Journal of Experimental Psychology:
Animal Behavior Processes
2005, Vol. 31, No. 2, 226–236

0097-7403/05/$12.00 DOI: 10.1037/0097-7403.31.2.226

The Outcome Specificity of Learned Predictiveness Effects: Parallels Between Human Causal Learning and Animal Conditioning

M. E. Le Pelley
Cardiff University

S. M. Oakeshott
Brooklyn College of the City University of New York

A. J. Wills
University of Exeter

I. P. L. McLaren
University of Cambridge

Two experiments examined the outcome specificity of a learned predictiveness effect in human causal learning. Experiment 1 indicated that prior experience of a cue–outcome relation modulates learning about that cue with respect to a different outcome from the same affective class but not with respect to an outcome from a different affective class. Experiment 2 ruled out an interpretation of this effect in terms of context specificity. These results indicate that learned predictiveness effects in human causal learning index an associability that is specific to a particular class of outcomes. Moreover, they mirror demonstrations of the reinforcer specificity of analogous effects in animal conditioning, supporting the suggestion that, under some circumstances, human causal learning and animal conditioning reflect the operation of common associative mechanisms.

In the field of animal conditioning, it is relatively well established that experience of a predictive (or nonpredictive) relation between a conditioned stimulus (CS) and an unconditioned stimulus (US) appears to affect the processing power devoted to learning about that CS on subsequent learning episodes. One model of such learned predictiveness effects in animal conditioning is that proposed by Mackintosh (1975), which states that the change in associative strength (ΔV) for CS A on each learning episode is given by

$$\Delta V_A = S\alpha_A(\lambda - V_A), \tag{1}$$

where V_A represents the associative strength of CS A, S is a constant learning-rate parameter, and λ is the asymptote of conditioning supportable by the US occurring on that trial. α_A represents the associability of Cue A. Mackintosh allowed associability to change as a result of experience of a cue's predictiveness, with animals proposed to devote more processing power to stimuli that are uniquely successful in their predictions. Specifically, Cue A maintains a high associability to the extent that it is a better predictor of the outcome of the current trial than are all other cues present. Conversely, associability decreases if the outcome is predicted by other events at least as well as by Cue A. The extent to which the outcome is predicted by Cue A is represented by the absolute value of the error term ($\lambda - V_A$). These ideas can therefore be encapsulated in the following rules:

$$\Delta\alpha_A > 0 \text{ if } |\lambda - V_A| < |\lambda - V_Z| \tag{2}$$

and

$$\Delta\alpha_A < 0 \text{ if } |\lambda - V_A| \geq |\lambda - V_Z|, \tag{3}$$

where V_Z is the associative strength of all stimuli other than Cue A present on that trial. The size of the change in associability on each trial is proportional to the magnitude of these inequalities.

It should be noted that this is not the only theory of associative learning to incorporate a variable associability mechanism. For instance, Pearce and Hall (1980) proposed an alternative approach that, although it takes a very different view of the way associability operates, is able to predict many of the same learned predictiveness effects as is the Mackintosh (1975) model. That said, the Pearce–Hall model is less readily applied to the particular experiments dealt with in the current article (see the General Discussion), and, hence, the approach to associability taken by the Mackintosh model provides our main focus in this article.

The suggestion that the experienced predictiveness of a CS may influence the ability of that CS to enter into subsequent associations is supported by substantial evidence in the field of animal conditioning (see Le Pelley, 2004, for a recent review). Some of the most convincing evidence comes from studies of blocking and unblocking. *Blocking* refers to the finding that the gain in excitatory strength accruing to a cue, B, following reinforcement of an AB compound is much reduced if Cue A has previously been trained as being a good predictor of that US (Kamin, 1969). According to the Mackintosh (1975) model, pretraining of Cue A establishes it as a good predictor of the outcome. The novel Cue B presented on subsequent AB trials is therefore a poorer predictor of the outcome than is Cue A, and its associability falls correspond-

M. E. Le Pelley, School of Psychology, Cardiff University, Cardiff, Wales; S. M. Oakeshott, Department of Psychology, Brooklyn College of the City University of New York; A. J. Wills, School of Psychology, University of Exeter, Exeter, England; I. P. L. McLaren, Department of Experimental Psychology, University of Cambridge, Cambridge, England.

This work was supported by Economic and Social Research Council Grant RG36281. We are grateful to Alison King and Rossy McLaren for running a number of the participants in this experiment.

Correspondence concerning this article should be addressed to M. E. Le Pelley, School of Psychology, Cardiff University, Cardiff CF10 3AT, Wales. Email: m.lepelley@psychol.cam.ac.uk

ingly, such that changes in V_B are smaller than in a control group that has not had pretraining with Cue A.

It is, of course, possible to explain blocking without appealing to a variable associability. For example, according to Rescorla and Wagner's (1972) model, learning is dictated by the discrepancy between the magnitude of the US occurring on the current trial and the extent to which the magnitude of that US is predicted by the CSs present. Blocking occurs because the presence of Cue A on AB trials ensures that this discrepancy is small, such that little is learned about Cue B.

Unblocking refers to experimental manipulations that are found to attenuate blocking. For instance, if Cue A is followed by a single delivery of food (US1) and an AB compound is then followed by two successive food deliveries (US1 and US2), more excitatory conditioning to Cue B is observed than for a standard blocking contingency (Dickinson & Mackintosh, 1979; Holland, 1984, 1988; a similar experiment using shock instead of food USs was presented by Dickinson, Hall, & Mackintosh, 1976). This unblocking as a result of the surprising addition of a reinforcer on compound trials is consistent with the view taken by the Mackintosh (1975) model. The added US2 on the first AB trial is surprising and is predicted (or, rather, not predicted) by Cue B as well as it is by Cue A. Therefore, the addition of US2 serves to maintain Cue B's associability, promoting its ability to enter into excitatory associations with US1.[1] Unblocking by reinforcer addition is also consistent with the Rescorla–Wagner (1972) model: The surprising increase in the overall magnitude of reinforcement on AB trials supports additional excitatory learning about Cue B. There are, however, two further aspects of unblocking that defy explanation in terms of the Rescorla–Wagner model, instead indicating the operation of a variable associability.

First comes evidence from a study of unblocking by reinforcer addition by Holland (1988), in which US2 was qualitatively different from US1 (sucrose solution vs. solid food pellets) and generated a conditioned response with a different response topography, such that, on test, it was possible to determine the nature of learning about the added CS on compound trials. This experiment indicated that the major function of the added US2 was to enhance the development of associations between the added CS and the original reinforcer, US1. This runs contrary to the Rescorla–Wagner model (1972), which views unblocking as resulting from excitatory conditioning of Cue B with respect to the surprising US2, not the predicted US1.

Second come demonstrations of excitatory unblocking as a result of the omission of an expected reinforcer (Dickinson et al., 1976; Dickinson & Mackintosh, 1979; Holland, 1984, 1988). In these studies, trials on which Cue A is followed by US1 and then US2 precede trials on which Cue AB is followed by US1 alone. Just as in the case of the unexpected addition of US2 on AB trials, omission of the expected US2 on AB trials leads to unblocking. This is, again, incompatible with the Rescorla–Wagner (1972) model: Given that the overall magnitude of reinforcement on AB trials is less than that predicted by the presence of Cue A, this model predicts that such training will endow Cue B with inhibitory properties. Instead, it seems that the surprising omission of US2 facilitates the association of Cue B with US1, as predicted by the Mackintosh (1975) model. On the initial AB trial, US2 is absent but expected on the basis of Cue A. Cue B is a better predictor of the absence of US2 than is Cue A, and, hence, its associability is maintained at a high level, allowing it to enter into association with the remaining US1.

In addition to ruling out the Rescorla–Wagner (1972) view of unblocking, Holland's (1988) study went some way toward characterizing the mechanism controlling CS processing. As mentioned earlier, the added US2 on compound trials was qualitatively different from the original US1 (sucrose solution vs. solid food pellets). Holland found that addition of this different US2 enhanced the development of associations between the added CS and US1 at least as well as if US1 and US2 were identical. Thus, it is clear that the associability of a cue is not entirely reinforcer specific. Instead, it seems that changes in the associability of a cue caused by experience of its predictive relationship with regard to a particular outcome can, under some circumstances at least, affect the rate of learning about that cue with regard to a different outcome.

However, Dickinson and Mackintosh (1979) demonstrated that this is not always the case. Although qualitatively different, the reinforcers used by Holland (1988) were both appetitive. Dickinson and Mackintosh, conversely, used reinforcers drawn from different affective classes: one appetitive (food pellets), and the other aversive (electric shock). They found that unblocking by addition and omission did not occur when US1 was appetitive and US2 was aversive, or vice versa.

In sum, it seems that learned predictiveness developed with respect to one US modulates learning with respect to a different US drawn from the same affective class but not learning with respect to a US drawn from a different affective class. This is important, because it indicates that the CS processing effects demonstrated by studies of unblocking in animal conditioning do not reflect the operation of a general attentional process (e.g., Sutherland & Mackintosh, 1971). Such an account anticipates that an event that is able to increase attention to a CS (e.g., surprising addition or omission of a reinforcer) promotes learning of a relation between that CS and any presented outcome. Instead, the data support the idea that CS processing effects in unblocking index a stimulus-specific learning-rate parameter that, although not completely reinforcer specific, is also not an entirely general property of a cue. In the preceding discussion, we have focused on the approach to such CS processing mechanisms in unblocking offered by the Mackintosh (1975) model; in the General Discussion, we consider an alternative approach to this effect suggested by Pearce and Hall (1980).

Dickinson, Shanks, and Evenden (1984) noted a number of similarities between the factors influencing human causal learning (acquisition of a causal judgment as a result of experience of the relation between a cause and an effect) and animal conditioning (development of conditioned responding as a result of experience of the relation between a CS and a US). This led them to suggest that, under some circumstances, at least, a common associative

[1] In fact, unblocking is still predicted even if generalization of Stage 1 Cue A–US1 learning means that Cue A begins Stage 2 as a better predictor of US2 than is Cue B. Although this results in a decline in Cue B's associability, this decline is slower than for a control group in which Compound AB is followed by US1 only, this being well predicted by the presence of Cue A.

mechanism might underlie the two. Does this parallel extend to learned predictiveness effects?

Although it is well established that stimulus-processing mechanisms exert an influence on human categorization and discrimination learning (e.g., Kruschke, 1996; Whitney & White, 1993; Wolff, 1967; Zeaman & House, 1974), it is only recently that the operation of such mechanisms has been demonstrated in causal learning (Kruschke & Blair, 2000; Le Pelley & McLaren, 2003; Le Pelley, Oakeshott, & McLaren, 2005; Lochmann & Wills, 2003). In support of the view taken by Dickinson et al. (1984), the results of these studies fit well with the predictions of associability models developed on the basis of animal studies. We consider in some detail the study by Le Pelley and McLaren, as it forms the basis of the experiments presented in this article. The design of this experiment is shown in Table 1: It used a multiple-outcomes allergy prediction paradigm, with human participants playing the part of an allergist trying to judge the likelihood that various foods would cause different types of allergic reaction in fictitious patients. The foods therefore constitute the cues (causes), and the various allergic reactions are the outcomes (effects). In Table 1 the letters *A–Y* represent different foods, and the numbers *1–4* represent different types of allergic reaction that patients could suffer as a result of eating these foods (nausea, dizziness, itch, and sweating).

On each trial of Stage 1, participants were told the contents of a meal eaten by Mr. X and were asked to predict the type of allergic reaction that he would suffer (given a choice of Allergy 1 or Allergy 2). Cues A and D consistently indicated the occurrence of Allergy 1, Cues B and C consistently indicated the occurrence of Allergy 2, and Cues V–Y provided no basis for discrimination between the two outcomes—they were paired with Allergies 1 and 2 an equal number of times. According to the Mackintosh (1975) model, the associability of Cues A–D should remain high over Stage 1 trials, as they are the best predictors of the outcome on each trial; the associability of Cues V–Y should decrease, as they are poorer predictors of the outcome on each trial.

In Stage 2, participants were told that they would be given information regarding foods and allergies for a new patient, Mr. Y. On each of the first four Stage 2 trial types shown in Table 1, a good predictor from Stage 1 (Cue A, B, C, or D) was paired with a poor predictor (Cue V, W, X, or Y) with which it had not been paired in Stage 1, and this novel compound was paired with a novel outcome: Compounds AX and CV were paired with Allergy 3, whereas Compounds BY and DW were paired with Allergy 4.[2]

Table 1
Design of Le Pelley and McLaren (2003) and Experiment 1

Stage 1	Stage 2	Test
AV → 1	AX → 3	AC
BV → 2	BY → 4	BD
AW → 1	CV → 3	VX
BW → 2	DW → 4	WY
CX → 2	EF → 3	EH
DX → 1	GH → 4	FG
CY → 2	IJ → 3	IJ
DY → 1	KL → 4	KL

Note. Letters *A* through *Y* represent different foods; Numbers *1* through *4* represent different types of allergic reaction that patients could suffer as a result of eating these foods.

Following Stage 2, participants were asked to rate the likelihood that various meals composed of two foods would cause each of Allergies 3 and 4. For each meal, participants provided two ratings: one of how likely that meal was to cause Allergy 3; the other of how likely it was to cause Allergy 4. In the approach taken by Dickinson et al. (1984), these causal judgment ratings provide an index of the strength of cue–outcome associations developed over the course of training. The question of interest is how well participants had learned the various Stage 2 mappings between foods and the allergies with which they were paired. Accordingly, for each test compound, the causal judgment rating on the Allergy 4 scale was subtracted from that on the Allergy 3 scale to yield a difference score, revealing the differential predictiveness of that compound for each of the two allergies—that is, the extent to which it predicted Allergy 3 more (or less) than Allergy 4.

Allergies 3 and 4 were novel at the outset of Stage 2, so the causal strength of all cues for these two outcomes should have begun this stage at zero. Of course, the different allergic reactions had some degree of similarity to one another, so there might have been generalization of associative strength to these novel outcomes from Stage 1 learning. Nevertheless, given the randomization and counterbalancing manipulations used in this study, this generalization could not have selectively differentiated between Allergy 3 and Allergy 4. Combined with the use of difference scores as outlined above, this ensured that generalization from Stage 1 could not have any systematic effect on differential learning in Stage 2. For example, even though in Stage 1 participants might have learned that Cue A was a good signal for Allergy 1, any generalization of associative strength from Stage 1 to Stage 2 with respect to Allergy 3 was equal to that with respect to Allergy 4, yielding a differential predictiveness (as assessed by the difference score) of zero. According to similar logic, the differential predictiveness for all cues began Stage 2 at zero. Thus, the use of difference scores allowed the assessment of the relative rates of selective learning about cue–outcome relations in Stage 2 by ensuring that all cues started from the same baseline.

Note that the objective cue–outcome contingency for all of Cues A–Y was identical during Stage 2: All cues were equally reliable as predictors of Stage 2 outcomes. Hence, any difference in selective learning of Stage 2 cue–outcome relations must reflect differences in the processing afforded to cues as a result of their treatment in Stage 1.

We have shown that the Mackintosh (1975) model predicts that, at the end of Stage 1, Cues A–D (good predictors in Stage 1) will have a higher associability than Cues V–Y (poor predictors in Stage 1). This promotes more rapid learning of associations between the good predictors and the Stage 2 outcomes than between the poor predictors and the same outcomes over the course of Stage 2. Therefore, on test, participants should judge Compound AC as a strong predictor of Allergy 3, Compound BD as a strong predictor of Allergy 4, Compound VX as a weak predictor of Allergy 3 and Compound WY as a weak predictor of Allergy 4. In other words, the discrimination between Compounds AC and BD should be greater than that between Compounds VX and WY. This

[2] The remaining four trial types in Stage 2 (EF, GH, IJ, and KL) and related test compounds (EH, FG, IJ, and KL) were included as filler items and are not discussed further in the current article.

was exactly the pattern seen in Le Pelley and McLaren's (2003) data. This result clearly fits with the suggestion that a common mechanism underlies animal conditioning and human causal learning: Not only does human causal learning reveal effects of learned predictiveness, these effects are in line with a model of CS-processing changes developed to account for the results of studies of animal conditioning.

The research reported in the current article aims to investigate this parallel further by probing the outcome specificity of this learned predictiveness effect in human causal learning. Learned predictiveness effects in animal conditioning transfer between outcomes drawn from the same affective class but not between outcomes from different affective classes, which rules out an account of these effects in terms of general attention to cues. Is the same true for analogous effects in human learning?

Le Pelley and McLaren's (2003) study went some way toward addressing this question by demonstrating that learned predictiveness effects in human causal learning, like those in animal conditioning, are not completely outcome specific. Experience of the relations between cues and Outcomes 1 and 2 during Stage 1 modulated subsequent learning about those cues with respect to (qualitatively different) Outcomes 3 and 4 during Stage 2. However, as in Holland's (1988) animal study, the outcomes used in the two stages were similar, all coming from the same affective class (all were aversive). Whether learned predictiveness effects in human causal learning transfer between outcomes from different affective classes remains to be established.

Experiment 1

Experiment 1 uses the same design as did Le Pelley and McLaren's (2003) earlier study (see Table 1) but varies the affective class of the Stage 1 and Stage 2 outcomes in a factorial design. Participants played the part of a dietician looking at the effects of various foods on fictitious patients. For different groups of participants, the results of eating the foods could be either aversive (allergic reactions) or appetitive (enjoyment reactions). The ALL–ALL group experienced allergic reactions in Stages 1 and 2 and, thus, essentially constituted a replication of Le Pelley and McLaren's (2003) earlier study, with minor differences in experimental instructions and assignment of allergies to outcomes. The ENJ–ENJ group experienced enjoyment reactions in Stages 1 and 2. In both of these groups, Stage 1 and Stage 2 outcomes were drawn from the same affective class (aversive for the ALL–ALL group; appetitive for the ENJ–ENJ group). Given the results of Holland's (1988) animal study and Le Pelley and McLaren's human study, we expected the experience of relations between cues and outcomes in Stage 1 to modulate learning of cue–outcome associations in Stage 2 for the ALL–ALL and ENJ–ENJ groups.

The ALL–ENJ group experienced allergic reactions in Stage 1 and enjoyment reactions in Stage 2; the ENJ–ALL group experienced enjoyment reactions in Stage 1 and allergic reactions in Stage 2. For both groups, Stage 1 and Stage 2 outcomes were drawn from different affective classes. If learned predictiveness effects in human causal learning, like those in animal conditioning, do not transfer between outcomes from different affective classes, then we expect to see no effect of Stage 1 predictiveness on learning about cues in Stage 2 for the ALL–ENJ and ENJ–ALL groups.

Method

Participants, apparatus, and stimuli. Two hundred ten volunteers aged 17–35 years took part in the experiment. Participants were randomly assigned to groups, with 43 participants each in the ALL–ALL, ENJ–ENJ, and ALL–ENJ groups and 41 in the ENJ–ALL group. All volunteers received a small honorarium payment for participation.

The experiment was run on a Macintosh computer with a 15-in. (38.1-cm) monitor. The 16 foods used were asparagus, banana, carrots, sardines, tomato, mushrooms, pasta, eggs, onion, dates, ham, lentils, garlic, rice, vinegar, and yogurt. These foods were randomly and independently assigned to Cues A–Y in the experimental design shown in Table 1 for each participant.

For the ALL–ALL and ALL–ENJ groups, Outcomes 1 and 2 were itch and sweating. Assignment of allergies to outcomes was counterbalanced across participants—for half of the participants in each group, Outcome 1 was itch and Outcome 2 was sweating, whereas for the other half, Outcome 1 was sweating and Outcome 2 was itch. For the ENJ–ENJ and ENJ–ALL groups, Outcomes 1 and 2 were licks lips and contented sigh, counterbalanced in similar fashion.

For the ALL–ALL and ENJ–ALL groups, Outcomes 3 and 4 were nausea and dizziness. For the ENJ–ENJ and ALL–ENJ groups, Outcomes 3 and 4 were broad smile and thumbs up. Once again, assignment of reactions to Stage 2 outcomes was counterbalanced across participants.

Binary counterbalancing of Stage 1 and Stage 2 outcomes, along with the presentation order of rating scales on test (see below), yielded eight counterbalance conditions per group. Given the exclusion of participants as detailed in the *Results and Discussion* section, which left 32 participants in each group, this counterbalancing was complete, with 4 participants per group in each counterbalance condition.

Procedure. At the outset of the experiment, the following instructions were given to all groups, with small differences appropriate to each group. Sections in brackets indicate text that differed among the groups according to whether the following stage of the experiment involved allergic or enjoyment reaction outcomes. In each case, the first option relates to the use of allergic reactions, whereas the second option relates to the use of enjoyment reactions.

> In this experiment you are asked to imagine that you are a dietician, investigating the effects of food on people. You have just been presented with a new patient, "Mr. X."
>
> In an attempt to discover the effects of various foods on Mr. X, you arrange for him to eat a number of different meals, each containing two foods. You observe the effects of these foods on Mr. X from behind a one-way mirror.
>
> Eating the foods contained in these meals causes Mr. X to [suffer/display] one of two different types of [allergic/favorable] reaction. After some of the meals Mr. X is seen to [start sweating/lick his lips]. After others he is seen to [develop an itch/give a contented sigh].
>
> On the following screens, you will be shown the contents of meals eaten by Mr. X, and will be asked to predict what kind of reaction will result from eating each meal. For each meal you will have a choice of two [allergic/favorable] reactions, ["Sweating" and "Itch"/"Licks Lips" and "Contented Sigh"]. Mark your prediction by clicking the option button next to one of them, and then click OK. You will then be told whether your prediction was correct or incorrect. If your prediction was incorrect, the computer will beep.
>
> You will have to guess at first, but with the aid of the feedback your predictions should soon start to become more accurate. Your reaction times are not important in this experiment: you may take as long as you like on each trial.

On each Stage 1 trial, participants saw the message "Meal [meal number] contains the following foods," followed by two foods, one above the other. Below this were two types of reaction, one above the other, with a radio button next to each. Participants entered their prediction as to which type of reaction would result when Mr. X ate the meal by clicking the button next to that reaction and then clicking an *OK* button. The screen then cleared, and immediate feedback was provided: If participants had made the correct decision, the word *Correct* appeared in a green box; if they had made the incorrect decision, the word *Wrong* appeared in a red box and the computer beeped.

Stage 1 comprised 14 blocks, with each of the eight trial types occurring once per block. Trial order within a block was randomized, with the constraint that there could be no immediate repetitions across blocks. For each trial type, the order of presentation of the foods on the screen (top vs. bottom) was counterbalanced across blocks. For example, for trial type AV $\rightarrow$ 1, there were seven presentations with Food A above Food V and seven presentations with Food V above Food A (the order of these presentations was randomized). The two reaction options presented on each Stage 1 trial were always Outcome 1 and Outcome 2. For each trial type, the order of presentation of these reactions on the screen (top vs. bottom) was counterbalanced across blocks. Therefore, for trial type AV $\rightarrow$ 1, there were seven presentations with Outcome 1 above Outcome 2 and seven presentations with Outcome 2 above Outcome 1 (again in random order).

After the 112 trials of Stage 1, the following message appeared on the screen (again, the first option in brackets relates to the use of allergic reactions in the following stage):

> Treatment of Mr. X is now finished. In the next stage, you will be studying a new patient, Mr. Y.
>
> Once again, in an attempt to discover the effects of various foods on Mr. Y, you arrange for him to eat a number of different meals, each containing two foods. Some of the foods are the same as those given to Mr. X, some are not. Again you observe the effects of these foods on Mr. Y from behind a one-way mirror.
>
> Eating the foods contained in these meals causes Mr. Y to [suffer/display] one of two different types of [allergic/favorable] reaction. After some of the meals Mr. Y is seen to [suffer from dizziness/give a broad smile]. After others he is seen to [feel nauseous/give a thumbs up sign].
>
> On the following screens, you will be shown the contents of meals eaten by Mr. Y, and will be asked to predict what kind of reaction will result from eating each meal. For each meal you will have a choice of two [allergic/favorable] reactions, ["Dizziness" and "Nausea"/"Broad Smile" and "Thumbs Up"]. As before, make your prediction by clicking the option button next to one of them, and then click OK.

The form of each Stage 2 trial was exactly the same as that for Stage 1, except that now the two outcome options were always Outcome 3 and Outcome 4. There were four blocks in Stage 2, with each of the eight trial types appearing once per block. Counterbalancing and randomization of trial order, food presentation order, and outcome presentation order were as for Stage 1.

At the completion of Stage 2, the following message appeared on the screen:

> You will now be shown a number of meals to be eaten by Mr. Y. On the basis of the contents of these meals, you are asked to rate the likelihood of each reaction that is typically observed in Mr. Y.
>
> Rate the likelihood of each reaction occurring on a scale from 0–10. A rating of 0 means that eating the meal is very unlikely to cause that type of reaction, while a rating of 10 means that eating the meal is very likely to cause that type of reaction. To enter your rating, click on the appropriate option button. Once you have rated the meal with respect to both reactions, click OK. Remember, you will have to rate each meal twice: once for each type of reaction.

For clarification, participants also had access to a card on which instructions on how to use the rating scale were printed. Each of the eight test compounds was presented in random order for rating. On each test trial, the message "Meal [meal number] contains the following foods" appeared, followed by the two foods and, below them, the message "How likely is it that the following effects will occur in Mr. Y after eating this meal?" Below that were two boxes placed side by side, with the title of Outcome 3 (e.g., *Nausea*) at the top of one and the title of Outcome 4 at the top of the other. In each box were 11 radio buttons labeled 0–10, one above the other, with 0 at the bottom and 10 at the top. Participants were cued to enter a rating for each reaction by clicking the appropriate radio button in each box. Once they had provided a rating for the current meal for each of the two reactions, they clicked an *OK* button to progress to the next trial. The order of presentation of the two foods in the meal (top vs. bottom) was randomized, and the order of presentation of the two allergy scales (left vs. right) was counterbalanced over participants (half of the participants had the rating scale for Outcome 3 on the left and Outcome 4 on the right; for the other half, these were reversed).

Results and Discussion

We could only expect to see any effect of differences in learned predictiveness during Stage 1 on Stage 2 learning if participants were able to learn the contingencies involved. We therefore imposed a selection criterion of 50% correct or more on all the trials of each stage. Eleven participants in each of the ALL–ALL, ENJ–ENJ, and ALL–ENJ groups and 9 participants in the ENJ–ALL group failed to meet this criterion, showing numerically worse than chance performance in Stage 1 or in Stage 2. The data for these participants were excluded from all further analyses, which left 32 participants per group. Following this exclusion, the mean percentage correct over all trials of Stage 1 was 79.6% for the ALL–ALL group, 77.1% for the ENJ–ENJ group, 76.7% for the ALL–ENJ group, and 76.8% for the ENJ–ALL group (all ts $<$ 1). The mean percentage correct over all trials of Stage 2 was 66.5% for the ALL–ALL group, 64.5% for the ENJ–ENJ group, 64.6% for the ALL–ENJ group, and 65.8% for the ENJ–ALL group (all ts $<$ 1).

We calculated difference scores for each of the test compounds, AD, BC, VX, and WY, for each participant by subtracting the causal judgment rating provided on the Allergy 4 scale from that on the Allergy 3 scale. Figure 1 shows mean difference scores for these compounds for the four groups of Experiment 1. The basic data for analysis concerned discrimination between Compounds AC and BD (given by the difference score for AC minus the difference score for BD) compared with discrimination between Compounds VX and WY (difference score for VX minus difference score for WY). In three of the four groups, Kolmogorov–Smirnov tests revealed that the data for at least one of these two discriminations were not normally distributed, $D_{\max}(32) = 0.214$, $p < .01$, which ruled out analysis of variance as a statistical approach. Consequently, we used nonparametric tests for data analysis. All probabilities given in these and all subsequent analyses are two-tailed unless otherwise specified.

The ALL–ALL group demonstrated significant discrimination between Compounds AC and BD, with AC eliciting higher difference scores than BD, Wilcoxon's $T(26) = 46.0$, $p < .01$. Discrimination between Compounds VX and WY was also significant, with VX eliciting higher scores than WY, $T(25) = 64.5$, $p < .01$. A learned predictiveness effect would be revealed if the discrimination between Compounds AC and BD was better than that

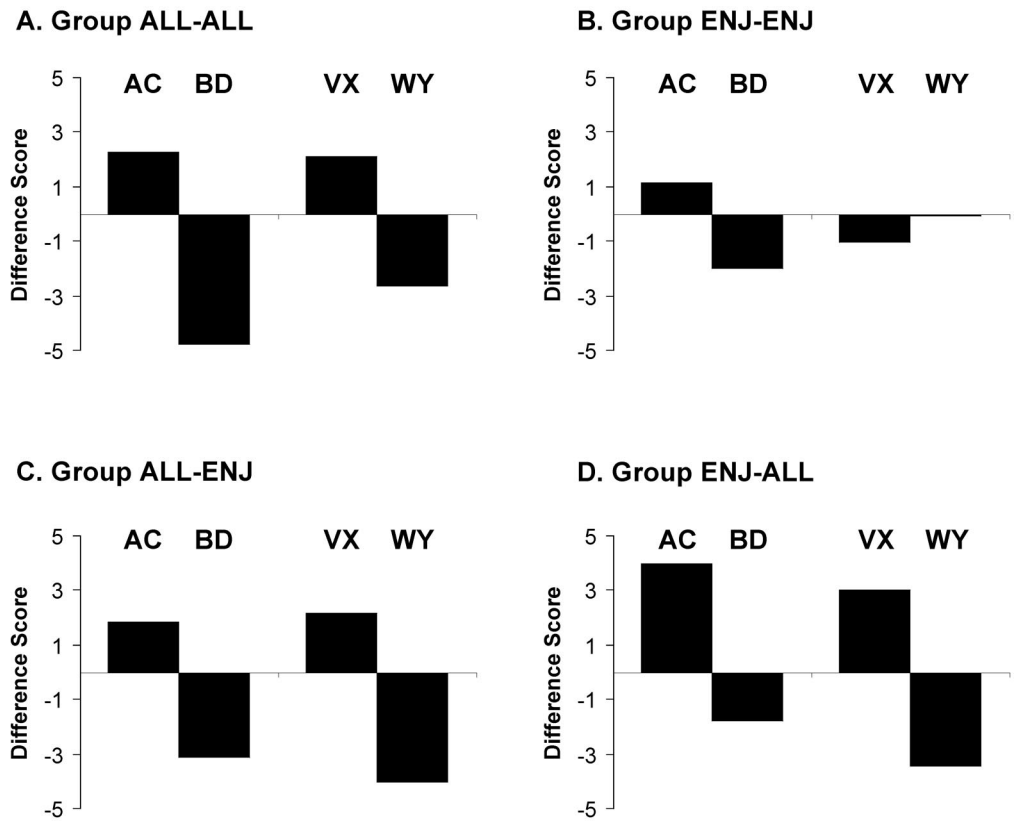

Figure 1. Difference scores for Compounds AC, BD, VX, and WY for the four groups of Experiment 1: the ALL–ALL group (Panel A), the ENJ–ENJ group (Panel B), the ALL–ENJ group (Panel C), and the ENJ–ALL group (Panel D). We calculated difference scores for each compound by subtracting the causal judgment rating for Outcome 4 from that for Outcome 3. A positive difference score indicates that the compound is perceived as predictive of Outcome 3; a negative score indicates that the compound is perceived as predictive of Outcome 4. A learned predictiveness effect is revealed by better discrimination (i.e., larger difference in difference scores) between Compounds AC and BD than between Compounds VX and WY. ALL–ALL = this group experienced allergic reactions in both Stage 1 and Stage 2; ENJ–ENJ = this group experienced enjoyment reactions in both Stage 1 and Stage 2; ALL–ENJ = this group experienced allergic reactions in Stage 1 and enjoyment reactions in Stage 2; ENJ–ALL = this group experienced enjoyment reactions in Stage 1 and allergic reactions in Stage 2.

between Compounds VX and WY—that is, if the difference in difference scores between Compounds AC and BD was significantly greater than that between Compounds VX and WY. Unfortunately, this comparison fell just short of significance, $T(27) = 121.0$, $p = .0501$ (one-tailed). A one-tailed test is appropriate here, as this is a replication of an earlier result with a group given essentially identical treatment (Le Pelley & McLaren, 2003).

For the ENJ–ENJ group, we observed significant discrimination between Compounds AC and BD, $T(26) = 96.5$, $p < .05$, but there was no reliable discrimination between Compounds VX and WY, $T(27) = 161.5$, $p > .10$. We observed a clearer learned predictiveness effect in this group, in that the discrimination between Compounds AC and BD was significantly better than that between Compounds VX and WY, $T(30) = 127.0$, $p < .05$. Thus, for both groups using outcomes drawn from the same affective class in each stage, we have some evidence that experience of the relation between cues and Stage 1 outcomes was able to modulate subsequent learning of the causal relations between cues and outcomes in Stage 2 (although the evidence from the ALL–ALL group is relatively weak).

The ALL–ENJ group exhibited significant discrimination between Compounds AC and BD, $T(28) = 68.0$, $p < .01$, and between Compounds VX and WY, $T(29) = 73.5$, $p < .01$. In contrast with the former groups, we observed no learned predictiveness effect in the ALL–ENJ group: Discrimination between Compounds AC and BD did not differ significantly from that between Compounds VX and WY, $T(28) = 197.0$, $p > .50$.

The ENJ–ALL group showed a pattern of performance similar to that of the ALL–ENJ group. That is, discrimination between Compounds AC and BD was reliable, $T(23) = 44.0$, $p < .01$, as was that between Compounds VX and WY, $T(26) = 17.5$, $p < .01$, but discrimination between AC and BD did not differ from that between VX and WY, $T(27) = 170.0$, $p > .50$. Thus, for both groups using outcomes drawn from different affective classes, experience of cue–outcome relations in Stage 1 had no effect on learning of new cue–outcome relations in Stage 2.

The data from Figure 1 are replotted in Figure 2. This figure shows the discrimination between Compounds AC and BD and between Compounds VX and WY, averaged for groups using outcomes drawn from the same affective class in Stages 1 and 2

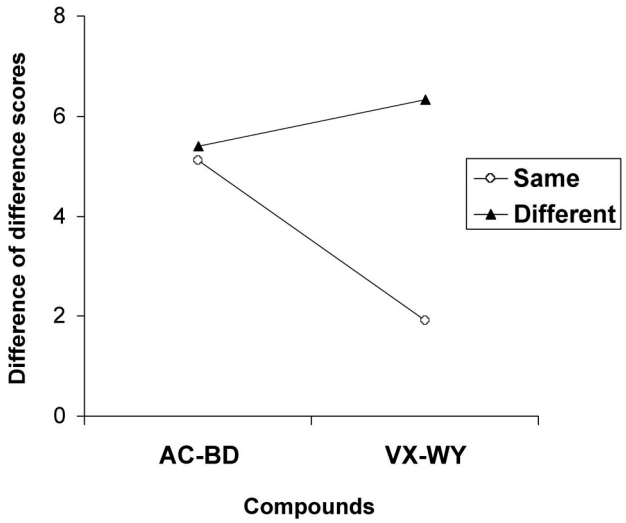

Figure 2. Difference in difference scores for Compounds AC and BD (AC–BD) and for Compounds VX and WY (VX–WY) for groups trained with outcomes drawn from the same affective class in Stages 1 and 2 (the ALL–ALL and ENJ–ENJ groups) and groups trained with outcomes drawn from different affective classes in Stages 1 and 2 (the ALL–ENJ and ENJ–ALL groups). Larger scores indicate better discrimination between compounds. ALL–ALL = this group experienced allergic reactions in both Stage 1 and Stage 2; ENJ–ENJ = this group experienced enjoyment reactions in both Stage 1 and Stage 2; ALL–ENJ = this group experienced allergic reactions in Stage 1 and enjoyment reactions in Stage 2; ENJ–ALL = this group experienced enjoyment reactions in Stage 1 and allergic reactions in Stage 2.

(the ALL–ALL and ENJ–ENJ groups, labeled *same*) and for groups using outcomes drawn from different affective classes (the ALL–ENJ and ENJ–ALL groups, labeled *different*). In this figure, large scores indicate good discrimination between compounds. A profound effect of learned predictiveness is seen in the *same* condition: Experience of differences in predictiveness during Stage 1 modulates development of causal judgments as a result of Stage 2 training, such that on test the AC–BD difference was greater than the VX–WY difference, $T(57) = 488.0$, $p < .01$. In the *different* condition, however, experience of differences in learned predictiveness between Cues A–D and V–Y during Stage 1 had no effect on learning about these cues in Stage 2: The AC–BD difference did not differ significantly from the VX–WY difference, $T(55) = 719.0$, $p > .50$. Most important, the difference between AC–BD and VX–WY was significantly greater in the *same* condition than in the *different* condition, Mann–Whitney $U(64, 64) = 1{,}574$, $p < .03$.

This latter analysis demonstrates a difference in learned predictiveness effects dependent on the particular outcomes used. The effect of earlier learning on later was greater when both earlier and later learning involved outcomes from the same affective class than when they involved outcomes from different affective classes (in fact, when outcomes from different classes were used, no effect at all was observed). The finding of a learned predictiveness effect in the *same* condition, despite the fact that the outcomes used in the two stages of the experiment were qualitatively different for each group involved, agrees with the results of Le Pelley and McLaren (2003) in indicating that associability is not an entirely outcome-specific property of a cue. The fact that no such effect was seen in the *different* condition, however, indicates that associability is also not an entirely general property of a cue.

Although the results of Experiment 1 are consistent with the idea that learned predictiveness effects in human learning are partially outcome specific, an alternative interpretation is possible. For the ALL–ALL and ENJ–ENJ groups, the experimental contexts of Stages 1 and 2 were relatively similar. In the ALL–ALL group, for example, participants were presented with information regarding allergic reactions in Stage 1 and information regarding other allergic reactions in Stage 2—albeit for a different patient, Mr. Y rather than Mr. X. The ALL–ENJ and ENJ–ALL groups, however, experienced a larger change of context between the two stages. Having dealt with allergic reactions in Stage 1, participants in the ALL–ENJ group were told that in Stage 2 they would be looking at enjoyment reactions. Perhaps it was this context change rather than any degree of outcome specificity that was the crucial determinant of whether learned predictiveness effects were observed. Even if attentional processing is a general property of a cue, potentially able to influence learning about that cue with respect to any outcome, the large change in context might have caused participants to reset their attention to all cues to some baseline level, such that no effect of learned predictiveness from Stage 1 was observed. The smaller change in context between Stages 1 and 2 for the ALL–ALL and ENJ–ENJ groups might have been insufficient to cause this resetting of attention, so learned predictiveness effects were observed.

Experiment 2

Experiment 2 seeks to determine whether learned predictiveness effects in human causal learning are truly outcome specific or merely context specific. The design is shown in Table 2. Stage 1 was exactly as for Experiment 1, with the Stage 1 outcomes (Outcomes 1 and 2) being allergic reactions for all participants. Stage 2 was again similar to that of Experiment 1, with one major difference: For all participants, the Stage 2 filler compounds (EF, GH, IJ, and KL) were associated with the class of compounds opposite to those of the experimental compounds (AX, BY, CV, and DW). For participants in the ALLEXP (allergic reactions paired with experimental compounds) group, Outcomes 3 and 4, paired with the experimental compounds in Stage 2, were allergic reactions, whereas Outcomes 3′ and 4′, paired with the filler compounds in Stage 2, were enjoyment reactions. For participants in the ENJEXP (enjoyment reactions paired with experimental compounds) group, this was reversed—Outcomes 3 and 4 were enjoyment reactions, whereas Outcomes 3′ and 4′ were allergic reactions.

This manipulation ensured that all participants experienced the same context change between Stages 1 and 2: For both groups, Stage 1 involved allergic reactions only, whereas Stage 2 involved both allergic reactions and enjoyment reactions. If differences in context change were responsible for the differences in learned predictiveness effects in Experiment 1, we would not expect to see any differences in learned predictiveness effects in the two groups of Experiment 2. If learned predictiveness effects in human causal learning were outcome specific rather than context specific, we

Table 2
Design of Experiment 2

Stage 1	Stage 2	Test
AV → 1	AX → 3	AC
BV → 2	BY → 4	BD
AW → 1	CV → 3	VX
BW → 2	DW → 4	WY
CX → 2	EF → 3′	
DX → 1	GH → 4′	
CY → 2	IJ → 3′	
DY → 1	KL → 4′	

Note. Letters *A* through *Y* represent different foods. For all participants, Numbers *1* and *2* represent different types of allergic reactions. For the ALLEXP group, Numbers *3* and *4* represent types of allergic reactions, whereas *3′* and *4′* represent types of enjoyment reactions. For the ENJEXP group, Numbers *3* and *4* represent types of enjoyment reactions, whereas 3′ and 4′ represent types of allergic reactions. ALLEXP = this group experienced allergic reactions paired with the experimental compounds in Stage 2; ENJEXP = this group experienced enjoyment reactions paired with the experimental compounds in stage 2.

would expect to see a learned predictiveness effect in the ALLEXP group but not in the ENJEXP group.

An unexpected aspect of the results of Experiment 1 is the weak nature of the learned predictiveness effect observed in the ALL–ALL group. This reached only borderline one-tailed significance, despite the inclusion of four times the number of participants as Le Pelley and McLaren's (2003) original study, which found a statistically more significant effect with an effectively identical design. Experiment 2 provides us with another opportunity to test the reliability of this basic learned predictiveness effect.

Method

Participants, apparatus, and stimuli. Forty volunteers aged 18–37 years took part in the experiment in return for a small honorarium payment. Participants were randomly assigned to groups, with 19 in the ALLEXP group and 21 in the ENJEXP group. The apparatus and food stimuli were as for Experiment 1.

For both groups, Outcomes 1 and 2 were itch and sweating, counterbalanced across participants. For the ALLEXP group, Outcomes 3 and 4 were nausea and dizziness, whereas Outcomes 3′ and 4′ were thumbs up and broad smile. Assignment of specific reactions for each of these pairs of outcomes was independently counterbalanced across participants. For the ENJEXP group, Outcomes 3 and 4 were thumbs up and broad smile and Outcomes 3′ and 4′ were nausea and dizziness (again, each independently counterbalanced across participants).

Binary counterbalancing of Stage 1 outcomes, Stage 2 experimental compound outcomes, and Stage 2 filler compound outcomes yielded eight counterbalance conditions for each group. Given the exclusion of participants, as detailed in the *Results and Discussion* section, which left 16 in each group, this counterbalancing was complete, with 2 participants per group in each counterbalance condition.

Procedure. Stage 1 of Experiment 2 was as for Experiment 1. At the outset of Stage 2, participants received further instructions on screen. Sections of these instructions referred specifically to allergic reactions (marked Sections A1 and A2 below); other sections referred specifically to enjoyment reactions (marked Sections B1 and B2). The ordering of these sections was randomly determined for each participant. That is, some participants received Section A1 before Section B1 and Section A2 before Section B2, whereas other participants received Section B1 before Section A1 and Section B2 before Section A2.

> Treatment of Mr. X is now finished. In the next stage, you will be studying a new patient, Mr. Y.
>
> Once again, in an attempt to discover the effects of various foods on Mr. Y, you arrange for him to eat a number of different meals, each containing two foods. Some of the foods are the same as those given to Mr. X, some are not. Again you observe the effects of these foods on Mr. Y from behind a one-way mirror.
>
> The foods contained in some of these meals cause Mr. Y to [Section A1] suffer one of two different types of allergic reaction. After some of these meals Mr. Y is seen to suffer from dizziness. After others he is seen to feel nauseous. [End Section A1]
>
> The foods contained in other meals cause Mr. Y to [Section B1] display one of two different types of favorable reaction. After some of these meals Mr. Y is seen to give a broad smile, after others he is seen to give a thumbs up. [End Section B1]
>
> On the following screens, you will be shown the contents of meals eaten by Mr. Y, and will be asked to predict what kind of reaction will result from eating each meal. [Section A2] For meals that cause allergic reactions you will have a choice of two allergic reactions, "Dizziness" and "Nausea." [End Section A2; Section B2] For meals that cause favorable reactions you will have a choice of two favorable reactions, "Broad Smile" and "Thumbs Up." [End Section B2] As before, make your prediction by clicking the option button next to one of them, and then click OK.

The remainder of Stage 2 was the same as for Stage 1, with the exception that on some trials participants were given a choice of two allergic reactions, whereas on other trials they were given a choice of two enjoyment reactions.

The test stage was as for Experiment 1, with two exceptions. First, only Compounds AC, BD, VX, and WY were presented on test—compounds made up of filler cues were not tested. Second, the order of presentation of the two allergy scales was randomly determined for each participant (as opposed to being counterbalanced over participants in Experiment 1) but remained consistent for all test trials.

Results and Discussion

As in Experiment 1, we could only expect to see learned predictiveness effects if participants were able to learn the contingencies of Stages 1 and 2. We therefore imposed a selection criterion of 50% correct or more over all of the trials of Stage 1 (which was the same for both groups) and over the trials containing experimental compounds in Stage 2. Three participants in the ALLEXP group and 5 in the ENJEXP group failed to meet these selection criteria; their data were excluded from all further analyses. This left 16 participants in each group. Selection used only a subset of trials in Stage 2 in case of differences in ease of learning about the different outcome types. Suppose that it is considerably easier to learn about allergic reactions than about enjoyment reactions. The ALLEXP group would therefore learn about the experimental compounds in Stage 2 better than would the ENJEXP group, even though both groups show similar levels of learning over all Stage 2 trials (as the ENJEXP group would learn better about the filler compounds). Given that good learning of the experimental compounds is important for learned predictiveness effects to be seen, applying a learning criterion over all trials could induce a bias toward seeing such effects in one group and not the other. Taking a criterion over experimental compounds only ensures that learning about these compounds is similar in both groups—following the exclusions we have described, the mean percentage correct on experimental compounds over Stage 2 was

67.6% for the ALLEXP group and 64.1% for the ENJEXP group ($t < 1$).

Figure 3 shows mean difference scores for the test compounds (AC, BD, VX, and WY) for the ALLEXP and ENJEXP groups. For consistency with the analysis of Experiment 1 and because of the skew in some comparisons of difference scores produced by the smaller numbers of participants in Experiment 2, nonparametric statistics were used for all analyses.

The ALLEXP group demonstrated significant discrimination between Compounds AC and BD, with AC eliciting higher difference scores than BD, $T(14) = 19.5$, $p < .05$. There was no reliable discrimination, however, between Compounds VX and WY, $T(12) = 35.5$, $p > .10$. A learned predictiveness effect was observed, in that the discrimination (difference in difference scores) between AC and BD was significantly greater than that between VX and WY, $T(13) = 14.0$, $p < .05$.

The ENJEXP group demonstrated significant discrimination between Compounds AC and BD, $T(13) = 10.5$, $p < .02$, and between Compounds VX and WY, $T(14) = 14.0$, $p < .02$. No significant learned predictiveness effect was observed: Discrimination between Compounds AC and BD did not differ reliably from that between Compounds VX and WY, $T(13) = 40.5$, $p > .10$.

Most important, the size of the learned predictiveness effect differed reliably in the two groups: The difference in discrimination between good predictor compounds (AC and BD) and poor predictor compounds (VX and WY) was significantly greater in the ALLEXP group than in the ENJEXP group, $U(16, 16) = 62.5$, $p < .02$. Thus, a difference in learned predictiveness was observed despite the fact that both groups experienced the same change in context between Stages 1 and 2, which rules out the suggestion that these effects reflect the operation of a general attentional mechanism that is context specific. Instead, this finding is consistent with an approach based on associability parameters that are specific to a particular class of outcomes.

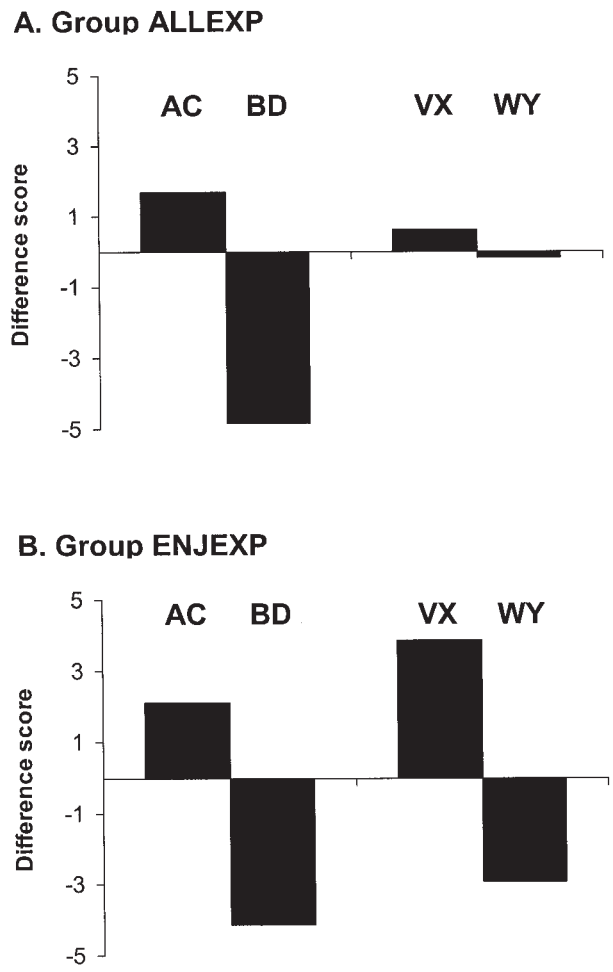

Figure 3. Difference scores for Compounds AC, BD, VX, and WY for the two groups of Experiment 2: the ALLEXP group (Panel A) and the ENJEXP group (Panel B). We calculated difference scores for each compound by subtracting the causal judgment rating for Outcome 4 from that for Outcome 3. A learned predictiveness effect is revealed by better discrimination (i.e., larger difference in difference scores) between Compounds AC and BD than between Compounds VX and WY. ALLEXP = this group experienced allergic reactions paired with the experimental compounds in Stage 2; ENJEXP = this group experienced enjoyment reactions paired with the experimental compounds in Stage 2.

General Discussion

Two experiments investigated the outcome specificity of a previously demonstrated learned predictiveness effect in human causal learning (Le Pelley & McLaren, 2003). In Experiment 1, learned predictiveness effects were demonstrated when the (qualitatively different) outcomes used in the two stages of the experiment were drawn from the same affective class but not when these outcomes came from different affective classes. However, in this study, changes in outcome class were confounded with changes in experimental context between Stages 1 and 2. Experiment 2 controlled for this latter factor by ensuring that all participants experienced the same change in context. Even under these conditions, learned predictiveness effects were influenced by the class of outcomes paired with the experimental cues in Stage 2.

These results rule out an interpretation of the original learned predictiveness effect in terms of a general attentional mechanism, as such a mechanism would modulate learning about a cue with respect to any presented outcome. Instead, the results support a view based on an associability mechanism that is (partially) reinforcer specific, with the associability developed by a cue with respect to a certain outcome able to modulate learning about that cue with respect to a different outcome from the same affective class but not an outcome from a different affective class. This conclusion is consistent with the pattern of results obtained from studies of animal conditioning (Dickinson & Mackintosh, 1979; Holland, 1988; see also Mackintosh, 1973). As such, the current data support Dickinson et al.'s (1984) claim that, under some circumstances at least, animal conditioning and human causal learning reflect the operation of a common, associative mechanism.

Although we have focused on the ability of Mackintosh's (1975) associability model to account for the results of the current experiments, this is not the only model of learned predictiveness effects in associative learning. For example, Pearce and Hall (1980) proposed another highly influential account that takes a quite different view of how associability operates. Their model states

that associability remains high for stimuli that are followed by surprising events (poor predictors) and declines for stimuli that are followed by expected events (good predictors). This model is also able to account for animal demonstrations of unblocking in terms of the surprise generated by addition or omission of US2 maintaining the associability of the added CS on compound trials. In addition, just as with the Mackintosh model, it is possible to modify the Pearce–Hall account to render associability specific to a particular class of reinforcer. However, the Pearce–Hall model is unable to account for the basic learned predictiveness effect in human causal learning demonstrated in this article and by Le Pelley and McLaren (2003). In the Pearce–Hall model, the associability of each cue of a compound is determined by the discrepancy between the magnitude of the US occurring on a trial and the extent to which the magnitude of that US is predicted by the compound (rather than by each cue individually). As such, it is the predictiveness of the compound that determines the associability of the elements of that compound. In the current experiments, however, all Stage 1 compounds were equally predictive of their respective outcomes, such that, according to the Pearce–Hall model, all cues have equal (low) associability at the end of Stage 1 and, hence, no differential learned predictiveness effect was expected in Stage 2. Consequently, we favor the approach to associability processes taken by the Mackintosh model as an account of the results of the current experiments. We also note that this model, unlike the Pearce–Hall model, is able to account for the results of a number of studies demonstrating learned predictiveness effects in human categorization and discrimination learning (e.g., Kruschke, 1996; Whitney & White, 1993; Wolff, 1967; Zeaman & House, 1974).

The possibility should be raised, however, that the effects of prior learning on later acquisition observed in the current experiments may reflect not the operation of associability processes but, instead, the influence of proactive interference, with memory of Stage 1 cue–outcome relations interfering with learning about those same cues and different outcomes in Stage 2. The explanation of why an effect is observed when the outcomes in Stages 1 and 2 are similar but not when they are very different falls out naturally from this position: Interference depends on the similarity between events (McGeoch & McDonald, 1931), and the events are more similar when the outcomes in the two stages are similar. The problem with this explanation is that it is hard to see, given the design of the experiment, how proactive interference could produce the necessary effect in the first place. The good predictors from Stage 1 doubtless have stronger associations to their respective outcomes than do the poor predictors, but this, in itself, tends to interfere with later learning about these good predictors rather than facilitate it. One possible way around this is to invoke a type of acquired distinctiveness effect, such that a cue that is strongly associated to Outcome 1 can then evoke the representation of Outcome 1 and associate that representation with the new outcome (see, e.g., Hall, Mitchell, Graham, & Lavis, 2003). However, training both AX and CV to Outcome 3 (and both BY and DW to Outcome 4) renders implausible any explanation of these results in terms of acquired distinctiveness. During Stage 2, AX tends to evoke a representation of Outcome 1 as a result of Stage 1 training, whereas CV tends to evoke Outcome 2. Therefore, neither of the Stage 1 outcomes is predictive of the current Outcome 3. Training of BY and DW with Outcome 4 in Stage 2 further ensures that Stage 1 Outcomes 1 and 2 are equally associated with Stage 2 Outcomes 3 and 4, such that acquired distinctiveness cannot exert a selective influence on learning of Stage 2 contingencies. The use of the current design thus ensures that any acquisition differences during Stage 2 reflect differences in cue processing.

If we grant that an associability account of our results along the lines suggested by the Mackintosh (1975) model is appropriate, it is tempting to draw further conclusions regarding the mechanisms of associability change from the current data. Specifically, the results address the issue of whether learned predictiveness effects are driven by an increase in the associability of good predictors from some starting value, a decrease in the associability of poor predictors, or both. In other words, did we observe better learning of the AC–BD discrimination relative to the VX–WY discrimination because participants were faster to learn about Cues A–D during Stage 2 than they would otherwise have been or because they were slower to learn about V–Y? Looking at the results of Experiment 1 (see Figure 2), we see that the magnitude of the discrimination between AC and BD was very similar in the *same* condition (in which a learned predictiveness effect was observed) and the *different* condition (in which no such effect was seen), $U(64, 64) = 1{,}974.0$, $p > .50$. Likewise, in the results of Experiment 2, discrimination between AC and BD was similar in the ALLEXP (learned predictiveness effect observed) and ENJEXP (no effect) groups, $U(16, 16) = 118$, $p > .50$. If the learned predictiveness effect resulted from an increase in processing of Cues A–D during Stage 1, speeding learning about these cues during Stage 2, then we would expect to see greater discrimination between AC and BD in conditions in which this learned predictiveness effect was observed. The fact that this was not found therefore indicates that the associability of good predictors does not rise much above a starting value. In contrast, the discrimination between VX and WY in the *same* condition of Experiment 1 was significantly worse than in the *different* condition, $U(64, 64) = 1{,}485$, $p < .01$. Likewise, the difference in discrimination between VX and WY in the ALLEXP and ENJEXP groups approached significance, $U(16, 16) = 82$, $p = .08$. This is the pattern expected if the processing power devoted to Cues V–Y had decreased as a result of experience of their nonpredictive nature during Stage 1, slowing learning about these cues during Stage 2. Overall, this pattern of results is consistent with a view in which the associability of good predictors is maintained at a high starting value but the associability of poor predictors decreases from this starting value.

If we are to explain our results in terms of an associability process, how should such an explanation be implemented? One approach is to postulate that there are two classes of outcome in the world, appetitive and aversive, and separate associabilities for any stimulus to each class of outcome. According to this view, associability is a property of the stimulus for that class of outcome. This fits the facts as known at present but lacks elegance. Another solution is to say that associability generalizes across outcomes. Thus, there is little transfer between dissimilar outcomes (e.g., appetitive to aversive) but much more between, for instance, two different types of allergic reaction. Further research is needed to differentiate between these two possibilities.

The results presented here replicate Le Pelley and McLaren's (2003) demonstration of learned predictiveness effects in human causal learning, in line with the predictions of the Mackintosh

(1975) model of animal conditioning. The current experiments extend that earlier work by demonstrating that this learned predictiveness effect reflects not a general attentional process but rather a (partially) outcome-specific associability mechanism. Moreover, the outcome specificity of the learned predictiveness effect in human causal learning exactly mirrors that observed in earlier studies of animal conditioning. As such, the current results provide further support for the idea that common mechanisms underlie both human causal learning and animal conditioning.

References

Dickinson, A., Hall, G., & Mackintosh, N. J. (1976). Surprise and the attenuation of blocking. *Journal of Experimental Psychology: Animal Behavior Processes, 2,* 313–322.

Dickinson, A., & Mackintosh, N. J. (1979). Reinforcer specificity in the enhancement of conditioning by posttrial surprise. *Journal of Experimental Psychology: Animal Behavior Processes, 5,* 162–177.

Dickinson, A., Shanks, D. R., & Evenden, J. L. (1984). Judgment of act-outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology, 36A,* 29–50.

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

Holland, P. C. (1984). Unblocking in Pavlovian appetitive conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 476–497.

Holland, P. C. (1988). Excitation and inhibition in unblocking. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 261–279.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science, 8,* 225–247.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review, 7,* 636–645.

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

Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *Quarterly Journal of Experimental Psychology, 56B,* 68–79.

Le Pelley, M. E., Oakeshott, S. M., & McLaren, I. P. L. (2005). Blocking and unblocking in human causal learning. *Journal of Experimental Psychology: Animal Behavior Processes, 31,* 56–70.

Lochmann, T., & Wills, A. J. (2003). Predictive history in an allergy prediction task. In F. Schmalhofer, R. M. Young, & G. Katz (Eds.), *Proceedings of EuroCogSci 03* (pp. 217–222). Hillsdale, NJ: Erlbaum.

Mackintosh, N. J. (1973). Stimulus selection: Learning to ignore stimuli that predict no change in reinforcement. In R. A. Hinde & J. S. Hinde (Eds.), *Constraints on learning* (pp. 75–96). New York: Academic Press.

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

McGeoch, J. A., & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *American Journal of Psychology, 43,* 579–588.

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

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

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning.* New York: Academic Press.

Whitney, L., & White, K. G. (1993). Dimensional shift and the transfer of attention. *Quarterly Journal of Experimental Psychology, 46B,* 225–252.

Wolff, J. L. (1967). Concept-shift and discrimination-reversal learning in humans. *Psychological Bulletin, 68,* 369–408.

Zeaman, D., & House, B. J. (1974). Interpretations of development trends in discriminative transfer. In A. D. Pick (Ed.), *Minnesota Symposia on Child Psychology* (Vol. 8, pp. 144–186). Minneapolis: University of Minnesota Press.

Received August 1, 2004
Revision received November 16, 2004
Accepted December 2, 2004 ■