggested that such associations are the basis of retrospective revaluation effects such as backward blocking (e.g., Chapman, 1991; Shanks, 1985). It is not, however, clear under exactly which conditions they will be formed. Aitken et al. (2001) observed no effect of within-compound associations in a food allergist task when compound conditioning followed training with the individual cues. Luque et al. (2013) also used a food allergist task but manipulated participants' existing knowledge of food pairings. For some participants, cue compounds consisted of foods not usually eaten together (such as grapes and noodles) whereas for others, foods were chosen because they go together (e.g., macaroni and cheese). There was no evidence of retrospective revaluation effects for the former group of participants even after 30 presentations of each cue compound. In our experiment foods were randomly assigned to different cues, and so it is unlikely that foods presented in compound were consistently ones commonly eaten together for a significant proportion of the participants.

Experiment 3

An alternative explanation for our results, and those of previous studies of compound extinction in human causal learning (Griffiths & Westbrook, 2012; Griffiths et al., 2017; Holmes et al., 2014), is that participants do not process cue compounds in an elemental manner. That is, they might treat compound stimuli as configurations of cues that are distinct from their component parts rather than simply summing the causal values of the cues in order to generate predictions about outcomes. Such configural processing would reduce generalization from the compound extinction trials to the test trials where the cues were presented individually. There is evidence for such a generalization decrement in those previous studies: prediction ratings taken for Cues A–D at test were somewhat higher than corresponding predictions from the end of the compound extinction phases, and we observed this same effect in Experiment 1. Interestingly, cues that were also extinguished in isolation received low ratings at test where no generalization decrement would be expected. In Pavlovian conditioning experiments with rats, Urcelay et al. (2009) also observed no effect of compound extinction training. They concluded that extinction in compound with a second excitor did in fact deepen extinction, but that effect was counteracted by a generalization decrement.

There is substantial evidence that in different situations people might process cue compounds either elementally or configurally (Melchers et al., 2004, 2005, 2008; Williams & Braker, 1999; Williams et al., 1994), but that they are naturally biased toward configural processing (Mehta & Russell, 2009; Williams & Braker, 1999; Williams et al., 1994). For example, Williams and Braker (1999) trained participants on a causal learning task in which they had to predict whether or not a "widget-pressing" machine was operating properly based on the illumination of indicator lamps. One group was given training designed to bias them toward elemental processing. They were initially exposed to C+, D+, and E− trials before compound training with CD+ and DE−; the outcomes associated with the individual elements and the compound were consistent. A second group was given compound training where this was not the case. They experienced the same C+, D+, and E− trials but their compound training consisted of CD− and DE+ trials. For a control group, the compound trials (FG+, HI−) were unrelated to the elemental trials. All groups were then trained on a discrimination between XY+ and YZ−, before test trials with the individual Cues X, Y, and Z. For the elemental group, outcome predictions were highest for X and lowest for Z, consistent with the predictions of the Rescorla–Wagner model. For both the configural and control groups, outcome predictions were the same for all three stimuli, suggesting that there was very little generalization between the compounds and individual cues.

In Experiment 3, we attempted to bias participants toward processing cue compounds in an elemental manner to facilitate generalization of learning between phases of training. If this manipulation was successful, it should have encouraged greater protection from extinction of Cue B relative to Cue D. To do this, we provided participants with a demonstration of how the causal value of different cues might combine to bias them toward elemental processing of cue compounds. The design of the experiment is shown in Table 2. During Phase 1, two cues (L and M) were individually paired with a minor allergic reaction, but their compound (LM) was paired with a serious reaction (L+, M+, LM++). To test whether participants learned about the additive nature of outcomes, they also received training with Cues N and O in Phase 2. N was followed by a serious reaction when presented alone (N++), but by no reaction in compound with Cue O (NO−). This training should establish Cue O as a conditioned inhibitor: it signaled the absence of an otherwise expected reaction. Hence, comparison of ratings for Cue O and other cues (F, G, and H) that had been paired with no outcome served as a manipulation check by allowing us to assess whether participants

Table 2
The Design of Experiment 3

Phase 1	Phase 2	Phase 3	Choice
A++	A−	AB−	A versus C
B++			B versus D
C++		CD−	
D++			
E−	E++		
F−	F−		
G−	G−		
H−	H−		
I−	I+		
	K+		
L+	L+		
M+			
LM++		LM++	
	N++		
	NO−		
	P+	P+	

Note. Treatment of Cues A–D was the same as in Experiment 1. Cues L and M in bold were included to demonstrate to participants that the causal value of cues summed when they were presented in compound.

summed the causal values of cues presented in compound with each other. As in Experiment 1, the critical test for differential protection from extinction was between Cue B and Cue D.

Method

Participants

Fifty-four undergraduate psychology students at the University of Hull participated in Experiment 3 in exchange for course credit. Their mean age was 20.4 years (range = 18–35, $SD = 3.9$) and 39 were female.

Design

The design of Experiment 3 is shown in Table 2. Cues A–D were treated in the same manner as in Experiment 1, as were filler Cues E, F, I, and K. Cues G and H were paired with no reaction (−) in Phases 1 and 2 but were not presented in Phase 3 and Cue J was not presented at all in Experiment 3. Cues L and M were paired with a minor reaction when presented alone, but a severe reaction when presented in compound during Phase 1. L+ and LM++ trials were given in Phases 2 and 3, respectively, to provide consistency in reactions to some cues across phases of the experiment. During Phase 2, Cue N was followed by a large reaction when presented alone, but no reaction in compound with Cue O. Cue P was paired with a minor reaction during Phases 2 and 3 and ensured that there were equal numbers of trials on which a reaction or no reaction occurred during Phase 3.

Stimuli and Materials

All stimuli and materials were the same as for Experiments 1 and 2.

Procedure

The procedure was the same as for Experiments 1 and 2, with the exception of some training contingencies, as shown in the Table 2. There were eight blocks of trials in each phase of the experiment, and each trial type was presented once during each block of trials. This meant that there was a total of 216 trials.

Data Analysis

Data were analyzed in a similar manner as for Experiment 1, and the same exclusion criterion was applied. There was a single test for aggregation (K vs. I) and additional tests for participants' understanding of cue interaction. In a test for conditioned inhibition, ratings for Cue O were compared with the mean ratings for Cues F, G, and H. The latter three cues had all been paired with no outcome during Phases 1 and 2. Cue O, however, had been paired with no outcome in the presence of N which predicted a severe reaction by itself. Hence, we expected that the ratings for Cue O would be lower than the average ratings for the other three cues. We also checked that participants learned the outcomes associated with Cues L and M and Compound LM during Phase 1. For each participant, we first determined which of Cues L and M they predicted the greatest allergic reaction for on the final block of trials in Phase 1 and compared that to their prediction for Compound LM using a paired t test.

Results

Exclusion Criterion

All 54 participants met the exclusion criterion.

Training

Participants learned that Compound LM was associated with a more severe allergic reaction than either of its constituent cues. On the final block of Phase 1, participants' mean prediction for Compound LM ($M = 4.79$, $SD = 0.92$) was significantly higher than for the greater of Cues L or M ($M = 3.64$, $SD = 0.78$), $t(53) = 6.14$, $p < .001$, $M_{\text{difference}} = 1.15$, 95% CI [0.78, 1.53], $\text{BF}_{10} > 10^5$.

Outcome predictions and confidence ratings for Cues A–D followed a very similar pattern as in Experiment 1 and are shown in the top two rows of Figure 4. At the beginning of Phase 3, ratings for AB were lower than for CD, $t(53) = 4.28$, $p < .001$, $M_{\text{difference}} = 0.97$, 95% CI [0.51, 1.42], $\text{BF}_{10} = 280$. There was no difference in the confidence ratings for the two compounds, $t(53) = 1.04$, $p = .303$, $M_{\text{difference}} = 0.17$, $\text{BF}_{10} = 0.25$.

Forced Choice Test

The results of the forced choice test were also very similar to those from Experiment 1. Of the 54 participants, 36 (67%) selected A as being safer than C ($p = .020$) and 25 (46%) thought that D was safer than B ($p = .683$). Of the 36 that selected A as safer than C, 14 (39%) also thought that D was safer than B ($p = .243$).

Ratings Test

The results of the rating test are summarized in the bottom row of Figure 4. Extinction training was effective; mean ratings for Cues A–D were lower than for Cue E, $t(53) = 5.62$, $p < .001$, $M_{\text{difference}} = 1.49$, 95% CI [0.96, 2.03], $\text{BF}_{10} > 10^4$. Two phases of extinction resulted in lower ratings than did one. Ratings for A were lower than the mean for Cues B–D, $t(53) = 4.23$, $p < .001$, $M_{\text{difference}} = 0.99$, 95% CI [0.52, 1.46], $\text{BF}_{10} = 241$. Extinction in compound with an already extinguished cue did not, however, protect B from extinction relative to Cue D. Ratings for B did not differ from ratings for Cue D, $t(53) = 0.36$, $p = .717$, $M_{\text{difference}} = 0.11$, $\text{BF}_{10} = 0.16$, or from the mean for Cues C and D, $t(53) = 0.99$, $p = .329$, $M_{\text{difference}} = 0.25$, $\text{BF}_{10} = 0.23$.

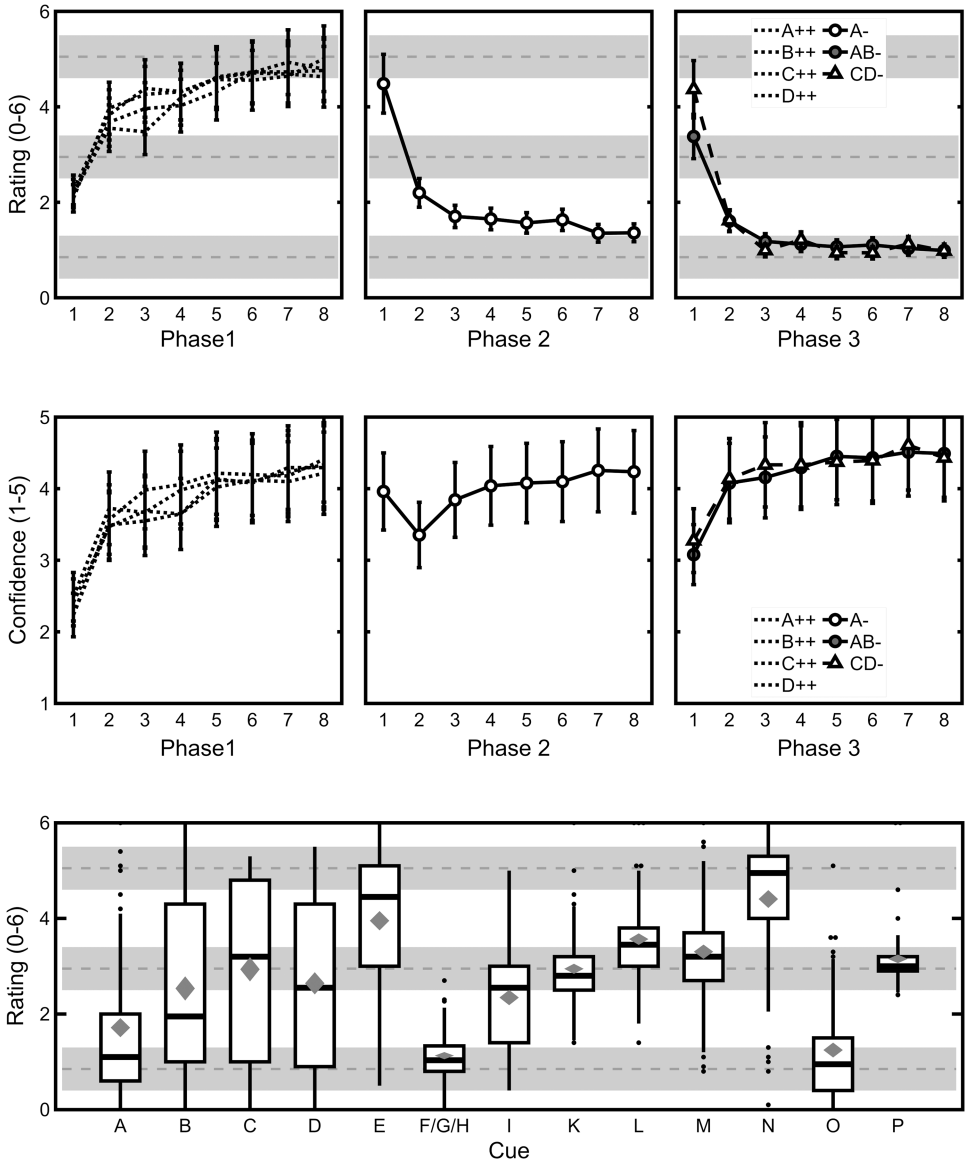
Once again, there was evidence that participants' ratings were based on their aggregated experience of the cues over different phases of the experiment. Mean ratings for I were lower than for K, $t(53) = 3.53$, $p = .001$, $M_{\text{difference}} = 0.60$, 95% CI [0.26, 0.94], $\text{BF}_{10} = 31.8$. It was not clear that participants fully understood the intended summative nature of outcomes. Although they did learn the L+ M+ LM++ discrimination during Phase 1, in the rating test there was no evidence that Cue O acquired inhibitory properties. Ratings for O did not differ from those for Cues F, G, and H which had been paired with no reaction, $t(53) = 0.70$, $p = .486$, $M_{\text{difference}} = 0.12$, $\text{BF}_{10} = 0.19$.

Discussion

Experiment 3 provided another replication of the failure to observe differential protection from extinction of two cues, B and D, extinguished in compound with partner cues with different causal strengths.

Figure 4

Results From Experiment 3



Note. The top panel shows participants' mean outcome predictions across the three phases of training for trials on which Cues A–D and their compounds (AB and CD) were presented. Participants' mean confidence ratings across the three phases are shown for the same stimuli in the middle panel. Test ratings are presented in the bottom panel for each cue.

In this experiment we attempted to bias participants toward elemental processing of cue compounds, but there was no strong evidence that this attempt was successful. Specifically, we detected no evidence that Cue O had gained conditioned inhibitory properties. It is possible, however, that this failure was a consequence of the rating scale that we used. We asked participants to predict Mrs. X's antibody level on a scale from 0.0 to 6.0, and the outcome level on "no reaction" trials was toward the bottom end of this scale (range = 0.4–1.3, M = 0.85). At test, ratings for Cue O and its control Cues F, G, and H were very low. Hence, it may have been the case that our test was not sensitive to small differences close to the bottom of the scale. That is, the test was subject to a floor effect. Alternatively, the scale might have affected participants' learning about Cue O. They had no experience of antibody levels below the range of the "no reaction" outcome and may have believed that they could not go any lower. Alternatively, participants may have understood that Cue O prevented the increase in antibody level that Cue N would otherwise have caused, but not then concluded than Cue O could result in a decrease in antibody level below a normal baseline. Indeed, there is evidence that the range of outcomes to which participants are exposed can affect learning in both Pavlovian conditioning (Mitchell & Lovibond, 2002) and causal learning tasks (Beckers et al., 2005; Lovibond et al., 2003).

Experiment 4

In Experiment 4, we attempted to address these issues by again encouraging elemental processing through training on a problem with an elemental solution, but also extending the bottom of the rating scale to avoid potential floor effects in participants' responses. Hence, a new rating scale was used that ran from 0.0 to 8.0. The ranges of antibody levels that corresponded to the three outcomes used in the previous experiments were increased ("−" = 2.5–3.4, "+" = 4.6–5.5, and "++" = 6.7–7.6), and a new negative ("−−") outcome was added to the bottom end of the scale (0.4–1.3). On screen, the bottom half of the scale (0.0–4.0) was marked as "no reaction." Participants were told that Mrs. X generally had a safe, baseline level of antibodies in her bloodstream and that some foods could cause a reduction in her antibody level below this baseline, although this was perfectly safe. At the beginning of each trial both the prediction and outcome sliders were set to some level within this baseline range (corresponding to the "−" outcome: 2.5–3.4). During each phase of the experiment, some cues were followed by the new "−−" outcome to make it clear to participants that foods could have a depressive effect on antibody levels. Finally, contingencies were arranged so that participants could observe that when a cue associated with the reduced antibody level (R−−) was presented in compound with a second cue that increased antibody level (Q+), their outcomes were additive (QR−). As in Experiments 1 and 3, the critical test for differential protection from extinction was a comparison between Cues B and D.

Method

Participants

Fifty-five undergraduate psychology students at the University of Hull participated in Experiment 4 in exchange for course credit. Their mean age was 20.4 years (range = 18–38, $SD = 3.9$) and 42 were female.

Design

The design of Experiment 4 is shown in Table 3. Cues A–D were treated in the same way as in Experiments 1 and 3, as were Cues G, I, K, L, N, O, and P. The demonstration of summation provided by Cues L and M in Experiment 3 was replaced by Cues Q and R. Q was paired with a minor outcome and R with a reduction in antibody level, when presented individually. In compound they were paired with no reaction (Q+, R−−, QR−). The intention of these trial types was to show participants that if one cue increased antibody level and another decrease it, they might cancel out the effects of each other. If participants learned that cues could interact in this way, we might have been more likely to observe deeper extinction to Cue A and greater protection from extinction of Cue B following their compound extinction in Phase 3. Cue T was included as an example of a cue that consistently lowered antibody levels throughout the experiment. S was followed by no outcome in Phase 2 and made sure that the number of cues followed by a reaction or no reaction in Phase 2 was the same as in Experiment 3.

Stimuli and Materials

All stimuli and materials were the same as for Experiments 1–3 with the following exceptions. The prediction and feedback sliders used during the training task were marked with the values 0.0–8.0 along their sides in intervals of 1.0, but they still moved in increments of 0.1. In the feedback panel, the green rectangle marked "None" extended from 0.0 to 4.0 on this scale. The orange and red rectangles marked "Minor" and "Severe" were moved up to be next to the values 4.1–6.0 and 6.1–8.0, respectively. During the rating test, the scale of the rating slider was also changed to 0.0–8.0.

Table 3
The Designs of Experiments 4 and 5

Study	Phase 1	Phase 2	Phase 3	Choice
Experiment 4 and Experiment 5	A++	A−	AB−	A versus C
	B++			B versus D
	C++		CD−	
	D++			
	E−	E++	E++	
	G−	G−		
		K+		
		N++		
		NO−		
		P+	P+	
	Q+			
	R−−			
	QR−			
		S−		
	T−−	T−−	T−−	
Experiment 4	I−	I+		
	L+	L+		
Experiment 5	WX+	WX+		
	YZ−	YZ−		

Note. The treatment of Cues A–D was the same as in Experiments 1 and 3. On some trials (−−) the level of antibody in the patient's blood was reduced following the consumption of the food(s). Cues Q and R in bold were included to demonstrate to participants that the causal value of cues summated when they were presented in compound. In Experiment 5, a compound rating test was completed following the choice test.

Procedure

The procedure was similar to that for Experiments 1–3, with the exception of some training contingencies, as shown in the Table 3. The following text was added to the end of instructions given to the participants:

> Mrs. X tends to have a low level of antibodies in her blood in between meals. You might find that some foods actually reduce antibody levels below this baseline whereas others might increase or have no marked effect on antibody levels. This reduction in antibody level is perfectly safe; Mrs. X only suffers from an allergic reaction when her antibody level rises beyond a certain level.

Four outcomes were used on training trials: no reaction with a reduction in antibody level (−−), no reaction with no change in antibody level (−), a minor reaction (+), and a severe reaction (++). The antibody levels associated with each of these outcomes was jittered around a central value in the following ranges: "−−" = 0.4–1.3, "−" = 2.5–3.4, "+" = 4.6–5.5, and "++" = 6.7–7.6.

Data Analysis

Cues O and R each signaled a reduction in antibody level below baseline, either explicitly (R−−), or as a result of compound training (N+, NO−). Each of these cues was also presented during only one stage of training. To determine whether participants understood that foods could reduce Mrs. X's antibody levels, mean ratings for Cues O and R were compared to ratings for Cue S which signaled no change in antibody level, also during a single phase of training. Ratings for O and R were also compared. To check that participants learned the outcomes associated with Cues Q and R and Compound QR during Phase 1, we compared the predictions for these three trial types on the final block of Phase 1 using an ANOVA and paired *t* tests.

Results

Exclusion Criterion

Participants were excluded if their mean outcome predictions for Cues A–D over the last four blocks of Phase 1 failed to exceed the mean value of the "minor" outcome ("+" = 5.05). Application of this criterion lead to the exclusion of one participant. Computer failure led to the loss of choice and rating test data from a second participant. Analysis of data from the remaining 53 participants is reported below.

Training

Participants learned the Q+ R−− QR− discrimination during Phase 1. Outcome predictions for Cues Q ($M = 4.68$, $SD = 1.21$) and R ($M = 1.48$, $SD = 0.82$) and their Compound QR ($M = 2.87$, $SD = 1.04$) differed significantly from each other on the final block of trials in Phase 1, $F(2, 104) = 110.04$, mean square error (MSE) = 1.24, $p < .001$, $\eta_p^2 = .679$, 90% CI [0.590, 0.733]. Paired *t* tests found that predictions for QR were significantly lower than for Q, $t(52) = 7.38$, $p < .001$, $M_{\text{difference}} = 1.81$, 95% CI [1.32, 2.30], $\text{BF}_{10} > 10^6$, and were significantly higher than those for R, $t(52) = 8.93$, $p < .001$, $M_{\text{difference}} = 1.38$, 95% CI [1.07, 1.70], Šidák corrected $\alpha = .025$, $\text{BF}_{10} > 10^9$.

Outcome predictions and confidence ratings for Cues A–D over the three phases of Experiment 4 are shown in the top two rows of Figure 5. As expected, and consistent with Experiments 1 and 3, outcome predictions for CD were higher on the first block of Phase 3 than predictions for AB, $t(52) = 4.62$, $p < .001$, $M_{\text{difference}} = 1.24$, 95% CI [0.70, 1.78], $\text{BF}_{10} = 790$. There was no difference in participants' confidence in their ratings for these cue compounds, $t(52) = 0.32$, $p = .749$, $M_{\text{difference}} = 0.04$, $\text{BF}_{10} = 0.16$.

Forced Choice Test

Thirty-nine (74%) of the 53 participants believed that Cue A was safer than Cue C ($p = .001$). Thirty-six participants (68%) chose D as safer than B ($p = .013$). Out of the 39 participants who chose A, 25 (64%) also chose D ($p = .108$)

Rating Test

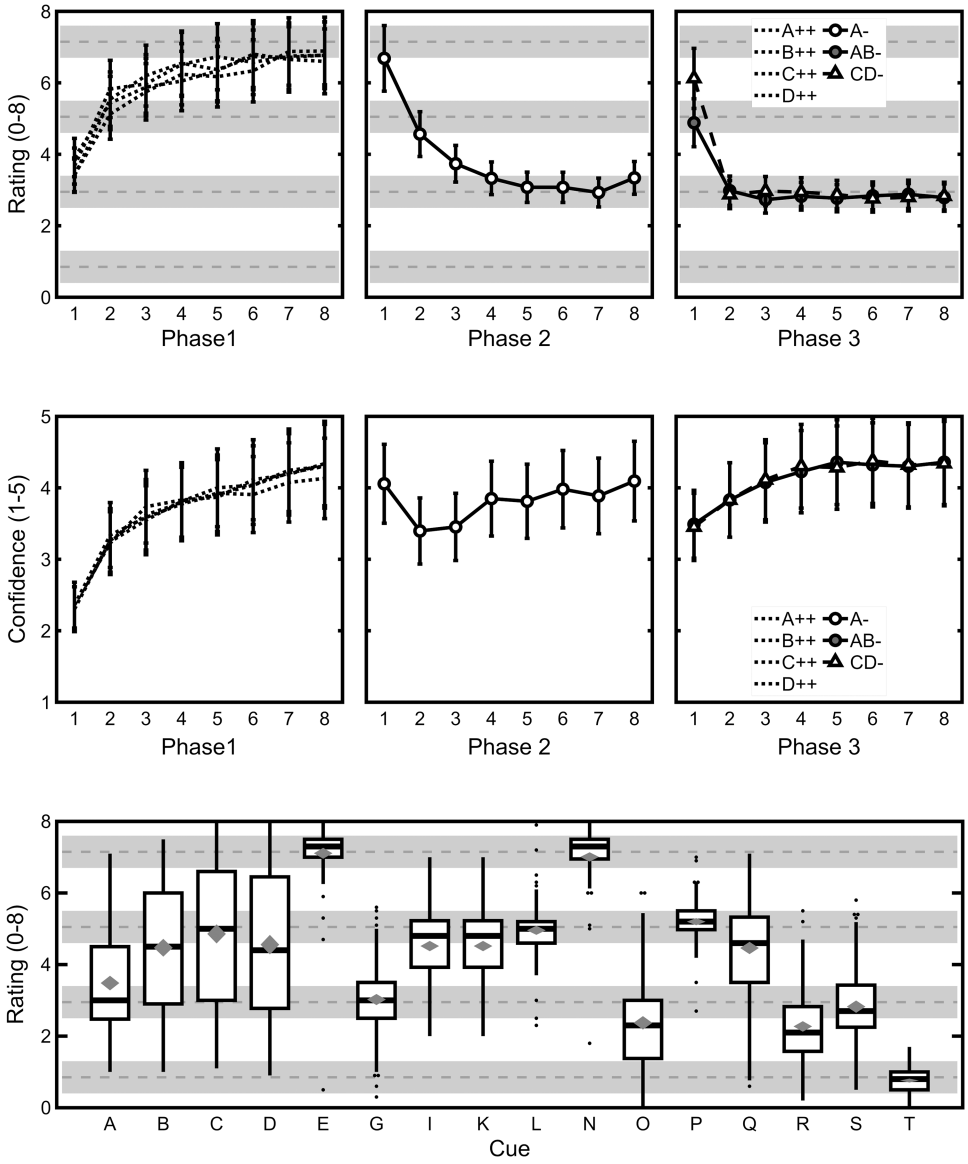
Consistent with the previous experiments, the rating test showed that extinction was effective. Ratings for Cues A–D, all of which underwent extinction, were lower than for Cue E which was paired with a serious outcome ("++") in Phases 2 and 3, $t(52) = 11.87$, $p < .001$, $M_{\text{difference}} = 2.77$, 95% CI [2.30, 3.24], $\text{BF}_{10} > 10^{13}$. Furthermore, two phases of extinction were more effective than a single phase; mean ratings for Cues B–D were higher than ratings for A, $t(52) = 4.07$, $p < .001$, $M_{\text{difference}} = 1.14$, 95% CI [0.58, 1.71], $\text{BF}_{10} = 148$. Once again, we observed that a cue extinguished in compound with an already extinguished cue was not protected from extinction relative to a cue extinguished in compound with an excitor. Ratings for B did not differ from ratings for Cue D, $t(52) < 1$, $M_{\text{difference}} = 0.09$, $\text{BF}_{10} = 0.16$, or from the mean ratings for Cues C and D, $t(52) = 1.01$, $p = .319$, $M_{\text{difference}} = 0.24$, $\text{BF}_{10} = 0.24$.

This time, Cue O appeared to have acquired conditioned inhibitory properties. Mean ratings for Cues O and R, each of which signaled a reduction in antibody, were lower than mean ratings for Cue S which was consistently paired with no change in antibody level, $t(52) = 2.47$, $p = .017$, $M_{\text{difference}} = 0.50$, 95% CI [0.09, 0.91], $\text{BF}_{10} = 2.36$. Ratings for Cues O and R were not significantly different from each other, $t(52) < 1$, $M_{\text{difference}} = 0.10$, $\text{BF}_{10} = 0.18$. Finally, there was only marginal evidence of aggregation of experience across phases of training. Ratings for Cue I were lower than those for Cue K, but this difference just failed to reach significance, $t(52) = 1.98$, $p = .053$, $M_{\text{difference}} = 0.33$, $\text{BF}_{10} = 0.91$.

Discussion

In Experiment 4, participants appeared to understand that when cues were presented in compound, their outcomes were additive. That is, they seem to have processed cue compounds in an elemental manner. Following training in which Cue N was paired with a severe allergic reaction when it was presented by itself but not when presented in Compound NO, Cue O acquired conditioned inhibitory properties. There was, however, little evidence for differential protection from extinction during compound stimulus presentation. Test Ratings for Cue B, which was extinguished in compound with an already extinguished cue were the same as for a cue extinguished in compound with another excitor (D). Participants were more likely to choose D over B as the safer food in the forced choice

Figure 5
Results From Experiment 4



Note. The top panel shows participants' mean outcome predictions across the three phases of training for trials on which the Cues A–D and their compounds (AB and CD) were presented. Participants' mean confidence ratings across the three phases are shown for the same stimuli in the middle panel. Test ratings are presented in the bottom panel for each cue.

test, but this effect went away when we considered only those participants who also believed that A was safer than C.

Experiment 5

There is evidence from pigeons (Pearce et al., 2002) that generalization of learning between individual cues and their compounds can be affected by the similarity between the cues and the background they are shown against when presented alone. That is, when Cue B underwent compound extinction alongside Cue A and was then tested alone, generalization of learning about Cue B may have been incomplete because Cue A was quite dissimilar to the plain background next to which B was presented at test. One way to reduce this generalization decrement is to use a testing procedure where participants are asked to rate cue compounds. In Experiment 5, participants were given additional test trials in which each of Cues A–D was presented in compound with another, familiar, cue. Hence, although the context in which the cues appeared was different between extinction and test, generalization decrement based upon the cues being presented in isolation should have been reduced. To achieve this, training trials with two compounds, WX and YZ, were given during Phases 1 and 2. The former compound was paired

with a minor allergic reaction (WX+) and the latter with no reaction (YZ−). On compound test trials, each of Cues A–D was presented in compound with one of the elements of Compound WX (AW, BX, CW, and DX) and with one of the elements of Compound YZ (AY, BZ, CZ, and DY). The critical comparison for these test trials was between compounds containing Cue B (BX, BZ) and those containing Cue D (DX, DY).

Method

Participants

Fifty-three undergraduate psychology students at the University of Hull participated in Experiment 5 in exchange for course credit. Their mean age was 22.4 years (range = 18–53, $SD = 7.6$) and 46 were female.

Design

The design of Experiment 5 is shown in Table 3. It was similar to Experiment 4 but differed in two important ways. First, participants received training during Phases 1 and 2 with Compounds WX and YZ. WX was paired with a minor reaction (+), whereas YZ was paired with no change in antibody level (−). Second, a compound rating test was administered immediately following the choice test. In this test, cues, A–D were presented in compound with cues from each of the Compounds WX and YZ. Each of Cues A–D was presented on two compound rating trials, once with either W or X and once with either Y or Z (AW, AY, BX, BZ, CW, CZ, DX, DY). Following the compound rating test, participants underwent a second test where each cue was presented alone.

Stimuli and Materials

All stimuli and materials were the same as for Experiments 1–4.

Procedure

The procedure was the same as for Experiment 4, except for some training contingencies shown in the Table 3 and an additional, compound, rating test. The compound rating test took place immediately following the choice test. It consisted of eight trials, on each of which participants were presented with two foods and asked to rate the likelihood that Mrs. X would suffer an allergic reaction the next time she ate that meal. The two foods and their names were presented one above the other in the same way as on compound training trials. Participants received one trial with each of the Compounds AW, AY, BX, BZ, CW, CZ, DX, and DY, in random order. Following the compound rating test, each of the 18 foods used in the experiment were rated individually.

Data Analysis

Ratings for the cue compounds were analyzed using a two-way repeated measures ANOVA. The compound cues were first categorized based upon which of Cues A–D they contained, and second depending upon the outcome that the partner cue had been paired with during training (W/X = "+," Y/Z = "−"). Hence, the two factors in the ANOVA were Cue (A, B, C, D), and Outcome ("+," "−"). If participants' ratings of these compound cues were sensitive to the training history of their components, we should predict a main effect of Outcome. Differential extinction of Cues A–D should have also resulted in a main effect of Cue. An interaction between these factors might be expected if the relationship between causal value and rating was nonlinear, resulting in greater or lesser sensitivity to differences in causal rating at different points on the scale. The main effect of Cue was further investigated using a set of planned t tests which followed the same logic as those used in previous experiments. The effectiveness of one or two phases of extinction was assessed by comparing average ratings for compounds containing Cue A with average ratings for all other compounds. Rating for compounds containing Cue B were then compared with ratings for compounds containing Cue D to determine whether Cue B enjoyed greater protection from extinction relative to Cue D.

Results

Exclusion Criterion

The same exclusion criterion was used as in Experiment 4. Three participants failed to meet the criterion, and only data from the remaining 50 were further analyzed.

Training

Participants learned the Q+ R−− QR− discrimination during Phase 1. Outcome predictions for Cues Q ($M = 4.41$, $SD = 1.14$) and R ($M = 1.35$, $SD = 0.79$) and their Compound QR ($M = 3.17$, $SD = 1.08$) differed significantly from each other on the final block of trials in Phase 1, $F(2, 98) = 136.31$, $MSE = 0.872$, $p < .001$, $\eta_p^2 = .736$, 90% CI [0.657, 0.781]. Paired t tests found that predictions for QR were significantly lower than for Q, $t(49) = 6.27$, $p < .001$, $M_{difference} = 1.24$, 95% CI [0.85, 1.64], $BF_{10} > 10^5$, and were significantly higher than those for R, $t(49) = 9.96$, $p < .001$, $M_{difference} = 1.82$, 95% CI [1.45, 2.19], Šidák corrected $\alpha = .025$, $BF_{10} > 10^{10}$.

Outcome predictions and confidence ratings for each phase of training are shown in the top two rows of Figure 6. At the beginning of Phase 3, outcome predictions for CD were significantly higher than those for AB, $t(49) = 5.80$, $p < .001$, $M_{difference} = 1.69$, 95% CI [1.10, 2.28], $BF_{10} > 10^4$. For the first time, there was also a difference in the confidence that the participants had in their predictions. Confidence was higher for CD than for AB, $t(49) = 2.71$, $p = .009$, $M_{difference} = 0.52$, 95% CI [0.13, 0.91], $BF_{10} = 3.98$.

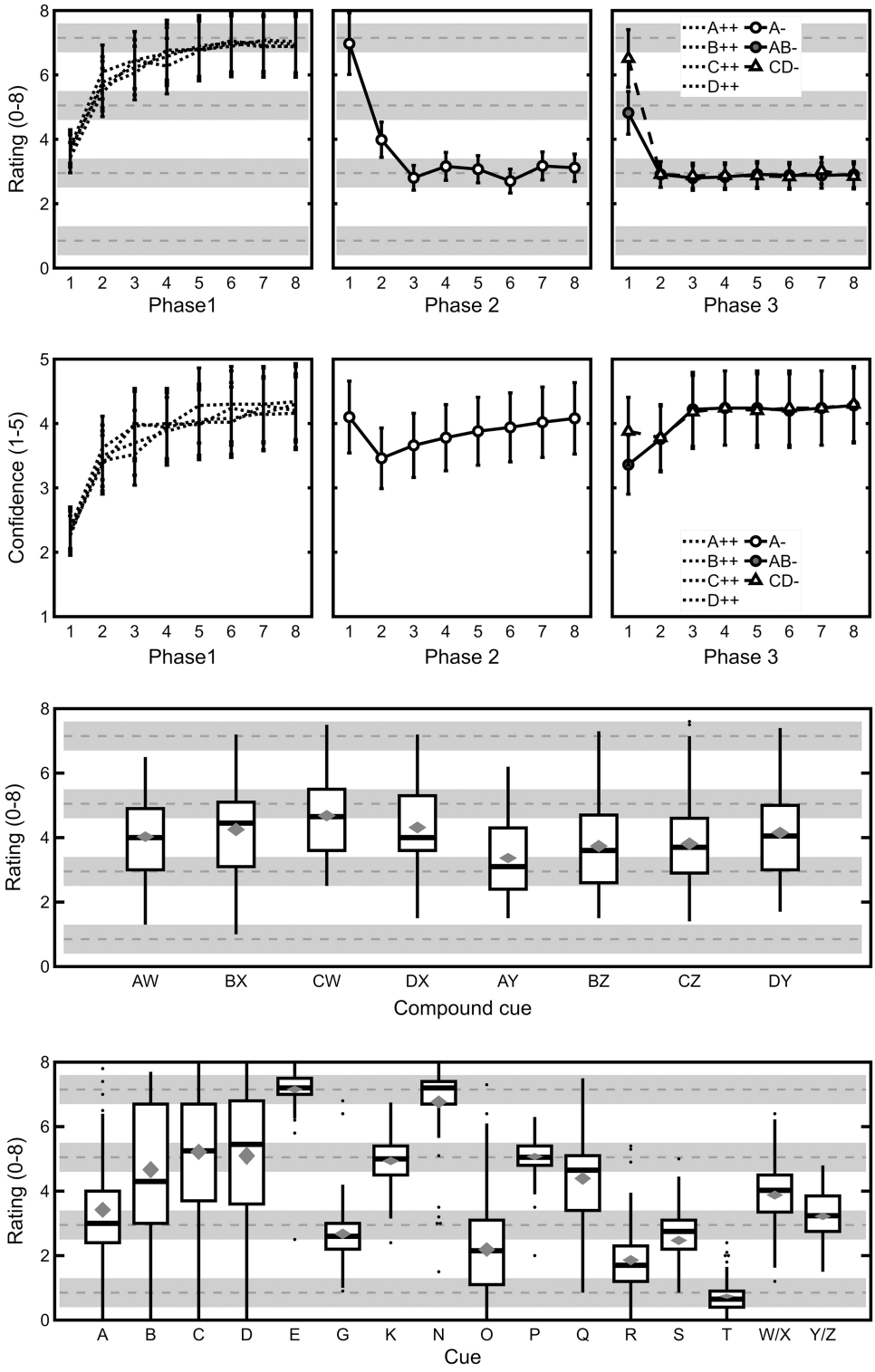
Forced Choice Test

A was chosen as being safer than C by 40 out of the 50 participants (80%, $p < .001$). Twenty-four participants (48%) thought that B was safer than D ($p = .888$). Of the 40 participants who selected A, 19 (47%) also thought that D was safer than B ($p = .875$).

Compound Cue Ratings

Ratings for the cue compounds are shown in the third row of Figure 6 and were analyzed using a two-way repeated measures ANOVA. There was a significant effect of Cue, $F(3, 147) = 3.55$, $MSE = 1.87$, $p = .016$, $\eta_p^2 = .068$, 90% CI [0.007, 0.127], and of outcome, $F(1, 49) = 14.54$, $MSE = 2.09$, $p < .001$, $\eta_p^2 = .229$, 90% CI [0.075, 0.377], but the interaction was not significant, $F(3, 147) = 2.40$, $MSE = 0.92$, $p = .070$.

Figure 6
Results From Experiment 5



Note. The top panel shows participants' mean outcome predictions across the three phases of training for trials on which Cues A–D and their compounds (AB and CD) were presented. Participants' mean confidence ratings across the three phases are shown for the same stimuli in the second panel. The third panel shows the predictions from the compound rating test trials, and the bottom panel shows the results from the individual rating tests.

Planned *t* tests were used to further analyze the main effect of cue. Ratings for compounds containing A were lower than for those containing Cues B–D, $t(49) = 2.42$, $p = .019$, $M_{difference} = 0.42$, 95% CI [0.07, 0.77], $BF_{10} = 2.15$, suggesting that two phases of extinction were more effective than one phase. There was no evidence that Cue B enjoyed greater protection from extinction than Cue D. Ratings for compounds containing B did not differ from those containing Cue D, $t(49) = 1.32$, $p = .193$, $M_{difference} = 0.24$, $BF_{10} = 0.35$.

Individual Cue Ratings

The bottom row of Figure 6 shows the results from the rating test with individual cues. Extinction training was effective. Ratings for Cue E, which was paired with a serious reaction in Phases 2 and 3, were higher than the mean ratings for Cues A–D which each underwent extinction, $t(49) = 10.80$, $p < .001$, $M_{difference} = 2.55$, 95% CI [2.08, 3.03], $BF_{10} < 10^{11}$. Two phases of extinction were more effective than one; ratings for Cue A were lower than the mean ratings for Cues B–D, $t(49) = 5.32$, $p < .001$, $M_{difference} = 1.57$, 95% CI [0.98, 2.17], $BF_{10} < 10^4$. There was no evidence, however, that Cue B was protected from extinction relative to Cue D. Ratings for B did not differ significantly from ratings for Cue D alone, $t(49) = 1.22$, $p = .229$, $M_{difference} = 0.43$, $BF_{10} = 0.31$, or from the mean ratings for Cues C and D, $t(49) = 1.75$, $p = .087$, $M_{difference} = 0.49$, $BF_{10} = 0.63$.

There was also evidence that participants learned that Cues O and R each caused a reduction in antibody level. Mean ratings for O and R were lower than for Cue S which had signaled no change in antibodies, $t(49) = 2.17$, $p = .035$, $M_{difference} = 0.45$, 95% CI [0.03, 0.86], $BF_{10} = 1.32$. There was no difference in the ratings for Cue R which had been explicitly paired with the "−−" outcome, and Cue O which had been trained as a conditioned inhibitor, $t(49) = 1.21$, $p = .233$, $M_{difference} = 0.33$, $BF_{10} = 0.30$.

Discussion

Results from the rating test with individual cues were very similar to those from Experiment 4. The combination of elemental training in Phase 1, and the use of an extended rating scale, did appear to inform participants that when cues were presented in compound their causal values were additive. Following N+ NO− training in Phase 2, O appeared to acquire inhibitory properties. There was, however, no evidence of differential protection from extinction for Cues B and D. Testing Cues A–D in compound with a neutral (Y or Z) or mildly excitatory (W or X) cue did not alter this result. Ratings for the compound cues were affected by the associative status of the added cue, but ratings for compounds containing B did not differ from those containing the control Cue D.

Meta Analysis

In Experiments 1, 3, 4, and 5, Cues A–D were treated in the same manner as they were in Experiment 1 of Griffiths et al. (2017). None of those five experiments revealed significant evidence that extinction of Cue A during the second phase of training conferred greater protection from extinction to Cue B during compound extinction in Phase 3 relative to control Cues C and D. We calculated a meta-analytic Bayes factor using the BayesFactor package in R (Morey & Rouder, 2022), combining data from all 279 participants across these five experiments. Evidence for the point null was compared with the one-sided prediction that ratings for Cue B were higher than the average of those for Cues C and D, using the standard Cauchy prior. The resulting Bayes factor indicated strong evidence in favor of the null hypothesis, $BF_{10} = 0.015$. When the analysis was restricted to the four experiments reported here and ratings for Cue B were compared with those for control Cue D, $BF_{10} = 0.027$, again indicating strong evidence for the null.

In fact, in each of these five experiments, ratings for Cue B were numerically lower than those for Cues C and D—the opposite of the predicted difference. In each case, the size of this differences was, however, very small. Griffiths et al. (2017) reported that a power analysis indicated that a sample size of 281 participants would have been required to detect an effect of the size that they reported (Cohen's $f = 0.11$) with Power $(1 - \beta) = .80$. Similar analyses using G*Power 3.1 (Faul et al., 2009) indicated that very large sample sizes would have been required to detect effects of the sizes we reported in Experiments 1 (Cohen's $d = 0.06$, $n = 2{,}291$), 3 ($d = 0.13$, $n = 439$), 4 ($d = 0.14$, $n = 414$), and 5 ($d = 0.25$, $n = 131$). To determine whether there was any systematic difference in rating across experiments, we performed a meta-analysis using the meta package (Balduzzi et al., 2019) in R. The mean and standard deviation of the difference between ratings for Cue B and the average of Cues C and D for Experiment 1 of Griffiths et al. (2017) were estimated from the reported 95% confidence intervals of the mean. Our Experiments 4 and 5 used a longer rating scale (0–8) than Experiments 1 and 3 (0–6), but differences in ratings were not transformed or normalized in any way for the meta-analysis. This was because the differences between antibody levels assigned to the outcomes associated with Cues A–D (++ and −) were the same across all experiments; the longer scale in the latter experiments was simply due to the inclusion of an additional, negative, outcome (−−) at the bottom of the scale.

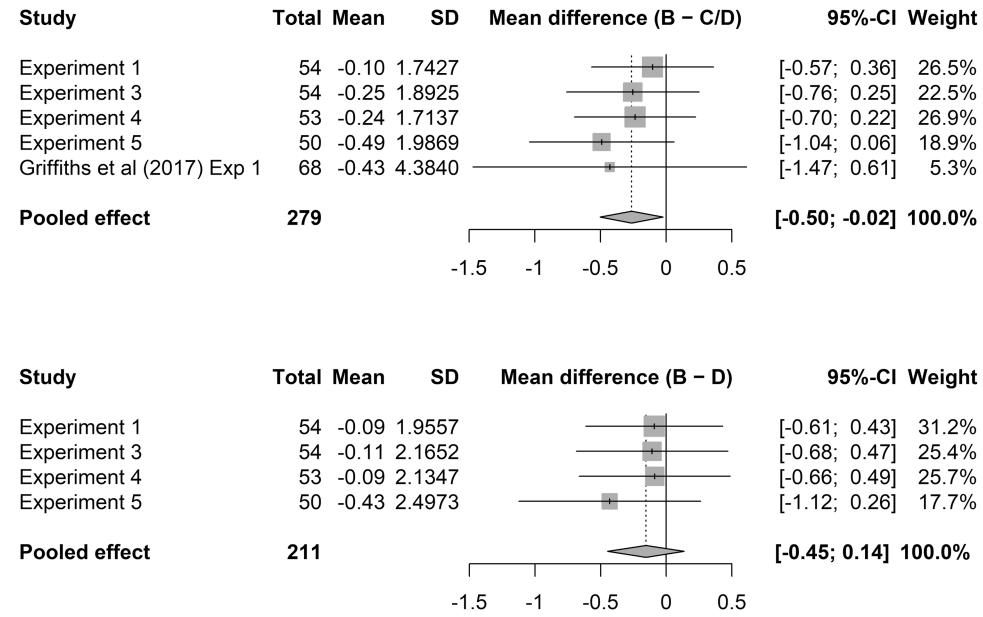
The results of the meta-analysis are shown in the top panel of Figure 7. Tests of heterogeneity revealed no significant differences in observed effect sizes across studies, $Q(4) = 1.22$, $p = .875$. The mean of the pooled difference scores was −0.26, 95% CI [−0.50, −0.02], indicating a small overall effect size (Cohen's $d = 0.13$), with a significant z of 2.16, $p = .031$. These results suggest that rather than Cue B being protected from extinction relative to Cues C and D, participants' ratings for B were consistently slightly lower than for Cues C and D. A slightly different picture emerged when the meta-analysis was restricted to a comparison of Cue B with the more appropriate control Cue D in the four experiments reported here (a comparison which was not reported by Griffiths et al., 2017), as can be seen in the bottom panel of Figure 7. Again, tests of heterogeneity revealed no difference in the observed effect size across studies, $Q(3) = 0.74$, $p = .862$. The mean of the pooled differences scores was smaller at 0.15, 95% CI [−0.44, 0.14], $d = 0.07$, and did not significantly differ from zero, $z = 1.04$, $p = .300$.

General Discussion

In five experiments participants received training in which four cues were first paired with a large outcome (A++, B++, C++, D++) and subsequently underwent compound extinction training (AB−, CD−). In each experiment, we found no evidence that the training history of a cue affected learning about the cue that it was

Figure 7

Forest Plots Showing the Mean Differences in Ratings for Cue B and the Average of Cues C and D (Top Panel) or Cue B Alone (Bottom Panel) for Each Experiment Included in the Meta-Analyses



Note. Exp = experiment; CI = confidence interval.

partnered with during extinction. Experiment 1 was a replication of Griffiths et al. (2017) Experiment 1. Between the first phase and the compound extinction phase, one of the cues underwent extinction by itself (A−). According to the Rescorla–Wagner model, the resulting reduction in the causal value (or associative strength) of Cue A should have decreased the prediction error for Compound AB at the start of the compound extinction phase, relative to that for Compound CD. As a consequence, B should have experienced a lesser reduction in causal value than Cue D. Consistent with the results of Griffith et al., we found no evidence for this in either a forced choice test, or a rating test. In Experiment 2, we investigated the effect of manipulating prediction error on extinction of an already extinguished cue. In the second phase of the experiment, three of the critical cues were extinguished alone, whereas the fourth continued to be paired with a large outcome (A−, B++, C−, D−). The Rescorla–Wanger model predicts that, at the start of compound extinction, the prediction error for Compound AB should have been greater than that for Compound CD. Hence, Cue A should have lost more causal value than Cue C. Forced choice and rating tests yielded no such effect.

Experiments 3–5 used the same basic design as Experiment 1 and tested different explanations for the lack of differential protection from extinction. In Experiment 3, participants were given training to illustrate that the outcome associated with a cue compound was the sum of the outcomes associated with the individual cues (L+, M+, LM++). This was intended to encourage elemental rather than configural processing of cue compounds and promote generalization of learning from the compound extinction training to the ratings of individual cues at test. We cannot be certain that this training did affect the way that participants treated Cues A–D, but there is evidence (Williams & Braker, 1999) that similar training with one set of cues can bias people toward processing compounds of other cues elementally. We extended the range of possible outcome values in Experiment 4, to make it clear that the causal value of cues could be reduced below a baseline level. This was further intended to encourage participants to process compounds elementally and to add together the causal value of individual cues when they were presented in compound. Finally, in Experiment 5 we added a compound rating test in which each of the critical cues was presented in combination with neutral or mildly excitatory cues. The purpose of this test was to reduce generalization decrement between compound extinction and test phases by presenting the critical cues in similar contexts in the two phases. Experiments 3–5 produced similar results to Experiment 1. There was no evidence that Cues B and D experienced differential protection from extinction by virtue of undergoing compound extinction in partnership with cues with different causal strengths. Bayes factor analyses found positive or strong evidence in favor of the null hypothesis in each experiment. This was true also for the compound rating test in Experiment 5. A significant proportion of participants in Experiment 4 chose D as being safer than B, consistent with differential protection from extinction. This effect went away, however, when we considered only those participants who also believed that C was safer than A and was not observed in either Experiment 3 or 5. Together, the 5 experiments reported here consistently failed to find evidence that the causal status of one cue affects extinction of a second cue when the two are presented in compound. These results are consistent with those from similar tasks reported elsewhere (Griffiths & Westbrook, 2012; Griffiths et al., 2017; Holmes et al., 2014).

In a causal learning task, Collins and Shanks (2002) found that participants sometimes based their causal judgements on the current associative, or causal, value of a cue (momentary strategy) and sometimes on their accumulated experience with that cue across training (integrative strategy). Which strategy they employed depended

upon how frequently they were asked to make these judgements. When judgements were made every 10 trials throughout training, participants adopted a momentary strategy, whereas when they were asked for a single judgement at the end of training, they used an integrative strategy. Griffiths and Westbrook (2012) suggested that one explanation for their failure to detect protection from extinction might be that participants based their test ratings of cues on their aggregated experience of those cues over different phases of training (see also Griffiths et al., 2017). In their experiments and ours participants were asked to make predictions about the effects of cues on the level of antibodies in Mrs. X's blood stream, but they were only explicitly asked how likely they believed that a cue would cause Mrs. X to experience an allergic reaction in the final rating test, consistent with the conditions that Collins and Shanks (2002) suggested promote an integrative strategy. Following the example of Griffiths et al. (2017), we stressed to participants that their ratings should be based on what they believed about the food "right now," but we cannot be sure that these instructions were effective. In Experiments 1 and 3–5, ratings for Cue B might not have differed from those for Cues C and D simply because participants experienced pairings of those cues with the same outcomes across phases of the experiment (i.e., outcome ++ in Phase 1, and outcome − in Phase 3). Aggregation of these experiences also explains why ratings for these cues were much higher at test than at the end of the final phase of training. When we explicitly tested for aggregation, however, we did not always find evidence consistent evidence for it, whereas we consistently observed the same pattern of responding to the critical Cues A–D. It is also the case that Griffiths and Westbrook (2012) found evidence for blocking of causal learning, and Holmes et al. (2014) found that extinguishing a cue in compound with a conditioned inhibitor did confer protection from extinction. Neither of these effects is consistent with an explanation of our results purely in terms of aggregation of experience.

A second explanation that Griffiths and Westbrook (2012) considered for their results was a state-based model of learning proposed by Redish et al. (2007) which employs a temporal difference reinforcement learning algorithm related to the error-correction rule of the Rescorla–Wagner model. Animals, or humans, learn about the value of taking an action in the state they are currently in based on reinforcement. A second component of the model categorizes observed cues to recognize the current state and to create new states when the situation changes sufficiently. Importantly, chronic negative prediction error (such as that experienced during extinction) promotes the formation of new states, whereas positive prediction error tends not to. Hence, one might expect the negative prediction error and change in cue configuration experienced during Phase 3 of our experiments to result in the creation of a new state. If the new state is initialized with a low value, no further extinction learning will occur since the prediction error will be low. Since limited learning about the critical cues occurs in Phase 3, there should be little difference in ratings for Cues B, C, and D. One might, however, expect that rating to these cues will be lower than for Cue E. This is because positive prediction error for Cue E in Phase 2 of our experiments should not result in the creation of a new state, but simply updating the value of the existing state. At test, ratings for Cue E should be based on the value of the single state that E evokes, whereas ratings for Cues B, C, and D might be based on the original and/or extinction states of those cues.

We might also explain the lack of protection from extinction in these experiments by appealing to within-compound associations. There is substantial evidence from animal conditioning (e.g., Brogden, 1939; Rescorla, 1981a; Rescorla & Durlach, 1981) and human causal learning (e.g., Dickinson & Burke, 1996; Wasserman & Berglan, 1998) that when two cues are presented in compound, associations may form between them. These associations can then allow the presentation of one of the cues to retrieve a memory of its companion. Furthermore, within-compound associations are more likely to develop when the compound is not immediately followed by an outcome (Holland, 1980; Honey & Hall, 1992; Rescorla, 1981b; Urcelay & Miller, 2009), as is the case in the compound extinction phase of our experiments. It is possible that in Experiments 1, 3, 4, and 5, Cue B enjoyed greater protection from extinction than Cue D due to the prior extinction of Cue A. As a result, and as predicted by the Rescorla–Wagner model, at the end of the third phase of each experiment, Cue B would have then had greater causal strength than Cue D, while Cue A had negative causal strength. At test, however, presentation of Cue B might have resulted in the retrieval of a memory of Cue A which influenced the rating given by participants. This could explain why ratings for Cue B were, in fact, lower than those for Cues C and D, as revealed by the meta-analysis of five similar experiments, although this difference was not significant when we considered only the comparison between Cues B and D in the four experiments reported here. A similar mechanism would explain why ratings for Cue A were not lower than those for Cue C in Experiment 2. Hence, disrupting the formation of within-compound associations might reveal the effects that we have failed to observe here. Given that within-compound associations form more readily in the absence of an outcome, an obvious way to achieve this is to pair the critical cues with an outcome during Phases 2 and 3 of the experiment. This might be achieved either by pairing the cues and compounds with the moderate (+) outcome, or by adopting an alternative procedure where participants are asked to choose between different allergic reactions that a food might cause (e.g., Quigley et al., 2023). In each case, negative prediction error between the cue and its Phase 1 outcome will result in extinction, but the presence of an outcome should reduce the potential for the formation of within-compound associations. The former approach might also allow us to assess the state-based model of Redish et al. (2007). The reduced prediction error and situational change resulting from the presentation of a diminished outcome, rather than no outcome, might discourage the creation of a new state and promote new learning during Phase 3.

Finally, we must consider the possibility that learning during extinction is not determined simply by a common error term, but by a combination of the common error and the difference between an individual cues' associative strength and the outcome. Rescorla (2000a, 2001) developed a testing procedure for assessing changes in the associative strength of cues with different training histories when they were later conditioned, or extinguished, in compound. For example, in the first stage of one experiment employing rats as subjects, Rescorla (2000a) first established Cues A and C as excitors by pairing them with food and Cues B and D as conditioned inhibitors by presenting them in compound with a third excitor in the absence of food (A+ C+ X+ BX− DX−). In the second stage of the experiment, the Compound AB was paired with food (AB+). Finally, responding to the Compounds AD and BC was assessed. Since each of these compounds comprised one cue that had originally been an excitor (A or C) and a second that had been a conditioned inhibitor (B or D) then, ignoring the effects of the Stage 2

training, responding to these two compounds should have been equal. Furthermore, according to the Rescorla–Wagner model, any change in the associative strength of Cues A and B as a result of Stage 2 training should have been equivalent, since the same common error term would apply to each. In fact, Rescorla found that responding was higher to Compound BC than to Compound AD, indicating that B underwent more associative change than A during compound conditioning. Other experiments involving compound extinction or presenting an excitor in combination with a neutral stimulus (Rescorla, 2000a, 2001) found that, in each case, the cue with the greater individual prediction error underwent greatest change in associative strength. To accommodate these results, Rescorla suggested that learning was governed by both common and individual error terms, perhaps in a multiplicative manner. This proposal has been incorporated into some models of associative learning (e.g., Brandon et al., 2003; Le Pelley, 2004).

This combination error term can explain our results to an extent. At the beginning of Phase 3 of Experiments 1 and 3–5, Cue A should have had low causal strength, whereas Cue B should have had a high causal strength. Hence, the discrepancy between Cue B's causal strength and the outcome (no reaction) was much larger than was the case for Cue A. As a result, Cue B should have undergone a much greater reduction in causal strength than Cue A. One might still expect Cue B to have experience a smaller reduction in causal strength than Cue D due to differences in the common error terms applied to each, but the magnitude of this difference depends upon exactly how common and individual error terms are combined. In Experiment 2, Cues A, C, and D should all have had low causal strength at the beginning of Phase 3, and so the individual error terms associated with each cue during compound extinction would have been very small resulting in minimal learning about any of the cues. It is possible that combining our experimental design with the compound testing procedure described by Rescorla (2000a, 2001) might reveal the relative contributions of common and individual error terms to the compound extinction of causal learning.

We mentioned in the introduction that the results of several previous studies could be partly explained in terms of Pearce's configural theory (Pearce, 1987, 1994, 2002). Pearce proposed that each pattern of stimulation, whether composed of a single cue or a compound of multiple cues, is represented as a unique configuration which may become associated with an outcome. Generalization of associative, or causal, strength between configurations is a function of the number of cues that they share and the number that are exclusive to each. Other authors (Kinder & Lachnit, 2003; Pearce et al., 2008) have further suggested that generalization might be influenced by the discriminability of individual cues, so that for some cues there will be very little generalization between elements and compounds whereas for others there will be much more. Hence, configural theory might go some way to explaining why Holmes et al. found no difference in test ratings for cues extinguished in compound with excitatory, neutral, or inhibitory partners if there was very little generalization of compound extinction to the tests with individual cues. In our experiments and previous similar ones (Griffiths & Westbrook, 2012; Griffiths et al., 2017; Holmes et al., 2014), there were, however, significant differences in ratings for Compounds AB and CD at the start of the compound extinction phase. These differences suggest that there was significant generalization between individual cues and cue compounds, undermining this as an explanation of why we failed to find evidence that Cue B was protected from extinction relative to Cues C and D.

Finally, we should briefly consider the role of the cover story given to our participants. The cover story that we used, involving tracking antibody levels and fluctuations in the patient's response due to changes in her treatment regimen, was copied from previous studies (Griffiths & Westbrook, 2012; Griffiths et al., 2017; Experiment 3), but is complicated and this might have contributed to the results. Our findings were, however, consistent with those of several other studies that have used different versions of the allergist task, and at least one using a very different learning procedure. The first two experiments reported by Griffiths and Westbrook (2012), employed a more standard allergist task where, during training, participants simply had to learn about two binary outcomes (the presence or absence of a reaction), a task also used by Holmes et al. (2014). These experimental all found results similar to ours—following compound extinction, cues' ratings were not affected by the causal status of the cue that they had been extinguished with. Similarly, Lovibond et al. (2000) found no difference in expectancy ratings or skin conductance levels to stimuli extinguished in compound with either a second excitatory stimulus or a conditioned inhibitor in a fear conditioning experiment. Although we cannot be certain that our cover story did not affect our results, this consistency across studies is encouraging.

In summary, we have replicated previous failures to find evidence for differential protection from extinction in human causal learning (Griffiths & Westbrook, 2012; Griffiths et al., 2017; Holmes et al., 2014). This absence of an effect survived attempts to bias participants toward elemental processing of cue compounds, manipulation of the task to encourage the development of negative causal strength (i.e., conditioned inhibition), and a compound testing procedure designed to reduce generalization decrement between the compound extinction and test phases. These experiments taken together with the existing literature provide compelling evidence that, in human causal learning, learning about a cue extinguished in compound is not affected by the causal status of its partner cue in the manner predicted by the Rescorla–Wagner model. Or, alternatively, a variety of different methods of collecting causal ratings after training may not accurately measure such learning. There are several potential explanations for these results. Our data do not allow us to choose between them, but each generates testable predictions.

References

Aitken, M. R., Larkin, M. J., & Dickinson, A. (2001). Re-examination of the role of within-compound associations in the retrospective revaluation of causal judgements. *The Quarterly Journal of Experimental Psychology B*, *54*(1), 27–51. https://doi.org/10.1080/02724990042000029

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence Based Mental Health*, *22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

Beckers, T., De Houwer, J., Pineño, O., & Miller, R. R. (2005). Outcome additivity and outcome maximality influence cue competition in human causal learning. *Journal of Experimental Psychology: Learning, Memory,* and Cognition, *31*(2), 238–249. https://doi.org/10.1037/0278-7393.31.2.238

Bouton, M. E., Westbrook, R. F., Corcoran, K. A., & Maren, S. (2006). Contextual and temporal modulation of extinction: Behavioral and biological mechanisms. *Biological Psychiatry*, *60*(4), 352–360. https://doi.org/10.1016/j.biopsych.2005.12.015

Brandon, S. E., Vogel, E. H., & Wagner, A. R. (2003). Stimulus representation in SOP: I theoretical rationalization and some implications. *Behavioural Processes*, *62*(1–3), 5–25. https://doi.org/10.1016/S0376-6357(03)00016-0

Brogden, W. J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology*, *25*(4), 323–332. https://doi.org/10.1037/h0058944

Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 837–854. https://doi.org/10.1037/0278-7393.17.5.837

Coelho, C. A. O., Dunsmoor, J. E., & Phelps, E. A. (2015). Compound stimulus extinction reduces spontaneous recovery in humans. *Learning & Memory*, *22*(12), 589–593. https://doi.org/10.1101/lm.039479.115

Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgements. *Memory & Cognition*, *30*(7), 1138–1147. https://doi.org/10.3758/bf03194331

Culver, N. C., Vervliet, B., & Craske, M. G. (2015). Compound extinction: Using the Rescorla–Wagner model to maximise exposure therapy effects for anxiety disorders. *Clinical Psychological Science*, *3*(3), 335–348. https://doi.org/10.1177/2167702614542103

Denniston, J. C., Savastano, H., & Miller, R. R. (2001). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 65–117). Erlbaum.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *The Quarterly Journal of Experimental Psychology Section B*, *49*(1b), 60–80. https://doi.org/10.1080/713932614

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

George, D. N. (2020). The representation of stimulus conjunction in theories of associative learning: A context-dependent added-elements model. *Journal of Experimental Psychology: Animal Learning and Cognition*, *46*(3), 185–204. https://doi.org/10.1037/xan0000252

George, D. N., Haddon, J. E., & Griffiths, O. (2024). *Absence of differential protection from extinction in human causal learning*. https://osf.io/78mf5/

Griffiths, O., Holmes, N., & Westbrook, R. F. (2017). Compound stimulus presentation does not deepen extinction in human causal learning. *Frontiers in Psychology*, *8*, Article 120. https://doi.org/10.3389/fpsyg.2017.00120

Griffiths, O., & Westbrook, R. F. (2012). A common error term regulates acquisition but not extinction of causal judgments in people. *Learning & Behavior*, *40*(2), 207–221. https://doi.org/10.3758/s13420-011-0056-0

Hendry, J. S. (1982). Summation of undetected excitation following extinction of the CER. *Animal Learning & Behavior*, *10*(4), 476–482. https://doi.org/10.3758/BF03212287

Holland, P. C. (1980). Second-order conditioning with and without unconditioned stimulus presentation. *Journal of Experimental Psychology: Animal Behavior Processes*, *6*(3), 238–250. https://doi.org/10.1037/0097-7403.6.3.238

Holmes, N. M., Griffiths, O., & Westbrook, R. F. (2014). The influence of partner cues on the extinction of causal judgments in people. *Learning & Behavior*, *42*(3), 289–303. https://doi.org/10.3758/s13420-014-0146-x

Honey, R. C., & Hall, G. (1992). CS memory after trace conditioning. *Learning and Motivation*, *23*(2), 145–155. https://doi.org/10.1016/0023-9690(92)90014-D

Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford University Press.

Kamin, L. J. (1968). Attention-like processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimuli* (pp. 9–32). University of Miami Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kinder, A., & Lachnit, H. (2003). Similarity and discrimination in human Pavlovian conditioning. *Psychophysiology*, *40*(2), 226–234. https://doi.org/10.1111/1469-8986.00024

Kremer, E. F. (1978). The Rescorla–Wagner theory: Losses in associative strength in compound conditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *4*(1), 22–36. https://doi.org/10.1037/0097-7403.4.1.22

Krypotos, A.-M., & Engelhard, I. M. (2019). Targeting avoidance via compound extinction. *Cognition and Emotion*, *33*(7), 1523–1530. https://doi.org/10.1080/02699931.2019.1573718

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology B*, *57*(3), 193–243. https://doi.org/10.1080/02724990344000141

Lovibond, P. F., Been, S.-L., Mitchell, C. J., Bouton, M. E., & Frohardt, R. (2003). Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. *Memory & Cognition*, *31*(1), 133–142. https://doi.org/10.3758/BF03196088

Lovibond, P. F., Davis, N. R., & O'Flaherty, A. S. (2000). Protection from extinction in human fear conditioning. *Behaviour Research and Therapy*, *38*(10), 967–983. https://doi.org/10.1016/S0005-7967(99)00121-7

Luque, D., Flores, A., & Vadillo, M. A. (2013). Revisiting the role of within-compound associations in cue-interaction phenomena. *Learning & Behavior*, *41*(1), 61–76. https://doi.org/10.3758/s13420-012-0085-3

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276–298. https://doi.org/10.1037/h0076778

McConnell, B. L., & Miller, R. R. (2010). Protection from extinction provided by a conditioned inhibitor. *Learning & Behavior*, *38*(1), 68–79. https://doi.org/10.3758/LB.38.1.68

McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, *28*(3), 211–246. https://doi.org/10.3758/BF03200258

Mehta, R., & Russell, E. (2009). Effects of pretraining on acquisition of novel configural discriminations in human predictive learning. *Learning & Behavior*, *37*(4), 311–324. https://doi.org/10.3758/LB.37.4.311

Melchers, K. G., Lachnit, H., & Shanks, D. R. (2004). Past experience influences the processing of stimulus compounds in human Pavlovian conditioning. *Learning and Motivation*, *35*(3), 167–188. https://doi.org/10.1016/S0023-9690(03)00044-4

Melchers, K. G., Lachnit, H., Üngör, M., & Shanks, D. R. (2005). Prior experience can influence whether the whole is different from the sum of its parts. *Learning and Motivation*, *36*(1), 20–41. https://doi.org/10.1016/j.lmot.2004.06.002

Melchers, K. G., Shanks, D. R., & Lachnit, H. (2008). Stimulus coding in human associative learning: Flexible representations of parts and wholes. *Behavioural Processes*, *77*(3), 413–427; discussion 451–413. https://doi.org/10.1016/j.beproc.2007.09.013

Mitchell, C. J., & Lovibond, P. F. (2002). Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity. *The Quarterly Journal of Experimental Psychology B*, *55*(4), 311–329. https://doi.org/10.1080/02724990244000025

Morey, R. D., & Rouder, J. N. (2022). *BayesFactor: Computation of Bayes factors for common designs*. https://CRAN.R-project.org/package=BayesFactor

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

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, *94*(1), 61–73. https://doi.org/10.1037/0033-295X.94.1.61

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, *101*(4), 587–607. https://doi.org/10.1037/0033-295X.101.4.587

Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning & Behavior*, *30*(2), 73–95. https://doi.org/10.3758/BF03192911

Pearce, J. M., Esber, G. R., George, D. N., & Haselgrove, M. (2008). The nature of discrimination learning in pigeons. *Learning & Behavior*, *36*(3), 188–199. https://doi.org/10.3758/lb.36.3.188

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*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Pearce, J. M., Redhead, E. S., & George, D. N. (2002). Summation in autoshaping is affected by the similarity of the visual stimuli to the stimulation they replace. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*(2), 175–189. https://doi.org/10.1037//0097-7403.28.2.175

Pearce, J. M., & Wilson, P. N. (1991). Effects of extinction with a compound conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*(2), 151–162. https://doi.org/10.1037/0097-7403.17.2.151

Pineño, O. (2007). Protection from extinction by concurrent presentations of an excitor or an extensively extinguished CS. *Psicológica*, *28*(2), 151–166.

Pineño, O., Zilski, J. M., & Schachtman, T. R. (2007). Second-order conditioning during a compound extinction treatment. *Learning and Motivation*, *38*(2), 172–192. https://doi.org/10.1016/j.lmot.2006.08.004

Quigley, M., Bradley, A., & Haselgrove, M. (2023). Schizotypy dimensions do not predict overshadowing. *Behavioural Brain Research*, *453*, Article 114631. https://doi.org/10.1016/j.bbr.2023.114631

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reberg, D. (1972). Compound tests for excitation in early acquisition and after prolonged extinction of conditioned suppression. *Learning and Motivation*, *3*(3), 246–258. https://doi.org/10.1016/0023-9690(72)90021-5

Redish, A. D., Jensen, S., Johnson, A., & Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychological Review*, *114*(3), 784–805. https://doi.org/10.1037/0033-295X.114.3.784

Rescorla, R. A. (1971). Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. *Learning and Motivation*, *2*(2), 113–123. https://doi.org/10.1016/0023-9690(71)90002-6

Rescorla, R. A. (1981a). Simultaneous associations. In P. Harzem & M. Zeiler (Eds.), *Advances in analysis of behavior* (Vol. 2, pp. 47–80). Wiley.

Rescorla, R. A. (1981b). Within-signal learning in autoshaping. *Animal Learning & Behavior*, *9*(2), 245–252. https://doi.org/10.3758/BF03197827

Rescorla, R. A. (2000a). Associative changes in excitors and inhibitors differ when they are conditioned in compound. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*(4), 428–438. https://doi.org/10.1037/0097-7403.26.4.428

Rescorla, R. A. (2000b). Extinction can be enhanced by a concurrent excitor. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*(3), 251–260. https://doi.org/10.1037/0097-7403.26.3.251

Rescorla, R. A. (2001). Unequal associative changes when excitors and neutral stimuli are conditioned in compound. *The Quarterly Journal of Experimental Psychology B*, *54*(1), 53–68. https://doi.org/10.1080/02724990042000038

Rescorla, R. A. (2003). Protection from extinction. *Animal Learning & Behavior*, *31*(2), 124–132. https://doi.org/10.3758/BF03195975

Rescorla, R. A. (2006). Deepened extinction from compound stimulus presentation. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(2), 135–144. https://doi.org/10.1037/0097-7403.32.2.135

Rescorla, R. A., & Durlach, P. J. (1981). Within-event learning in Pavlovian conditioning. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Memory mechanisms* (pp. 81–111). Erlbaum.

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). Appleton.

Robinson, D., Hayes, A., & Couch, S. (2023). *broom: Convert statistical objects into tidy tibbles*. https://CRAN.R-project.org/package=broom

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *The Quarterly Journal of Experimental Psychology Section B*, *37*(1b), 1–21. https://doi.org/10.1080/14640748508402082

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2023). *afex: Analysis of factorial experiments*. https://CRAN.R-project.org/package=afex

Stanley, D. (2021). *apaTables: Create American Psychological Association (APA) style tables*. https://CRAN.R-project.org/package=apaTables

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review*, *114*(3), 759–783. https://doi.org/10.1037/0033-295X.114.3.759

Thomas, B. L., & Ayres, J. J. B. (2004). Use of the ABA fear renewal paradigm to assess the effects of extinction with co-present fear inhibitors or excitors: Implications for theories of extinction and for treating human fears and phobias. *Learning and Motivation*, *35*(1), 22–52. https://doi.org/10.1016/S0023-9690(03)00040-7

Urcelay, G. P., Lipatova, O., & Miller, R. R. (2009). Constraints on enhanced extinction resulting from extinction treatment in the presence of an added excitor. *Learning and Motivation*, *40*(4), 343–363. https://doi.org/10.1016/j.lmot.2009.04.003

Urcelay, G. P., & Miller, R. R. (2009). Potentiation and overshadowing in Pavlovian fear conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(3), 340–356. https://doi.org/10.1037/a0014350

Vervliet, B., Vansteenwegen, D., Hermans, D., & Eelen, P. (2007). Concurrent excitors limit the extinction of conditioned fear in humans. *Behaviour Research and Therapy*, *45*(2), 375–383. https://doi.org/10.1016/j.brat.2006.01.009

Wagner, A. R. (2003). Context-sensitive elemental theory. *The Quarterly Journal of Experimental Psychology Section B*, *56*(1b), 7–29. https://doi.org/10.1080/02724990244000133

Wasserman, E. A., & Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgement: The role of within-compound associations. *Quarterly Journal of Experimental Psychology*, *51B*(2), 121–138. https://doi.org/10.1080/713932675

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Williams, D. A., & Braker, D. S. (1999). Influence of past experience on the coding of compound stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*(4), 461–474. https://doi.org/10.1037/0097-7403.25.4.461

Williams, D. A., Sagness, K. E., & McPhee, J. E. (1994). Configural and elemental strategies in predictive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 694–709. https://doi.org/10.1037/0278-7393.20.3.694

Witnauer, J. E., & Miller, R. R. (2012). Associative status of the training context determines the effectiveness of compound extinction. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(1), 52–65. https://doi.org/10.1037/a0026333

(*Appendix follows*)

Appendix

List of Foods

apple
avocado
banana
beans
bread
broccoli
carrot
cauliflower
cheese
cherry
chicken
chocolate
corn
cucumber
egg
garlic
grapes
pepper
ham
ice cream
lamb chop
leek
lettuce
mackerel
melon
mushroom
onion
orange
parsnip
pasta
peach
pear
peas
pineapple
potato
rice
salmon
steak
strawberry
tomato

Received September 27, 2023
Revision received March 12, 2024
Accepted March 13, 2024 ■