

Overshadowing, but not relative validity, between the elements of an outcome during human associative learning

Martyn Quigley^{a,1}, Mark Haselgrove^{b,*}

^a Swansea University, UK

^b The University of Nottingham, UK

ARTICLE INFO

Keywords:

Associative learning
Competition
Overshadowing
Relative validity
Outcome

ABSTRACT

Overshadowing and relative validity constitute two phenomena that inspired the development of the Rescorla-Wagner model in 1972. They demonstrate that cues will interact with one another for an association with the presence or absence of an outcome. Here, three experiments sought to explore whether these two effects extended to outcomes using a food allergist paradigm with human participants. In Experiment 1 (overshadowing) participants received trials in which a cue was followed by a compound of two outcomes (A-O1O2). Test trials revealed that participants learned less about the A-O2 association than they did between a control cue C, which had been paired with O2 in isolation (C-O2) in training – thus demonstrating an outcome overshadowing effect. In Experiment 2 (relative validity) participants received true discrimination trials, in which A was paired with an O1O3 compound and B was paired with an O2O3 compound, and pseudo discrimination trials, in which C and D were paired on 50% of the trials with an O4O6 compound and on the remaining trials with an O5O6 compound. Consequently, O3 is less well predicted by A and B relative to O1 and O2, whereas O6 is equally well predicted by C and D relative to O4 and O5. Despite the relative validity of A and B for O3 being less than the relative validity of C and D for O6, the ratings of A and B for O3 were the same as C and D for O6. This failure to observe an outcome relative validity effect was reproduced in Experiment 3, which replicated Experiment 2, but with an adjustment made to the number of training trials given to participants. These results are discussed in terms of a real-time development of the Rescorla-Wagner model provided by Wagner (1981).

1. Introduction

In their influential theory, Rescorla and Wagner (1972) proposed that variations in the effectiveness of reinforcement and nonreinforcement could be used to understand the circumstances under which Pavlovian conditioning occurs. They suggested that the strength of the association between a conditioned stimulus (CS, or cue) and a reinforcer (or outcome) is updated on each trial as a function of, what would now be called, reinforcement prediction error - the difference between the asymptote of learning supported by the reinforcer (or outcome) and the current associative strength of all cues present on that trial. The model was, and continues to be, extremely successful. It provided a simple explanation for existing cue-competition phenomena such as overshadowing (e.g., Pavlov, 1927), blocking (e.g., Kamin, 1968) and relative validity (e.g., Wagner, Logan, Haberlandt, & Price, 1968) and correctly predicted new phenomena such as inhibition from over-

expectation (Kremer, 1978). The theory also regularly serves as the starting point for many other, more complex, theories of conditioning and associative learning (e.g., Esber & Haselgrove; 2011; Le Pelley, 2004), and has greatly assisted our understanding of the role of dopamine neurons in the primate midbrain (Waelti, Dickinson & Schultz, 2001), as well as our understanding of learning in humans (Shanks, 1995). Given the centrality of the role of reinforcement in the theory, the Rescorla-Wagner theory has, perhaps surprisingly, a relatively simple conception of reinforcement representation. Reinforcers (or outcomes) are either present on a trial (in which case their value is 1 in the algorithm of the theory), or they are absent (in which case their value is 0), and their effectiveness for driving learning is modulated by the value of the prediction error. This contrasts with the way in which the theory represents CSs, or cues. Here, rather than stimuli being just present or absent, CSs are decomposed into individual elements, each of which may compete for an association with the outcome to generate effects

* Corresponding author at: School of Psychology, University of Nottingham, University Park, Nottingham NG7 2RD, UK.

E-mail addresses: m.c.quigley@swansea.ac.uk (M. Quigley), mark.haselgrove@nottingham.ac.uk (M. Haselgrove).

¹ School of Psychology, Vivian Tower, Swansea University, Swansea SA2 8PP, UK.

such as blocking, overshadowing and relative validity. The question we ask here is do the elements of an outcome behave similarly? Do they interact with each another to differentially associate with a preceding cue?

1.1. Outcome blocking (and facilitation)

First reported by Kamin (1968), blocking remains one of the most influential cue-competition phenomena within the study of conditioning and learning. In a typical blocking experiment, a cue is paired with an outcome in stage 1 (A+) before being accompanied by a novel cue in stage 2, whilst continuing to be paired with the same outcome (AB+). In a final test, responding to the novel cue B is found to be weaker than in a control group who, for example, had the stage 1 training with A omitted, or included stage 1 training with A being paired with an alternative outcome. Blocking has been demonstrated using a variety of different learning procedures. For example, blocking has been observed in rats using appetitive and aversive Pavlovian conditioning (e.g.: Jones & Haselgrove, 2013; Kamin, 1968), taste-preference and taste-aversion learning (e.g.: Dwyer, Haselgrove & Jones, 2011; Willner, 1978), and spatial learning (e.g., Pearce, Graham, Good, Jones, & McGregor, 2006; Rodrigo, Chamizo, McLaren, & Mackintosh, 1997). Furthermore, there is evidence of blocking in species as diverse as honeybees (e.g., Blaser, Couvillon, & Bitterman, 2006), goldfish (Tennant & Bitterman, 1975), and humans (Dickinson, Shanks, & Evenden, 1984; Le Pelley, Oakeshott & McLaren, 2005).

Across all the studies just described, the locus of blocking has been between the elements of cues – element A of a cue blocks learning about element B. The question of whether blocking is observed between the elements of an outcome has been investigated much less frequently, however. That said, Rescorla (1980) did explore this question using a second-order autoshaping procedure with pigeons in which the presentation of visual cues was reinforced by the delivery of food. In an experimental group, a cue was paired on some trials with a single outcome (A – O1), while on other trials the same cue was paired with a compound of two outcomes (A – O1O2). For a control group, only the outcome compound trials (A – O1O2) were presented. The results from a test stage revealed that the control group, who had no exposure to A – O1, demonstrated greater learning about the association between A – O2 than the experimental group. These results therefore indicate an ‘outcome blocking’ effect - learning about the A-O1 association attenuated the formation of the A-O2 association. Miller and Matute (1998) have also provided a demonstration of “outcome blocking” in a series of sensory preconditioning experiments with rats.

“Outcome blocking” has also been investigated in human participants, however, these studies have produced more heterogeneous results. Cobos et al. (2002) reported the absence of an outcome interaction effect when examining the impact of causal structure (i.e., whether causes or effects are presented first) and cue structure (whether multiple cues are presented prior to a single event or whether a single cue is followed by multiple events) in a causal judgement task. Flach et al. (2006), however, reported outcome interaction effects when employing a response priming task (Elsner & Hommel, 2001). In response priming, participants are first asked to make two responses, each of which leads to a different visual stimulus (i.e., stage 1). Subsequently, participants are presented with the visual stimuli from stage 1 and asked to make a specified response as quickly as possible (i.e., a test stage). Response times during the test stage are typically faster when the stimulus–response mappings are congruent with the mappings from the training stage relative to when they are incongruent. In their study, Flach et al. included a stage 2 where an auditory stimulus was presented alongside the visual stimulus, thus forming an outcome compound. At test participants provided longer response times to the auditory element of the outcome compound relative to a control group that did not receive stage 1 training. Thus, demonstrating an outcome blocking effect comparable to that reported in animals (e.g., Miller and Matute, 1998;

Rescorla, 1980). Flach et al. also reported an additional experiment, where the auditory stimuli were presented in stage 1, prior to being paired in compound with visual stimuli in stage 2. At test, the opposite effect was observed, that is, an *outcome facilitation effect* occurred where pretraining with an element of an outcome compound produced quicker response times to the added visual stimulus at test.

Experiments conducted in our laboratory (Quigley & Haselgrove, 2020) have revealed a similar heterogeneity of results. In a series of three allergy-prediction experiments, we consistently observed, like Flach et al. (2006), an outcome facilitation effect. That is, learning in stage 1 about an element of an outcome (A–O1) enhanced learning about a novel outcome (O2) when these outcomes were presented as an outcome compound in stage 2 (A–O1O2). This effect was observed relative to a control stimulus that in stage 1 received B–O3 trials prior to B–O1O2 trials in stage 2. Interestingly, however, we were also able to observe an outcome blocking effect with a subtle procedural modification in which participants were presented with an additional set of control trials (C–O1O2) during Stage 2 (that had not been presented in stage 1). That is, participants displayed more learning about the C–O2 association than about the A–O2 association, thus displaying an outcome blocking effect (A < C) alongside an outcome facilitation effect (A > B).

1.2. Outcome overshadowing and relative validity

Two other cue-competition effects have been instrumental in shaping our understanding of conditioning and associative learning that, along with blocking, motivated the development of the Rescorla-Wagner model. These are overshadowing and relative validity. In overshadowing, a single cue (e.g., A) and a cue compound (e.g., BC) are both paired with an outcome during a single stage of training. Although both the single cue and the compound each reliably predict an outcome, the single cue will typically enter into a stronger association with an outcome than either one of the elements of the cue compound (unilateral overshadowing; see Mackintosh, 1976), or both elements of the cue compound (reciprocal overshadowing, see Sánchez-Moreno et al., 1999). Overshadowing is a robust property of associative learning and has been demonstrated in rats in studies of conditioned suppression (Kamin, 1968; Mackintosh, 1971), flavor-aversion and flavor-preference learning (Dwyer, Haselgrove, & Jones, 2011; Revusky, 1971), appetitive conditioning (Holland, 1999) and spatial learning (Pearce, Graham, Good, Jones, & McGregor, 2006). It has also been demonstrated in a diverse range of species from goldfish (Tennant & Bitterman, 1975) to humans (Chamizo, Aznar-Casanova, & Artigas, 2003).

In relative validity, training trials with just cue compounds are presented to subjects/participants. For some of the trials in this task, referred to as the true-discrimination (TD) trials, the compounds predict the presence (AX+) or absence (BX-) of an outcome, resulting in the common element (X) being paired with the outcome 50% of the time. For other trials, referred to as pseudo-discrimination (PD) trials, there is no distinctive compound, nor element, which reliably predicts the presence or absence of the outcome, that is, two cue compounds are paired with the outcome 50% of the time (CY-/+, DY-/+). What is typically observed when employing these tasks, is better learning about the association between the common element and the outcome following the pseudo-discrimination trials (Y) than the true-discrimination trials (X), despite the fact that X and Y are paired with the presence and absence of the outcome equally frequently. Relative validity was first reported by Wagner et al. (1968) using appetitive instrumental and aversive Pavlovian conditioning in rats, as well as conditioned eyelid closure in rabbits, but it has since been observed in autoshaping in pigeons (Pearce, Esber, George & Haselgrove, 2008) as well as in judgments of causality in human participants (Van Hamme, Kao & Wasserman, 1993).

In keeping with the “outcome blocking” literature, much less research has explored overshadowing or relative validity in outcomes. Arcediano et al. (2005) employed an outcome overshadowing procedure

using a food allergist paradigm similar to that reported by Quigley and Haselgrove (2020). Here, one food was paired with a single allergy, or a food was paired with a compound of allergies. At test, participants provided higher ratings to the food paired with a single outcome than the food paired with the outcome compound, thus demonstrating an outcome overshadowing effect. Evidence for the presence of relative validity between the elements of an outcome is equally scant, only one such study, by Matute et al. (1996), appears to have directly tested whether relative validity can appear when multiple outcomes are presented following a single cue. In this study a medical decision-making task was employed in which participants were presented with a medicine and tasked with identifying the extent to which it caused three fictitious syndromes (O1, O2 and O3) presented in compounds of two. In a PD condition the medicine (A) was presented 50% of the time prior to each of two outcome compounds (O1 O3 and O2 O3). In a TD condition consumption of the medicine would reliably predict one outcome compound (O1 O3) but not the other outcome compound (O2 O3). At test, participants' ratings of the strength of the association between the medicine (A) and the common outcome (O3) were comparable in the PD and TD condition. That is, the absence of relative validity was observed among the elements of outcomes, although this appeared to be influenced by the nature of the test question. In this study, however, stimuli (i.e., cues and outcomes) were presented simultaneously in a list like manner, thus it is possible that participants treated the task as a diagnostic task and therefore treated the outcomes as cues.

Taken together the results of the above studies suggest that while interaction effects are not unique to cues, there is heterogeneity in the literature regarding their presence in outcomes. Furthermore, there is a particular scarcity of studies that have examined outcome overshadowing and outcome relative validity. Consequently, we sought to assess whether outcome interaction effects can be obtained using both outcome relative validity and outcome overshadowing procedures, employing a relatively standard allergy prediction task that has been employed elsewhere (e.g.: Le Pelley & McLaren, 2003; Lochmann & Wills, 2003), and which we have previously used to investigate outcome blocking/facilitation (Quigley & Haselgrove, 2020). If outcome overshadowing can be observed in a comparable manner to cues, it is anticipated that better learning would be displayed toward an outcome element, as opposed to the outcome element that has been presented in compound with an additional element. If outcome relative validity can be observed among outcomes in a manner comparable to cues, it is anticipated that the common outcome-element on the pseudo-discrimination trials would be learned about better than the common outcome-element on the true-discrimination trials.

2. Experiment 1

Experiment 1 employed a variant of the food-allergist task reported by Quigley and Haselgrove (2020) in which participants were asked to imagine themselves as a food allergist tasked with identifying which foods caused certain allergic reactions. In the study by Quigley and Haselgrove, this task was used to study blocking and facilitation between the elements of an outcome, in the current experiment, however, we use it to examine outcome overshadowing. Participants were presented with trials that comprised cues which reliably predicted a single outcome element, along with trials in which cues predicted a compound outcome. In keeping with the experiments reported by Quigley and Haselgrove both causal judgement and diagnostic judgement tasks were employed in different groups. As such, participants were either informed about the foods which a fictitious patient ('Mr. X') had consumed and then presented with information about the reaction the patient experienced (the causal task), or they were presented with the reaction a patient experienced and then informed about the foods the patient had eaten (the diagnostic task).

The design of Experiment 1 can be seen in Table 1. As both causal and diagnostic versions of the task were used, the stimuli which served as

Table 1
Design of Experiment 1.

	Training	Test
Overshadowing	A – O1 O2 B – O3 O4	A: O1 – O4 B: O1 – O4
Control	C – O2 D – O4	C: O1 – O4 D: O1 – O4

Note. A – D refers to cues, O1 – O4 refer to outcomes. A and B trials represent the overshadowing trials, while C and D trials represent the control trials.

cues (i.e., A-D), were represented by foods in the causal version of the task, and reactions in the diagnostic version of the task. Correspondingly, then, the stimuli which served as outcomes (i.e., O1-O4) were reactions in the causal task and foods in the diagnostic task. In stage 1, four cues (A – D) were presented alongside four outcomes (O1 – O4), as outlined in Table 1 and participants were tasked with learning the relationships between these stimuli. Two of the cues reliably predicted outcome compounds, respectively A – O1O2 and B – O3O4, while two cues predicted outcome elements, C – O2 and D – O4. Trials with A and B represented the outcome overshadowing trials as outcome compounds were presented on these trials, while trials with C and D served as control trials as single outcomes were presented on these trials. During a subsequent test stage, participants were presented with a screen which required them to provide ratings concerning how predictive the cues (A – D) were of each of the outcomes. If an outcome overshadowing effect is present, then participants' ratings of the A – O2 and B – O4 relationships would be lower than their ratings of the C – O2 and D – O4 relationships.

2.1. Method

2.1.1. Participants

Thirty-two participants (21 females; 11 males) were recruited from the University of Nottingham's School of Psychology, with an equal number of participants in Group Causal ($n = 16$) and Group Diagnostic ($n = 16$). Participants ranged from 18 to 28 years of age ($M = 20.84$; $SEM = 0.56$) and had normal or corrected-to-normal vision. Participants received course credit for their participation or a £3 inconvenience allowance. The study received institutional ethical approval from the University of Nottingham's Psychology ethics committee.

2.1.2. Apparatus and stimuli

All stimuli were presented (and responses recorded) in the experimental software package PsychoPy2, v1.83.01 (see Peirce, 2007; Peirce, 2008), running on Windows 7 on a standard desktop computer (screen size: 27 cm × 46 cm; h × w). There were four pictures of foods: broccoli, cabbage, onion and potato. The reactions Mr. X experienced were diarrhoea, fever, skin rash and vomiting presented as text. Each of the foods and reactions were assigned to the letters and outcomes in Table 1 using a Latin-Squared counterbalancing technique. For the causal judgment version of the task, participants were presented with a food picture in the centre of the screen (9 mm high and 8 mm wide) against a grey background (see Fig. 1 top panel) and the reactions were presented beneath a white scale (length: 35 cm) in capitalised white Arial text (size: 32). As such, there were four options beneath the scale which participants could select from (e.g.: O2; O1 and O2; O4; O3 and O4). For the diagnostic version of the task participants were presented with the reaction in the centre of the screen and foods were presented beneath the white scale. Participants could select their choice by positioning a cursor on the scale above the outcome they wished to select. Participants could only select four points on the scale. The order of the outcomes beneath the scale was counterbalanced between participants.

At test each of the cues were presented individually (positioned on the left of the screen) and participants were asked to rate how predictive each cue was of each of the outcomes. Each outcome was positioned,

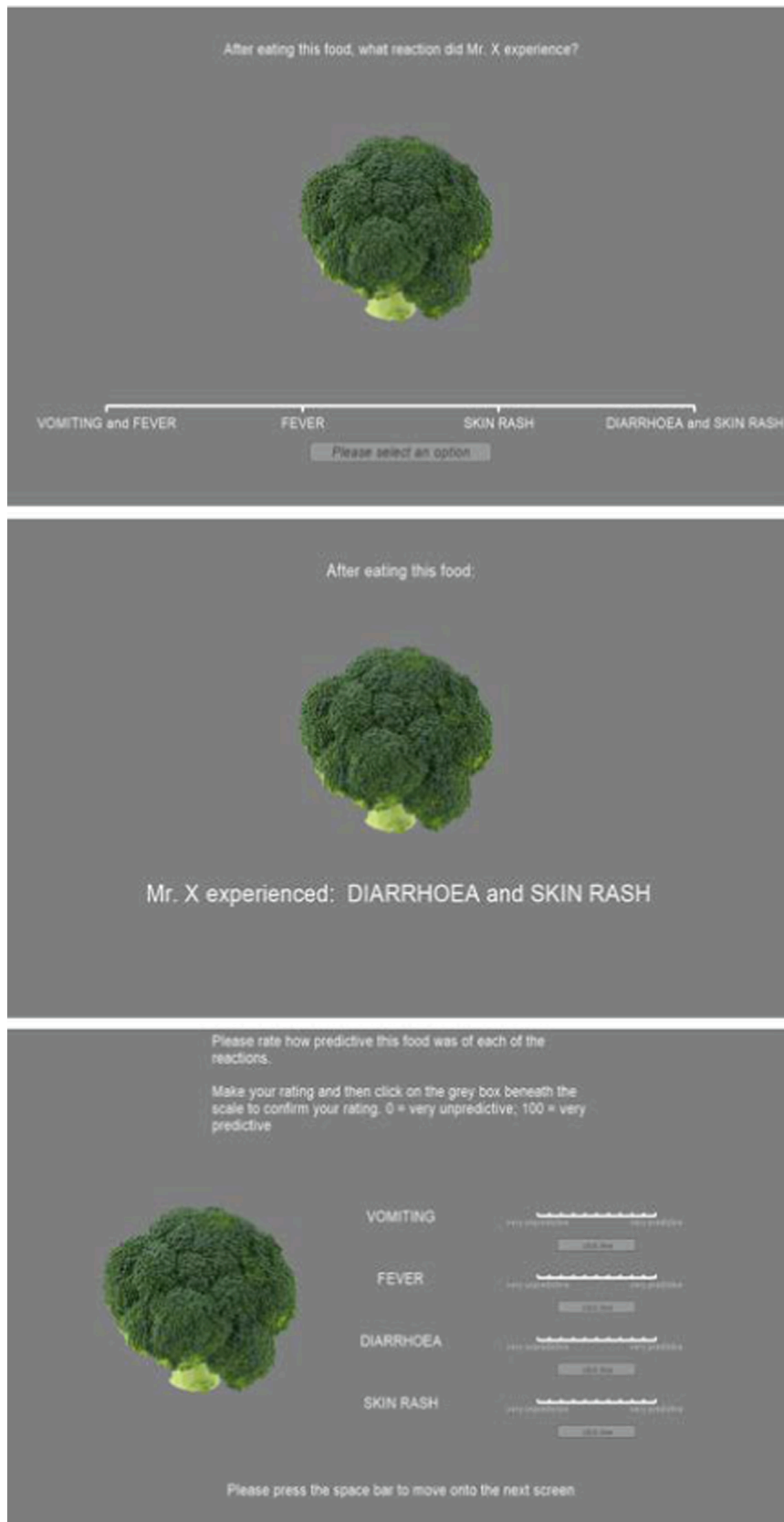


Fig. 1. Examples of each screen presented during Stage 1 (top), the response they received (middle) and during the test stage (bottom) for the causal version of the task.

alongside the cue, on the right of the screen (see Fig. 1). Participants made their ratings for each cue by moving a cursor on a Likert scale which ranged from 0 to 100 [‘0’ = ‘Very unpredictable’, ‘100’ = Very predictive’].

2.1.3. Procedure

Instructions.

All participants were tested in-person, individually in a small testing room. Following presentation of ethical and informed-consent information, participants were presented with the following text:

“In this experiment we would like you to imagine that you are an allergist (i.e., someone who investigates reactions to foods). You have just been presented with a new patient “Mr. X”, who suffers from different types of reactions as a result of eating certain foods. In an attempt to discover the relationship between the reactions Mr. X experiences and the different types of foods he has eaten, you observe the type of FOODS / REACTIONS² he has EATEN / EXPERIENCED and try and work out which REACTIONS / FOODS he had EXPERIENCED / EATEN. On the following screens, you will be shown the FOODS / REACTIONS Mr. X has EATEN / EXPERIENCED, and you will be asked to predict which REACTIONS / FOODS he had EXPERIENCED / EATEN. Each REACTION / FOOD will be presented at the bottom of the screen. Make your prediction by selecting one of the REACTIONS / FOODS at the bottom of the screen. You will then be provided with feedback about what REACTION / FOOD Mr. X had EXPERIENCED / EATEN. You will have to guess at first, but with the aid of the feedback your predictions should soon start to become more accurate.”

Training Stage.

Participants were exposed to 32 trials in total. Participants received 8 pairings of each cue – outcome pairing (A – O1O2, B – O3O4, C – O2 and D – O4). There were two blocks of trials with 4 presentations of each trial type in each block. Trial order was randomised within each block and there was no break between blocks. Participants in Group Causal were presented with the following text at the top of the screen: “Mr. X has eaten the following food, what did he experience?”. To move onto the next trial participants made a response by selecting a reaction/s (in word form) beneath the white scale. Participants in Group Diagnostic were presented with the following text at the top of the screen: “Mr. X has experienced the following reaction, what did he eat?”. To move onto the next trial participants made a response by selecting an image of a food beneath the white scale. Once participants had made a response, they were presented with feedback (Fig. 1, Centre Panel). Participants could take as much time as they needed when selecting the outcome. The feedback screen was presented for 1.5 s, before the next trial commenced automatically. Once participants had made their choice and received feedback on all trials, they moved onto the Test Stage.

2.2. Test stage

Prior to the test stage, participants were presented with the following text: “You will now be asked to rate how predictive each food/reaction was of each of the reactions/foods. Please press the SPACE BAR to continue”. Each of the foods/reactions were then presented at test, one per screen (see Fig. 1, Lower panel.). On each screen participants were presented with the text: “Please rate how predictive this food/reaction was of each of the reactions/foods [0 = very unpredictable; 100 = very predictive]. Make your rating and then press the spacebar to proceed”. Participants were then required to make a rating for each of the outcomes using the Likert scales positioned to the right of the reactions. No time limit was imposed for responding during the test stage. Once they

had rated each cue’s ability to predict each outcome, they were informed that the experiment was complete.

2.3. Results

2.3.1. Training data

The top and bottom panels of Fig. 2 show the mean proportion of correct responses for the cues (A-D) across the Training Stage of Experiment 1 for Groups Causal and Diagnostic respectively. These data have been collapsed into 2 blocks each containing 4 trials of each cue-outcome pairing (A - O1O2, B - O3O4, C - O2, D - O4) to be consistent with the training stages of the relative validity tasks described in Experiments 2 and 3. As can be seen, participants in both groups quickly learnt the relationships between the cues and the outcome compounds (i.e., A – O1O2 / B – O3O4) and the cues and the outcome elements (i.e., C – O2 / D – O4), demonstrating a high proportion of correct responses after Block 1 and reaching near asymptote on the second block.

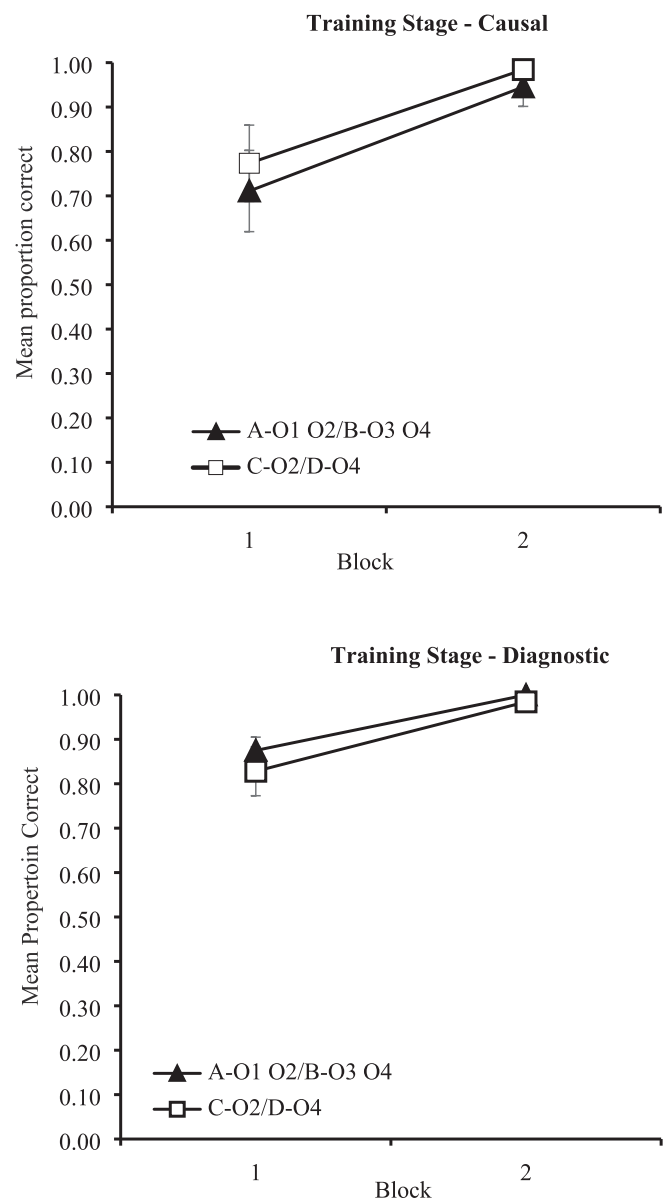


Fig. 2. Mean proportion of correct responses for A - O1O2/ B - O3O4 and C - O2 / D - O4 for Group Causal (Top panel) and Group Diagnostic (Bottom panel). Error bars represent 1 +/-SEM.

² The text either read FOOD or REACTION or EATEN or EXPERIENCED. The first option prior to the forward slash (/) refers to the causal version of the task, while the second options refers to the diagnostic version of the task.

Participants appeared to learn the associations quicker in Group Diagnostic than Group Causal.

A three-way mixed ANOVA was performed on these data with a between-subject factor of Group (Causal vs Diagnostic), and within-subject factors of Stimuli (A/B vs C/D) and Block (1 – 2). The ANOVA revealed a main effect of Block, $F(1, 30) = 89.17$, $MSE = 0.01$, $p < .001$, $\eta_p^2 = 0.75$, a main effect of Group, $F(1, 30) = 7.58$, $MSE = 0.02$, $p < .05$, $\eta_p^2 = 0.20$, and significant interactions between Block and Group, $F(1, 30) = 4.55$, $MSE = 0.01$, $p < .05$, $\eta_p^2 = 0.13$, and Group and Stimuli, $F(1, 30) = 10.22$, $MSE = 0.00$, $p < .01$, $\eta_p^2 = 0.25$. All other interactions were non-significant, largest $F(1, 30) = 1.37$, $MSE = 0.00$, $p = .25$, $\eta_p^2 = 0.04$. To explore the source of the interactions simple main effects were conducted. These analyses revealed that participants in Group Diagnostic had a higher mean proportion of correct responses after one block of training than those in Group Causal, $F(1, 30) = 6.84$, $MSE = 0.01$, $p < .05$, $\eta_p^2 = 0.19$. Furthermore, participants in Group Causal learned more about the relationship between the cues and the outcome elements, than the cues and the outcome compounds, $F(1, 30) = 9.64$, $MSE = 0.00$, $p < .05$, $\eta_p^2 = 0.39$.

2.3.2. Test data

Fig. 3 shows participants' outcome specific ratings at test. These ratings assess whether participants' ratings were higher or lower for the specific outcomes of interest (in this case, O2 and O4). To provide an outcome specific measure of participants' ratings for cues A – D and O2 and O4 a difference score was calculated. Here, the rating for the outcome which a cue had never been paired with – but was otherwise treated in an identical manner – was subtracted from the rating for the

outcome the cue had been paired with. For example, participants' ratings for A – O4 (which A had never been paired with, yet B had) were subtracted from their ratings for A – O2 (which A had been paired with). For the control stimulus C, participants' ratings for C – O4 were subtracted from their ratings for C – O2. As can be seen in Fig. 3, participants' ratings for A/B – O2/O4 were lower than their ratings for C/D – O2/O4. Participants in Group Diagnostic also provided slightly higher scores for both sets of stimuli compared to Group Causal.

A 2×2 mixed model ANOVA of mean ratings with a between-subjects factor of Group (Causal vs Diagnostic) and a within-subjects factor of Outcome (A/B – O2/O4 vs C/D – O2/O4) revealed a significant main effect of the factor Outcome, $F(1, 30) = 18.54$, $MSE = 619.77$, $p < .001$, $\eta_p^2 = 0.38$. However, the main effect of Group was non-significant, $F(1, 30) = 1.99$, $MSE = 1825.72$, $p = .17$, $\eta_p^2 = 0.06$. The interaction between these factors was also non-significant, $F(1, 30) = 0.00$, $MSE = 619.77$, $p = .99$, $\eta_p^2 = 0.00$. These results reveal that participants provided lower ratings to A/B – O2/O4 than C/D – O2/O4 regardless of task type, therefore demonstrating an outcome overshadowing effect.

2.3.3. Reciprocal vs unilateral overshadowing

In previous studies which have reported outcome overshadowing (Arcediano, Matute, Escobar & Miller, 2005) the nature of the outcome overshadowing effect was not discussed. That is to say, whether overshadowing occurred due to both elements of the compound being rated comparably low (i.e., reciprocal overshadowing) or due to only one element of the outcome compound being rated lower than the outcome element (i.e., unilateral overshadowing) was not considered. Yet the

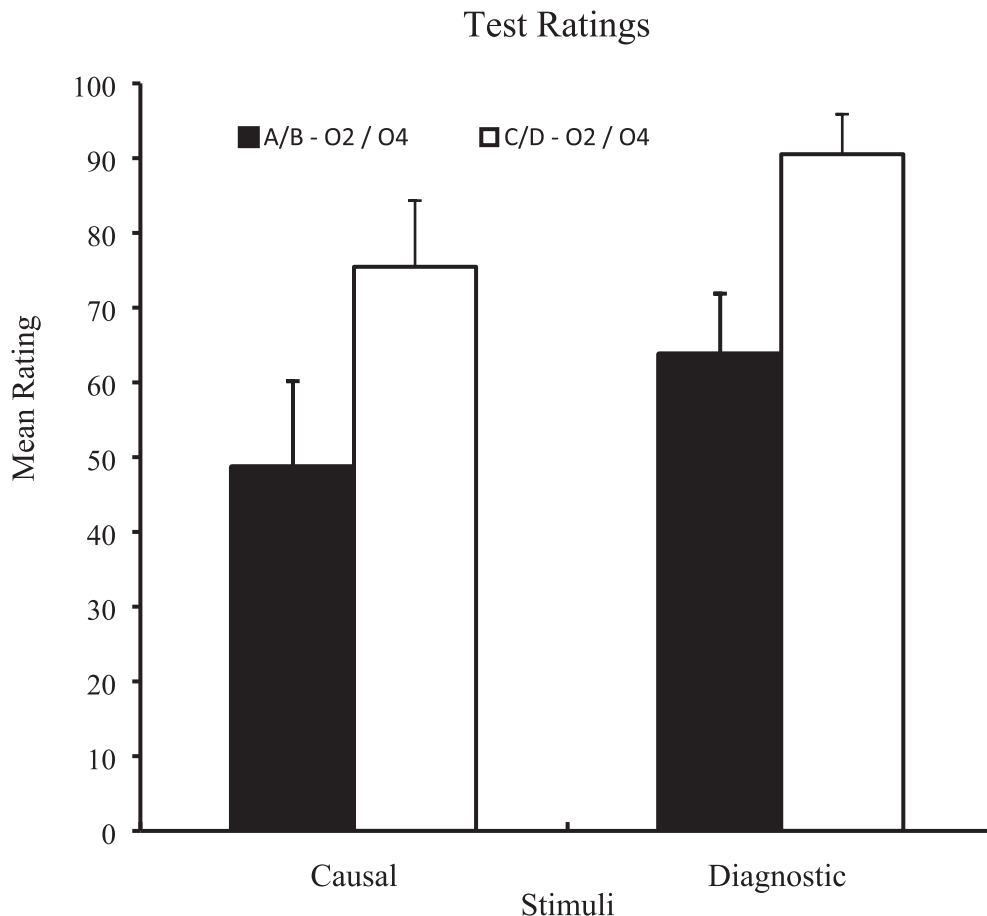


Fig. 3. Mean difference ratings for A/B – O2 / O4 and C/D – O2 / O4 during the test stage of Experiment 1 for groups causal and diagnostic. Error bars represent 1+/1 SEM.

nature of the overshadowing effect in cues has been a key test of predictions derived from theoretical models to account for this effect. For instance, according to the Rescorla and Wagner model (1972) overshadowing arises due to both stimulus elements receiving a comparable share of the limited associative strength that is supported by the US, thus overshadowing should be reciprocal in nature. Yet, according to the Mackintosh model (1975), the overshadowing effect reported in cues arises due to the extent to which a component of a compound assumes the role of being the best predictor of the outcome, which is determined by the salience or associability of the relevant stimuli. According to Mackintosh, the stimulus with the lower salience will be learned about less than the stimulus with the higher stimulus. However, the stimulus with the higher salience will be learned about to the same extent as the cue presented on its own (Mackintosh, 1976). This effect is referred to as unilateral overshadowing.

To determine the unilateral or reciprocal nature of the outcome overshadowing effect, further analysis was conducted on the ratings participants provided to cues A/B and the outcome elements of the outcome compounds O1O2 and O3O4. An analysis was performed on the test data collapsed across the variable of group (given the absence of a group effect) where the overshadowing outcome was defined, on a participant-by-participant basis, as the outcome element of the compound (i.e., O1O2 and O3O4) which received the highest rating, regardless of whether the outcome was also presented on its own on a control trial, whilst the overshadowed outcome was defined as the lower rated element of the compound. Bonferroni corrected t-tests (adjusted $p = .017$) revealed that the overshadowing outcome received a higher rating ($M = 72.16$, $SD = 34.37$) than the overshadowed outcome ($M = 49.61$, $SD = 43.26$; which would be expected as the highest rated outcome was always selected as the overshadowing outcome), $t(31) = 4.14$, $p < 0.001$. More interestingly, however, both elements of the outcome compound were rated lower than the control outcome ($M = 83.00$, $SD = 29.79$), smallest $t(31) = 2.58$, $p < 0.017$. These results therefore demonstrate a reciprocal overshadowing effect.

2.3.4. Discussion

Experiment 1 demonstrated that when employing an outcome overshadowing procedure, participants rated the outcomes which were presented on the overshadowing trials lower than the outcome elements which appeared on the control trials. These results provide evidence of an overshadowing effect which operates in a similar manner in both cues and outcomes. Further analysis concerning the nature of the outcome overshadowing effect also revealed that ratings for both elements of the outcome compound were lower than the outcome element presented on its own, thus demonstrating a reciprocal overshadowing effect. Additionally, the type of task participants completed appeared to have a mild influence upon their learning about these tasks in the training stage (as Group Diagnostic displayed better learning than Group causal during training) but not at test. Group Causal also learned about the outcome elements better across the course of training than the outcome compounds. One account of this effect is to simply suggest that the processing demands of learning about the additional outcome present in the outcome compound results in poorer learning. However, this would not account for why this effect was not present in the diagnostic task. Instead, the effect, in conjunction with the fact that participants displayed better overall learning in the training stage of the diagnostic task, would appear to support the idea that participants are able to better learn about multiple causes when the effects are presented first (in a diagnostic task), as is suggested by causal model theory (Waldmann & Holyoak, 1992). According to causal model theory when participants are first presented with the effect of a cause – as is the case in a diagnostic task – participants are more likely to consider that there are multiple possible causes as opposed to when completing a causal judgment task.

3. Experiment 2

Experiment 1 demonstrated that participants were more successful in learning the relationship between a cue and an outcome when the outcome comprised a single element rather than two elements. That is to say, we observed an outcome overshadowing effect. These results are consistent with the findings of Arcediano et al. (2005) and provide further evidence of an interaction effect amongst the elements of an outcome. Given the presence of outcome interaction effects in both blocking and overshadowing designs, these results could be taken as evidence to suggest that comparable effects may be observed under other circumstances – such as in tasks that translate the relative validity of the elements of cues (Wagner et al, 1968) to the relative validity of the elements of an outcome. As noted in the introduction, however, few previous studies have explored whether a relative validity effect can be observed amongst outcomes. Consequently, Experiment 2 sought to explore whether an outcome interaction effect can be observed when employing an outcome relative validity procedure with the same food allergist task as used in Experiment 1 and in Quigley and Haselgrove (2020). In keeping with these previous studies, the effect of the direction of causality was also explored with both causal judgement and diagnostic judgement tasks being employed. The design of Experiment 2 can be seen in Table 2.

During the training stage, four cues (A – D) were presented which were paired with four different outcome compounds (O1O3, O2O3, O4O6 and O5O6). Two of these cues, A and B, reliably predicted outcome compounds, O1O3 and O2O3 respectively. These stimuli represented the true-discrimination trials as one element of each outcome compound was only ever predicted by A (i.e., O1) or B (i.e., O2), while the other element of the compound (i.e., O3) was predicted by both. The remaining two cues, C and D, were presented as often as A and B, however, cues C and D were each paired with outcome compounds O4O6 and O5O6, half the time they were presented. Two features of this experimental design are worthy of note. First, A and B are paired with O3 as frequently as C and D are paired with O6. Second, and more importantly, O3 is less well predicted by A and B relative to O1 and O2. In contrast, O6 is equally well predicted by C and D relative to O4 and O5. Thus, the relative validity of cues A and B for O3 is less than the relative validity of cues C and D for O6.

In keeping with Experiment 1, participants were tasked with learning these relationships between the cues and outcomes across the course of the training stage of Experiment 2. At test, they were then presented with each of the cues (individually) and asked to make their ratings for each of the outcomes (O1 – O6). If a comparable outcome-interaction effect to that observed in the relative validity of cues is observed, then the common outcome which occurs on the pseudo-discrimination trials (O6) would receive a higher rating during cues C and D than the common outcome on the true-discrimination trials (O3) during cues A and B.

Table 2
Design of Experiment 2.

	Training	No. of trials	Test
True Discrimination	A – O1O3	24	
	B – O2O3	24	A: O1 - O6
Pseudo Discrimination	C – O4O6	12	B: O1 - O6
	C – O5O6	12	C: O1 - O6
	D – O4O6	12	D: O1 - O6
	D – O5O6	12	

Note. A – D refer to cues, O1 – O6 refer to outcomes. A and B represent the true-discrimination trials, while C and D trials represent the pseudo-discrimination trials.

3.1. Method

3.1.1. Participants

Forty-eight participants (37 females; 11 males) were recruited from

the University of Nottingham's School of Psychology. Participants ranged from 18 to 36 years of age ($M = 21.94$; $SEM = 0.60$). Participants received course credit for their participation or a cash payment of £3. Participants were randomly allocated (on the condition that there were

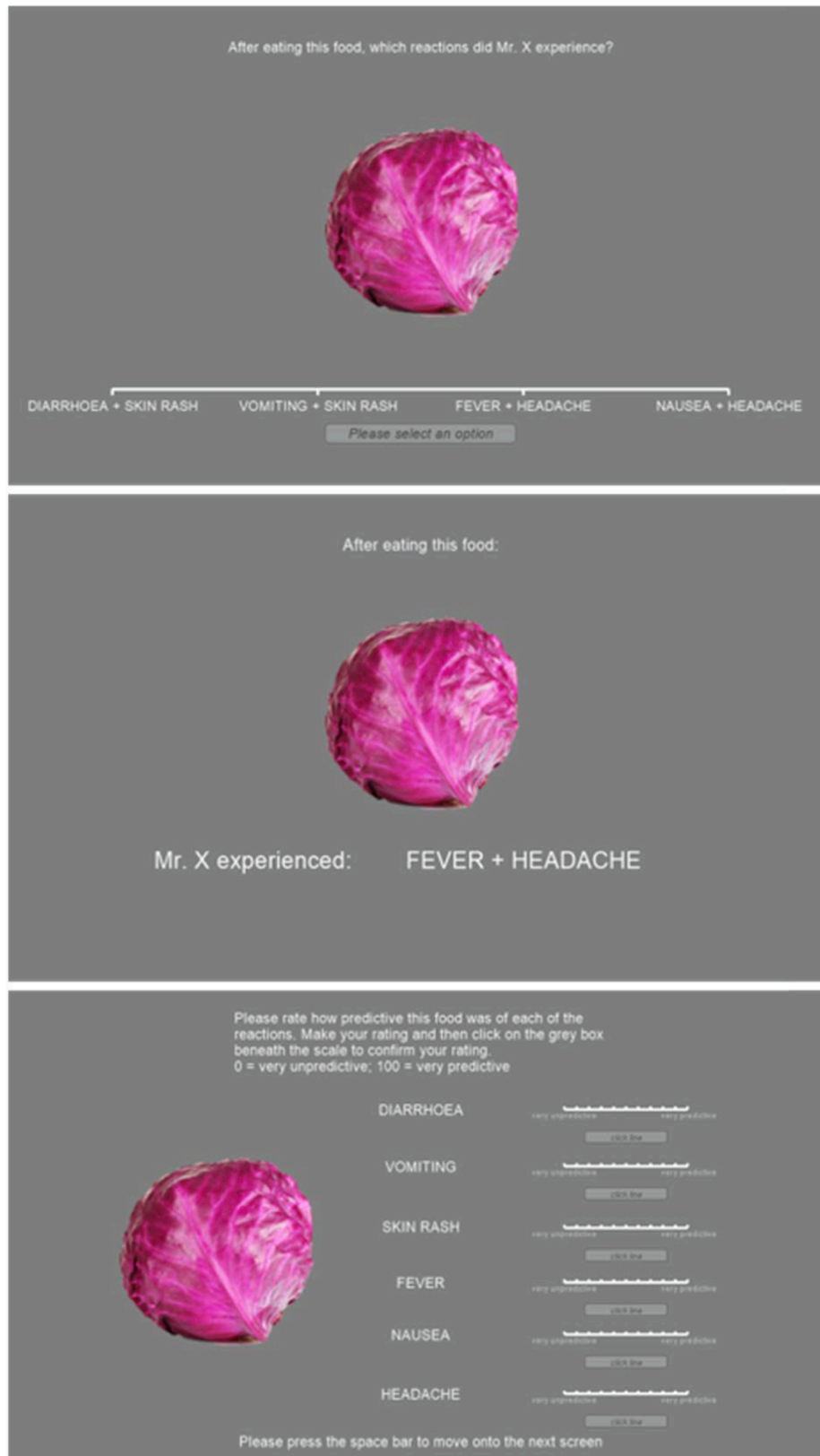


Fig. 4. Example of a training trial (top), feedback screen (middle) and a test screen (bottom) presented to Group causal in Experiment 2.

an equal number of participants in each group) to either complete either the Group causal ($n = 24$) or Group diagnostic ($n = 24$).

3.1.2. Apparatus and stimuli

Each of the foods and reactions were assigned to the letters and outcomes in Table 2 using a Latin-Squared counterbalancing technique. The six food stimuli used were: broccoli, cabbage, mushroom, pepper, potato, tomato. The six reactions included were: diarrhoea, fever, headache, nausea, skin rash and vomiting. Both the causal and diagnostic versions of the tasks were identical to that employed in Experiment 1, with the only change being that the outcomes were always presented in compound form. At test each of the cues (either foods or reactions dependent on the version of the task) were presented individually on the left of the screen, on the right each of the outcomes were presented (see Fig. 4). Next to each outcome a Likert scale was displayed, given that there were six outcomes presented, six scales were presented on the right of the screen. The order of these test rating screens was counterbalanced.

3.1.3. Procedure

Both groups experienced 96 trials in total with each trial type being presented 24 times across the training stage (see Table 2). There were six blocks of trials with 4 presentations of each trial type (A - O1O3, B - O2O3, C - O4O6/O5O6, D - O4O6/O5O6) in each block. Trial order was randomised. Once the training stage was completed, participants proceeded to a test stage. Aside from the number of trials the procedure for the training stage was the same as Experiment 1, except that in this experiment there were two additional outcomes employed to form the four outcome compounds. The Test stage was also identical to Experiment 1, except that on each rating screen participants were presented with six outcomes to rate as opposed to four.

3.2. Results

3.2.1. Training stage

Fig. 5 show the mean proportion of correct responses for the different cue-outcome pairings (i.e., stimuli) for Group Causal (top panel) and Group Diagnostic (bottom panel). The data have been collapsed into 6 blocks each containing 4 trials of each cue-outcome pairing (A - O1O3, B - O2O3, C - O4O6/O5O6, D - O4O6/O5O6). Mean proportion correct is averaged across A - O1O3 and B - O2O3 trials and averaged across C - O4O6, C - O5O6, D - O4O6 and D - O5O6 trials for each block of trials during Stage 1. As can be seen, regardless of the task completed (e.g., causal/diagnostic) participants displayed the same pattern of results, rapidly learning the relationships between the cue-outcome associations for the true-discrimination trials (i.e., stimuli A/B), reaching a performance asymptote by the end of training, yet remaining below 0.5 on the pseudo-discrimination trials (i.e., stimuli C/D). A three-way mixed model ANOVA was performed with a between-subject factors of task type (Causal vs Diagnostic), and within-subject factors of stimuli (A/B vs C/D) and block (1 - 6). The ANOVA revealed a main effect of block, $F(5, 194.58) = 18.01$, $MSE = 0.03$, $p < .001$, $\eta_p^2 = 0.28$, a main effect of stimuli, $F(1, 46) = 1245.39$, $MSE = 0.03$, $p < .001$, $\eta_p^2 = 0.96$, and a significant interaction between these two factors, $F(4.29, 197.20) = 6.75$, $p < .001$, $\eta_p^2 = 0.13$. There was no effect of task type, $F(1, 46) = 0.05$, $MSE = 0.07$, $p = .83$, $\eta_p^2 = 0.00$, and all other interactions were non-significant, largest $F(1, 46) = 0.70$, $MSE = 0.028$, $p = .41$, $\eta_p^2 = 0.02$. To further explore the source of the interaction between block (1-6) and stimuli (A/B vs C/D) simple main effects were performed. There was an effect of stimuli on all blocks, smallest $F(1, 46) = 131.75$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = 0.73$.

3.2.2. Test stage

One participant in Group Diagnostic failed to provide ratings for all the required outcomes and is therefore not included in the subsequent analyses. Fig. 6 shows participants' mean outcome-specific ratings

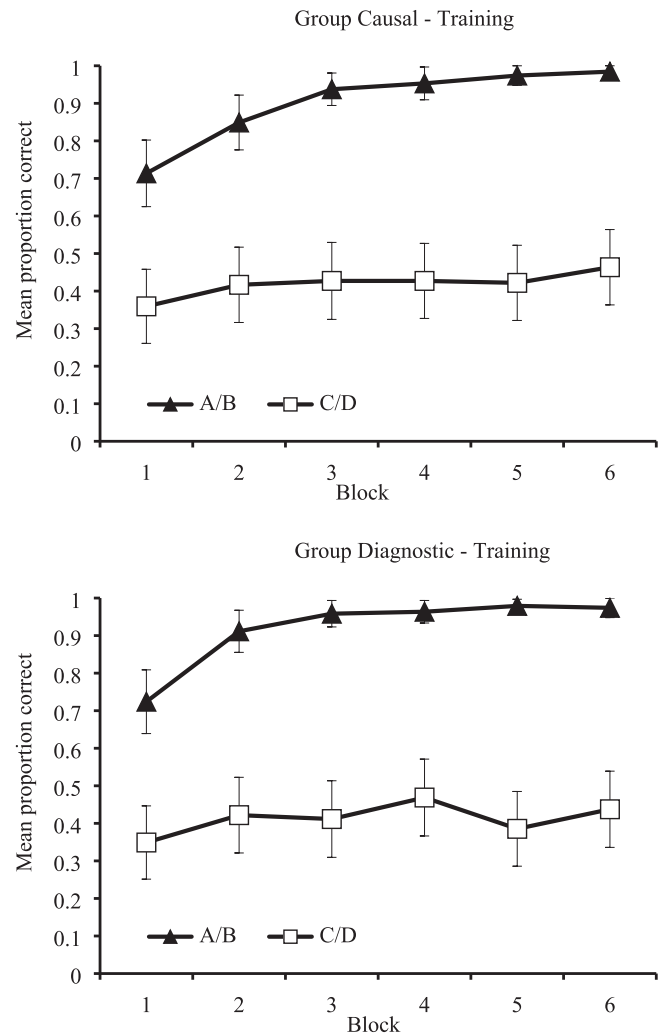


Fig. 5. Mean proportion of correct responses for A - O1O3/ B - O2O3 and C - O4O6/O5O6 and D - O4O6/O5O6 for Group Causal (Top panel) and Group Diagnostic (Bottom Panel) during the training stage of Experiment 2. Error bars represent SEM.

during the final test stage. These ratings were difference scores calculated for the relationships between cues A to D and outcomes O1 - O6. The key data of interest relate to O3 and O6, however, for the sake of completeness we also present ratings for O1, O2, O4 and O5. Difference scores were calculated by subtracting the rating for the outcome that a cue was not paired with from the rating for the outcome that the cue was paired with. For example, for stimuli A and B, participants' ratings for these cues and the outcome element that was presented on the pseudo-discrimination trials (O6) was subtracted from their ratings for these cues and the outcome that was presented on the true-discrimination trials (O3). For stimuli C and D, participants' ratings for these cues and the outcome element that was presented on the true-discrimination trials (O3) was subtracted from their ratings for these cues and the outcome that was presented on the pseudo-discrimination trials (O6). The same was done for O1, O2, O4 and O5. A 2×4 mixed model ANOVA with a between-subjects factor of Group (Causal vs Diagnostic) and a within-subjects factor of cue-outcome pairing (A/B - O1/O2, A/B - O3, C/D - O4/O5, C/D - O6), revealed no effect of Group, $F(1, 45) = 1.07$, $MSE = 3512.47$, $p = .31$, $\eta_p^2 = 0.02$ and no interaction between group and cue-outcome pairing, $F(3, 135) = 0.07$, $MSE = 642.27$, $p = .97$, $\eta_p^2 = 0.00$. However, there was an effect of outcome, $F(3, 135) = 4.40$, $MSE = 642.273$, $p < .01$, $\eta_p^2 = 0.09$. Crucially, Bonferroni corrected post-tests revealed that ratings for A/B-O3 and C/D-O6 did not differ $t = 0.13$, p

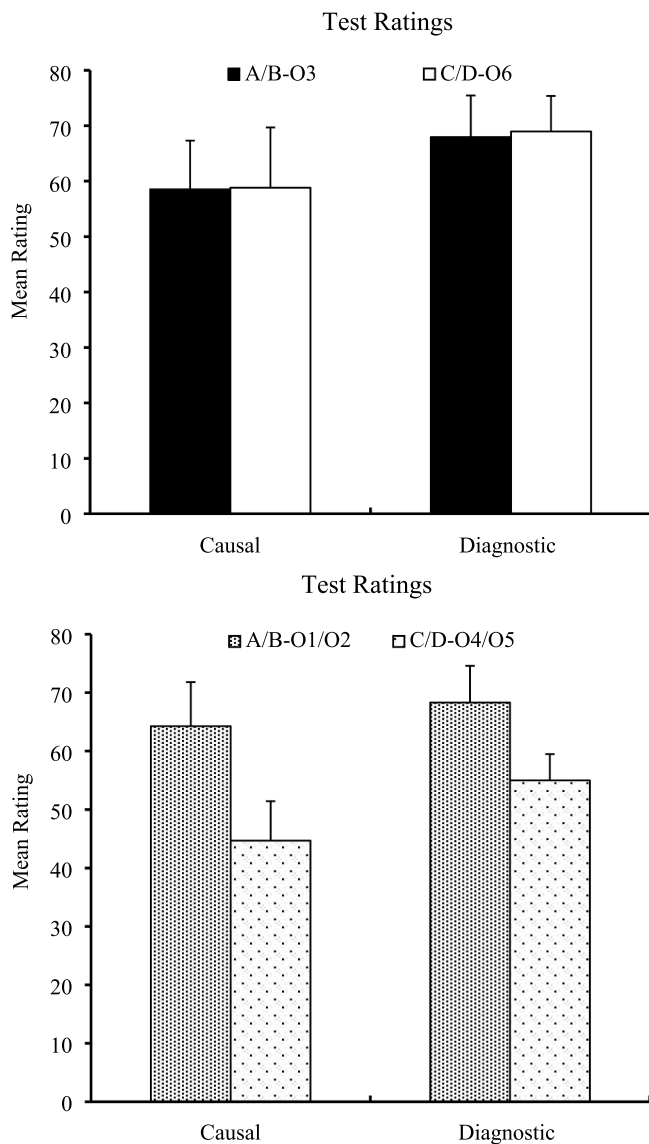


Fig. 6. Mean difference ratings for A/B – O3 and C/D - O6 (top panel) and A/B - O1/O2 and C/D - O4/O5 (bottom panel) for Group Causal and Group Diagnostic during the test stage of Experiment 2. Error bars represent SEM.

=.92, thus demonstrating no evidence of an outcome relative validity effect. However, ratings for C/D-O4/O5 were lower than A/B-O1/O2 ($t = 3.36, p < 0.01$) and C/D-O6 ($t = 2.71, p < 0.05$), this is unsurprising given that A/B consistently predicted O1/O2 and C/D consistently predicted O6, yet C/D inconsistently predicted O4/O5. All other comparisons were non-significant (smallest $p = .07$).

3.2.3. Discussion

In Experiment 2 participants completed an outcome relative-validity procedure in which on true-discrimination trials, one element of an outcome compound, O3, was less well predicted by cues A and B relative to elements O1 and O2. In contrast, on pseudo-discrimination trials, one element of an outcome compound, O6, was equally well predicted by cues C and D relative to O4 and O5. As noted in the introduction to this experiment, this establishes the relative validity of cues A and B for O3 to be less than the relative validity of cues C and D for O6. Despite this difference in relative validity, participants ratings of the relationship between cues A/B and O3 were comparable to the ratings of the relationship between C/D and O6. These results demonstrate the absence of an outcome interaction effect in an outcome relative-validity paradigm.

These results are surprising given the findings from Experiment 1, which revealed an overshadowing effect between the elements of an outcome. They are also surprising given the experiments reported by Quigley and Haselgrove (2020) which employed the same task as in the current studies, and which provided evidence for blocking and facilitation between the elements of an outcome. Furthermore, it is clear from Fig. 6 that the association between A/B and O1/O2 was greater than the association between C/D and O4/O5, and yet this differential stimulus control over the unique features of the outcome compound did not influence learning about the common feature of the outcome compound. Given the relative paucity of studies that have examined relative validity between the elements of an outcome, we therefore sought to examine this matter further.

4. Experiment 3

The results of Experiment 2 are inconsistent with the findings reported in the previous outcome-interaction experiments that have been conducted in our laboratory. All of these experiments have observed an interaction effect of one kind (facilitation: Quigley and Haselgrove, 2020, Experiments 1 to 3) or another (competition: Quigley & Haselgrove Experiment 2; Current Experiment 1). Models of associative learning, such as that proposed by Rescorla and Wagner (1972) explain cue interaction effects, such as blocking, overshadowing and relative validity, as different manifestations of a common underlying mechanism. Therefore, if we accept that interaction effects in outcomes have a degree of symmetry with interaction effects in cues (Gunther, Miller & Matute, 1997) then the presence of outcome interaction effects with blocking and overshadowing procedures suggests that we should also observe an interaction effect with an outcome relative validity procedure. However, this was not the case in Experiment 2. It is possible, however, that an additional factor may have influenced the results of Experiment 2. It could be, for instance, that the duration of the training which participants completed prior to the Test Stage influenced the results. Given the complexity of the task (relative to Experiment 1), the duration of training in Experiment 2 was extended from 32 trials to 96 trials (with each cue-outcome pairing being presented 24 times across the course of the experiment as opposed to 8 times), which is three times the length of Experiment 1, and also three times the length of the blocking stages of the experiments reported in Quigley and Haselgrove (2020). One motivation for having relatively shorter training stages in these previous experiments was to avoid any interaction effects being obscured by participants reaching a performance asymptote.

Experiment 3, therefore, sought to assess whether the extended amount of training in Experiment 2 precluded the observation of an outcome relative validity effect by reducing the amount of training participants received prior to the test. The experimental design was identical to that of Experiment 2 (see Table 3) with only the amount of training changing. If the duration of training in Experiment 2 was responsible for the comparable cue-outcome ratings provided at test, then the reduction in training may reveal differences in learning about cues A and B relative to cues C and D.

Table 3 Design of Experiment 3.

	Training	No. of trials	Test
True Discrimination	A – O1O3	8	
	B – O2O3	8	A: O1 - O6
Pseudo Discrimination	C – O4O6	4	B: O1 - O6
	C – O5O6	4	C: O1 - O6
	D – O4O6	4	D: O1 - O6
	D – O5O6	4	

Note. A – D refer to cues, O1 – O6 refer to outcomes. A and B represent the true-discrimination trials, while C and D trials represent the pseudo-discrimination trials.

4.1. Method

4.1.1. Participants

Forty-eight participants (40 females; 8 males) were recruited from the University of Nottingham's School of Psychology. Participants ranged from 18 to 22 years of age ($M = 19.00$; $SEM = 0.11$). All other details were the same as Experiment 2, with an equal number of participants in both Group Causal ($n = 24$) and Group Diagnostic ($n = 24$).

4.1.2. Apparatus, stimuli and procedure

The apparatus and stimuli used were the same as Experiment 2. The procedure was also the same as Experiment 2, with the only difference being that participants were presented with 32 trials during training (see Table 3) as opposed to the 96 trials included in Experiment 2 with each trial type being presented 8 times. There were two blocks of trials with 4 presentations of each trial type in each block. Trial order was randomised within each block and there was no break between blocks.

4.2. Results

4.2.1. Training stage

Fig. 7 shows the mean proportion of correct responses for cues A to D across training. In keeping with Experiment 1 and 2 the data have been collapsed into 2 blocks each containing 4 trials of each cue-outcome pairing (A - O1O3, B - O2O3, C - O4O6/O5O6, D - O4O6/O5O6). As in Experiment 2, participants quickly learnt the relationships between the cues and outcomes on the true-discrimination trials (i.e., stimuli A/B) but not the pseudo-discrimination trials (i.e., stimuli C/D), with participants' mean proportion of correct responses remaining below 0.5 during the pseudo-discrimination trials. The task type (causal vs diagnostic) participants completed also seemed to have little impact on their performance, which is consistent with Experiment 2.

A three-way mixed model ANOVA was conducted on the training data with a between-subject factor of Group (Causal vs Diagnostic), and within-subject factors of Stimuli (A/B vs C/D) and Block (1–2). The ANOVA revealed a main effect of Stimuli, $F(1, 46) = 282.52$, $MSE = 0.03$, $p < .001$, $\eta_p^2 = 0.86$, a main effect of Block, $F(1, 46) = 24.34$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = 0.35$, and a significant interaction between these two factors, $F(1, 46) = 20.31$, $p < .001$, $\eta_p^2 = 0.31$. There was no effect of task type, $F(1, 46) = 0.02$, $MSE = 0.04$, $p = .89$, $\eta_p^2 = 0.00$, and all other interactions were non-significant, largest $F(1, 46) = 0.25$, $MSE = 0.02$, $p = .62$, $\eta_p^2 = 0.01$. Simple main effects revealed there was an effect of stimuli on both blocks with participants producing a higher mean proportion of correct responses on the A/B trials, smallest $F(1, 46) = 125.58$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = 0.73$.

4.2.2. Test stage

Fig. 8 shows participants' outcome specific ratings during the final test for cues A to D and outcomes O1 - O6. As can be seen ratings were higher overall in Group Diagnostic relative to Group Causal. However, ratings for the A/B-O3 relationship were comparable to the C/D-O6 relationship, again suggesting the absence of an outcome relative validity effect. A 2×4 mixed model ANOVA with a between-subjects factor of Group (Causal vs Diagnostic) and a within-subjects factor of cue-outcome pairing (A/B - O1/O2, A/B - O3, C/D - O4/O5, C/D - O6) confirmed these impressions. There was a small effect of group $F(1, 46) = 4.11$, $MSE = 3786.22$, $p = .05$, $\eta_p^2 = 0.08$, with Group Diagnostic providing higher scores than Group Causal. Crucially, however, there was no effect of outcome, $F(3, 138) = 2.05$, $MSE = 863.26$, $p = .11$, $\eta_p^2 = 0.04$ and no interaction between group and outcome, $F(3, 138) = 1.24$, $MSE = 863.26$, $p = .30$, $\eta_p^2 = 0.03$. As such, there was no evidence of an outcome relative validity effect.

4.2.3. Discussion

Experiment 3 reproduced the absence of an outcome relative-validity effect that we observed in Experiment 2, with participants again

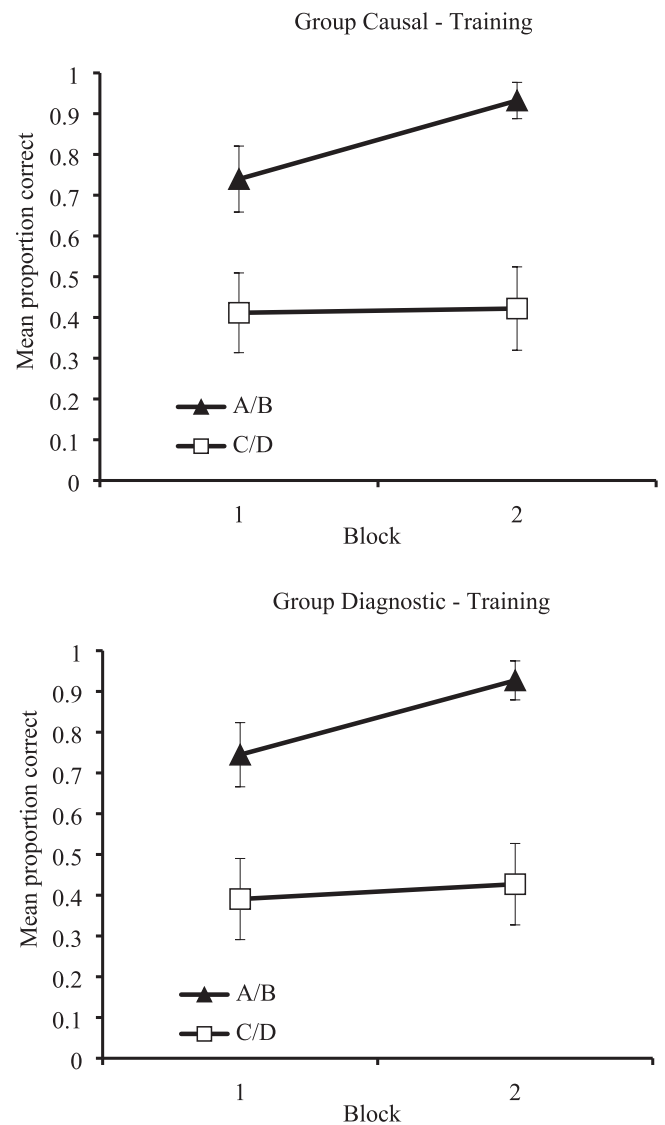


Fig. 7. Mean proportion of correct responses for trials with A - O1O3/B - O2O3 and C - O4O6/O5O6 and D - O4O6/O5O6 for Group Causal (Top panel) and Group Diagnostic (Bottom Panel) during the training stage of Experiment 3. Error bars represent SEM.

providing comparable ratings for the relationships between A/B and O3 (true discrimination trials) and the relationship between C/D and O6 (pseudo-discrimination trials). This effect was observed even though the duration of training was reduced, thus demonstrating that participants' extended training in Experiment 2 was unlikely to have influenced the results. It is notable, however, that ratings for C/D-O4/O5 were comparable to A/B-O1/O2, which is somewhat unexpected given their different predictive history and also the successful solution of the task during training (Fig. 7). One possible explanation for these results is that participants spontaneously configure compounds of stimuli after relatively few training trials (a la Experiment 3) which would permit the solution of the task but attenuate the observation of stimulus control of behaviour by its elements at test. With longer training, as in the case of Experiment 2 this configural representation of the outcome compound may be replaced by a more elemental representation of the compound, which will continue to support the solution of the task, but also permit the observation of individual stimulus control of behaviour when the elements of the compound are presented at test (Bellingham & Gillette, 1981).

Given the null effect obtained in Experiments 2 and 3 an additional

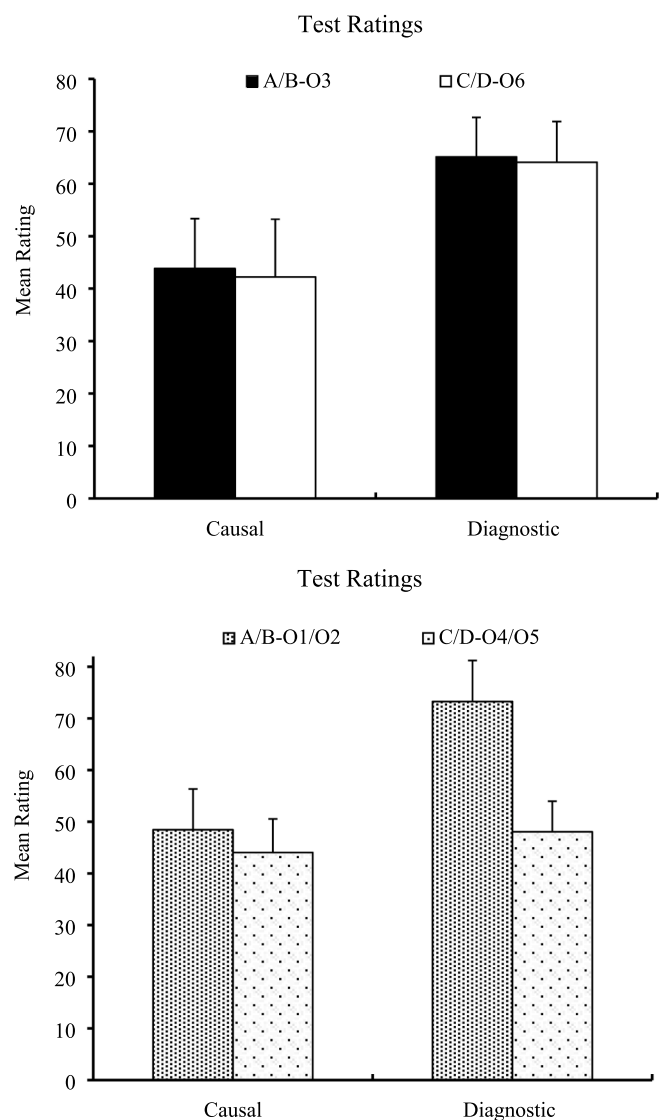


Fig. 8. Mean difference ratings for A/B-O3 and C/D O6 (top panel) and A/B-O1/O2 and C/D-O4/O5 (bottom panel) for Group Causal and Group Diagnostic during the test stage of Experiment 3. Error bars represent SEM.

Bayesian repeated measures ANOVA was conducted on the combined test data of the ratings of outcome elements O3 and O6, using default priors to estimate the Bayes Factors (Rouder, Morey, Speckman, & Province, 2012). This analysis allows us to evaluate the weight of evidence for the alternative hypothesis over the null (BF_{10}). Values > 1 provide evidence for the alternative hypothesis, < 1 represent evidence for the null hypothesis, and values $= 1$ do not provide evidence for either hypothesis (Lee & Wagenmakers, 2013). For the factor outcome (O3 vs O6) the $BF_{10} = 0.16$, which provides moderate evidence for the null hypothesis. For the interaction effect between the factors outcome and group the $BF_{10} = 0.05$, which also provides evidence for the null hypothesis. There was, however, anecdotal evidence in favour of the alternative hypothesis for the factor group $BF_{10} = 1.52$. As such, these results provide further evidence of an absence of interaction effects amongst the elements of an outcome when employing an outcome relative validity task.

5. General discussion

Three experiments investigated the presence of interaction effects among the elements of an outcome compound using outcome

overshadowing and relative validity procedures. In Experiment 1 an outcome overshadowing design was employed in which, during training, different cues were followed either by compounds of outcomes or by a single-element outcome. In a test stage, the elements of the outcomes received higher ratings when they had been trained in isolation with a cue than when they were trained as part of a compound, thus demonstrating an outcome overshadowing effect. In Experiments 2 and 3 an outcome relative validity procedure was employed in which on true-discrimination trials, one element of an outcome compound, O3, was less well predicted by cues A and B relative to elements O1 and O2. In contrast, on pseudo-discrimination trials, one element of an outcome compound, O6, was equally well predicted by cues C and D relative to O4 and O5. In both experiments, which differed only in terms of the amount of training given, there was no difference in participants' ratings of the relationship between A or B and O3, relative to C or D and O6 at test, thus demonstrating the absence of an outcome relative validity effect. A Bayesian analysis of the pooled data from Experiments 2 and 3 supported the null hypothesis.

The results of Experiment 1 are a challenge to classical theories of associative learning (e.g.: Esber & Haselgrove, 2011; Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972) which take a simplistic view of reward/outcome representation, and which conceive of the outcome as either present or absent. Instead, and in keeping with our earlier studies (Quigley & Haselgrove, 2020) Experiment 1 suggests that the elements of an outcome can interact to differentially support learning with a preceding cue. However, this general statement needs qualification, as Experiments 2 and 3 both failed to observe an outcome relative validity effect, and which when taken on their own, would be consistent with classical theories of learning such as the Rescorla-Wagner model (1972). Indeed, it is worth noting that relative validity effects in cues are only predicted by the Rescorla-Wagner model (1972) when it is assumed that the learning rate for reinforcement is greater than for non-reinforcement. If, however, these rates are comparable (as would be the case in Experiments 2 and 3 where an outcome is presented on every trial) then a relative validity effect would not be expected to occur. The results of Experiment 1 and Quigley and Haselgrove (2020) pose a greater challenge for such models, however, and an alternative approach to outcome representation therefore seems necessary (e.g., Delamater, 2012; Delamater & Oakeshott, 2007).

Quigley and Haselgrove (2020) proposed that the outcome-interaction effects that they observed could be understood in terms of the theory proposed by Wagner (1981). According to this theory, stimuli comprise multiple elements each of which can occupy one of three different states: a primary state (A1), a secondary state (A2) or an inactive state (I). Before being presented, a stimulus is in the inactive state where it is unable to attract attention, produce a response or engage in learning. Upon the presentation of a stimulus, however, the elements of the stimulus enter an A1 state, when the stimulus is regarded as being in the centre of attention and can produce its strongest response. From A1, the elements of the stimulus decay rapidly into the A2 state when the elements produce a weaker response and are considered to be at the periphery of an organism's focus. Elements decay from A2 back into inactivity relatively slowly which effectively provides A2 with a greater capacity than A1. As noted, stimulus elements can enter their A2 state through rapid decay from A1, however, stimulus elements can go directly from inactivity to the A2 state by being associatively activated by another stimulus, and importantly, the state(s) in which the elements of two stimuli reside is key to the type of association which may form between them. When the elements of two stimuli are both in an A1 state, an excitatory association will form between them; if, however, the elements of a cue are in A1 and the elements of an outcome are in A2, then an inhibitory association will form between them. According to Wagner's theory overshadowing will be observed among the elements of an outcome compound. This follows because the extent to which both elements of a compound outcome can be in A1 at the same time as the preceding cue will be restricted due to capacity limits,

reducing the extent to which an excitatory association can form between the cue and both elements of the outcome. However, this reduction will be less when only one outcome element is presented following the cue. Interestingly, this analysis predicts the presence of reciprocal overshadowing. That is, ratings of the relationship between a cue and *both* elements of an outcome compound should receive comparably low ratings relative to the case where a cue is followed by on a single outcome element, which is precisely the effect that we observed in Experiment 1.

The effect of outcome relative validity training that can be derived from Wagner's (1981) theory is a little more complex. According to Wagner's theory, on true-discrimination trials (A - O1O3 and B - O2O3) the ability of A or B to become associated with O3 will be restricted because the presence of O1 and O2 on these trials will preclude O3 from fully entering the limited capacity of A1 – that is to say, overshadowing will occur. The same effect will also occur on the pseudo-discrimination trials in which C and D are paired with an O4O6 compound on half of the trials and an O5O6 compound on the remaining trials. However, because C and D are followed on some trials with O4 and O5, but on other trials by their absence (e.g., O4 is absent on a C - O5O6 trial) there is also the opportunity for an inhibitory association to form between these cues and these outcomes. What will be the impact of this inhibition? One possibility is for the net excitation between C/D and O4/O5 to simply be lower (i.e., the inhibition between C/D and O4/O5 is simply subtracted from the excitation between these cues and outcomes). As training continues, then, the extent to which O4 and O5 can be activated from inactivity into A2 following cues C and D will be relatively lower (Wagner, 1981, p.13). Consequently, O4 and O5, will better activate from inactivity into A1 and, given the limited capacity of this state, will better overshadow learning about O6 (relative to overshadowing of O3 by O1 O2). Consequently, Wagner's theory appears to predict the presence of an outcome facilitation effect here, rather than a competition effect.

It is possible that two competing effects are operating simultaneously within our sample; a facilitation effect in accordance with the model proposed by Wagner (1981, noted above) and a competition effect produced as a result of something else. One candidate for this "something else" is outcome predictiveness. A growing body of recent literature has demonstrated that the predictability of an outcome can influence subsequent new learning featuring the outcome (Griffiths et al., 2015; Quigley et al., 2019; Thorwart et al., 2017). For instance, Griffiths et al. (2015) demonstrated – using a comparable task to that employed in the current studies – that a well predicted outcome is learned about better than a less-well predicted outcome when paired with a novel cue. In their study, outcomes were also presented as part of an outcome compound thus providing a parallel to our relative validity experiments. Considering Griffiths et al. results, it is therefore possible that the O6 presented on the pseudo discrimination trials enjoyed an enhanced associability as it was a well-predicted outcome, relative to O4 and O5 which were also presented on both trials with C and D. The same does not hold true for O3 on the true discrimination trials though as O1 and O2 are at least as equally well predicted by A and B as O3, and arguably better predicted as their presence is differentially predicted by A and B.

Outcome predictiveness can be regarded as a *global* effect on outcome processing as experience changes the effectiveness by which an outcome can support learning *that transfers to entirely novel learning contexts* - such as instances in which that outcome is later signalled by new cues (e.g., Griffiths et al., 2015). The mechanism that is provided by Wagner's (1981) theory however is *local*, for here the effectiveness by which an outcome can support learning is determined by the extent to which it is signalled by a specific cue or cues on that trial. These local and global influences on outcome processing may sometimes complement one another, whilst on other occasions they may be in conflict (as in the case of outcome relative validity). The variables that determine the relative dominance of either local or global outcome processing

mechanisms on learning remains to be determined,

An alternative account of the current results could also be provided by considering the procedural requirements of the outcome relative validity task. In Experiments 2 and 3 the common outcomes appearing on the true-discrimination trials and the pseudo-discrimination trials are presented more frequently than other outcomes in the experiment. This is also the case with the outcomes of interest in the outcome overshadowing experiment. Unlike the outcome overshadowing experiment, however, the outcomes of interest in the relative validity task are presented alongside several different outcomes which are either inconsistently predicted (on the pseudo-discrimination trials) or presented less frequently (on the true-discrimination trials) than the two primary outcomes of interest (O3 and O6). As such, it is possible that participants are inadvertently directed to focus on the two outcomes of primary interest, as these outcomes are either presented the most or predicted consistently, and therefore participants learn about these stimuli to the same extent (e.g., Popov & Reder, 2020).

It is also possible that within-compound associations may have played a role in Experiments 2 and 3. Other things being equal, a glance at Table 3 suggests that the within-compound associations between O3 and O1 and O2, and between O6 and O4 and O5 are equivalent. O3 is paired with the presence and absence of O1 and O2 as often as O6 is paired with the presence and absence of O4 and O5. However, other things are not equal. The pseudo-discrimination trials sustain a prediction error throughout training, as these trials can never be solved. It is conceivable that this prediction error influenced the extent to which within-compound associations form between the elements of the compound outcome. Indeed, Holland (1980) has shown in studies of animal conditioning that a surprising US can interfere more with associative learning between preceding stimuli than a predicted US. Future research should be undertaken to further examine the role of within-compound associations in the context of outcome competition tasks such as those we report here. Similarly, future research could also examine the potential impact of contextual associations. For example, O3 is never presented in the absence of the experimental context, however, it is presented in the absence of A and B (for example, O3 is presented in the absence of B on the A-O1O3 trials). Consequently, the experimental context can be viewed as a better predictor of O3 than A or B. This being the case we might expect the context to block learning about the A/B-O3 association to some degree. The question of interest, of course, is whether this context blocking to O3 would be greater or less than experienced by O6. Future research could overcome this issue by employing an experimental design in which the key outcome of interest (e.g., O3) is presented on both true and pseudo-discrimination trials (e.g., A - O1O3, B - O2O3, C - O1O3, C - O2O3).

In conclusion, the current results join a growing literature which demonstrates that stimulus-compound interaction effects – a hallmark of associative learning, and the catalyst to the development of the Rescorla and Wagner (1972) theory, are not just restricted to cues. Training that provides analogues of blocking, overshadowing and relative validity in the domain of outcome compounds reveals similar effects. Sometimes. The parameters that determine this "sometimes" remain to be fully understood.

Author Note

This work was supported by the Economic and Social Research Council (ES/I021108/1) and contributed to Martyn Quigley's doctorate degree by funding a studentship.

CRediT authorship contribution statement

Martyn Quigley: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Writing - review & editing. **Mark Haselgrove:** Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Arcediano, F., Matute, H., Escobar, M., & Miller, R. R. (2005). Competition between antecedent and between subsequent stimuli in causal judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 228–237. <https://doi.org/10.1037/0278-7393.31.2.228>
- Bellingham, W. P., & Gillette, K. (1981). Spontaneous Configuring to a Tone-Light Compound Using Appetitive Training. *Learning and Motivation*, 12(4), 420–434. [https://doi.org/10.1016/0023-9690\(81\)90003-5](https://doi.org/10.1016/0023-9690(81)90003-5)
- Blaser, R. E., Couvillon, P. A., & Bitterman, M. E. (2006). Blocking and pseudoblocking: New control experiments with honeybees. *The quarterly journal of experimental psychology*, 59(1), 68–76. <https://doi.org/10.1080/17470210500242938>
- Chamizo, V. D., Aznar-Casanova, J. A., & Artigas, A. A. (2003). Human overshadowing in a virtual pool: Simple guidance is a good competitor against locale learning. *Learning and Motivation*, 34(3), 262–281. [https://doi.org/10.1016/S0023-9690\(03\)00020-1](https://doi.org/10.1016/S0023-9690(03)00020-1)
- Cobos, P. L., López, F. J., Cano, A., Almaraz, J., & Shanks, D. R. (2002). Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology: Animal Behavior Processes*, 28(4), 331–346. <https://doi.org/10.1037/0097-7403.28.4.331>
- Delamater, A. R. (2012). On the nature of CS and US representations in Pavlovian learning. *Learning & Behavior*, 40(1), 1–23. <https://doi.org/10.3758/s13420-011-0036-4>
- Delamater, A. R., & Oakeshott, S. (2007). Learning about multiple attributes of reward in Pavlovian conditioning. *Annals of the New York Academy of Sciences*, 1104(1), 1–20. <https://doi.org/10.1196/annals.1390.008>
- Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology*, 36(1), 29–50. <https://doi.org/10.1080/14640748408401502>
- Dwyer, D. M., Haselgrove, M., & Jones, P. M. (2011). Cue interactions in flavor preference learning: A configurational analysis. *Journal of Experimental Psychology: Animal Behavior Processes*, 37(1), 41–57. <https://doi.org/10.1037/a0021033>
- Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of experimental psychology: human perception and performance*, 27(1), 229–240. <https://doi.org/10.1037/0096-1523.27.1.229>
- Esber, G. R., & Haselgrove, M. (2011). Reconciling the influence of predictiveness and uncertainty on stimulus salience: A model of attention in associative learning. *Proceedings of the Royal Society B: Biological Sciences*, 278(1718), 2553–2561. <https://doi.org/10.1098/rspb.2011.0836>
- Flach, R., Osman, M., Dickinson, A., & Heyes, C. (2006). The interaction between response effects during the acquisition of response priming. *Acta Psychologica*, 122(1), 11–26. <https://doi.org/10.1016/j.actpsy.2005.09.001>
- Gunther, L. M., Miller, R. R., & Matute, H. (1997). CSs and USs: What's the difference? *Journal of Experimental Psychology: Animal Behavior Processes*, 23(1), 15–30. <https://doi.org/10.1037/0097-7403.23.1.15>
- Holland, P. C. (1980). Second-order conditioning with and without unconditioned stimulus presentation. *Journal of Experimental Psychology: Animal Behavior Processes*, 6, 238–250. <https://doi.org/10.1037/0097-7403.6.3.238>
- Holland, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or blocking cues. *The Quarterly Journal of Experimental Psychology: Section B*, 52(4), 307–333. <https://doi.org/10.1080/027249999393022>
- Jones, P. M., & Haselgrove, M. (2013). Overshadowing and associability change: Examining the contribution of differential stimulus exposure. *Learning & Behavior*, 41(1), 107–117. <https://doi.org/10.3758/s13420-012-0089-z>
- Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior: Aversive Stimulation. Coral Gables, Florida* (pp. 9–31). Florida: University of Miami Press.
- Kremer, E. F. (1978). The Rescorla-Wagner model: 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>
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. New York, NY, US: Cambridge University Press.
- 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, Comparative and Physiological Psychology*, 57(3), 193–243. <https://doi.org/10.1080/02724990344000141>
- Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *The Quarterly Journal of Experimental Psychology*, 56B(1), 68–79. <https://doi.org/10.1080/02724990244000179>
- 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(1), 56–70. <https://doi.org/10.1037/0097-7403.31.1.56>
- Lochmann, T., & Wills, A. J. (2003). Predictive history in an allergy prediction task. In *Proceedings of EuroCogSci 03: The European Conference of the Cognitive Science Society* (pp. 217–222). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Mackintosh, N. J. (1971). An analysis of overshadowing and blocking. *Quarterly Journal of Experimental Psychology*, 23(1), 118–125. <https://doi.org/10.1080/00335557143000121>
- 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>
- Mackintosh, N. J. (1976). Overshadowing and stimulus intensity. *Animal Learning and Behaviour*, 4, 186–192. <https://doi.org/10.3758/BF03214033>
- Matute, H., Arcediano, F., & Miller, R. R. (1996). Test question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 182–196. <https://doi.org/10.1037/0278-7393.22.1.182>
- Miller, R. R., & Matute, H. (1998). Competition between outcomes. *Psychological Science*, 9(2), 146–149. <https://doi.org/10.1111/1467-9280.00028>
- Griffiths, O., Mitchell, C. J., Bethmont, A., & Lovibond, P. F. (2015). Outcome predictability biases learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, 41(1), 1. <https://doi.org/10.1037/xan0000042>
- Pavlov, I. P. (1927). *Conditioned reflexes* (G. V. Anrep, Trans.). London: Oxford University Press.
- 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., Graham, M., Good, M. A., Jones, P. M., & McGregor, A. (2006). Potentiation, overshadowing, and blocking of spatial learning based on the shape of the environment. *Journal of Experimental Psychology: Animal Behavior Processes*, 32(3), 201–214. <https://doi.org/10.1037/0097-7403.32.3.201>
- 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>
- Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Peirce, J. W. (2008). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics*, 2, 10. <https://doi.org/10.3389/neuro.11.010.2008>
- Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological review*, 127(1), 1. <https://doi.org/10.1037/rev0000161>
- Quigley, M. C., Eatherington, C. J., & Haselgrove, M. (2019). Learned changes in outcome associability. *Quarterly Journal of Experimental Psychology*, 72(2), 209–221. <https://doi.org/10.1080/17470218.2017.1344258>
- Quigley, M., & Haselgrove, M. (2020). Interactions between the elements of an outcome in human associative learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, 46(3), 297–313. <https://doi.org/10.1037/xan0000248>
- Rescorla, R. A. (1980). *Pavlovian second-order conditioning: Studies in associative learning*. Hillsdale, NJ: 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). New York: Appleton-Century-Crofts.
- Revusky, S. (1971). The role of interference in association over a delay. *Animal memory*, 155–213.
- Rodrigo, T., Chamizo, V. D., McLaren, I. P. L., & Mackintosh, N. J. (1997). Blocking in the spatial domain. *Journal of Experimental Psychology: Animal Behavior Processes*, 23(1), 110–118. <https://doi.org/10.1037/0097-7403.23.1.110>
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374. <https://doi.org/10.1016/j.jmp.2012.08.001>
- Sánchez-Moreno, J., Rodrigo, T., Chamizo, V. D., & Mackintosh, N. J. (1999). Overshadowing in the spatial domain. *Animal Learning & Behavior*, 27(4), 391–398. <https://doi.org/10.3758/BF03209976>
- Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge University Press.
- Tennant, W. A., & Bitterman, M. E. (1975). Blocking and overshadowing in two species of fish. *Journal of Experimental Psychology: Animal Behavior Processes*, 1(1), 22–29. <https://doi.org/10.1037/0097-7403.1.1.22>
- Thorwart, A., Livesey, E. J., Wilhelm, F., Liu, W., & Lachnit, H. (2017). Learned predictiveness and outcome predictability effects are not simply two sides of the same coin. *Journal of Experimental Psychology: Animal Learning and Cognition*, 43(4), 341. <https://doi.org/10.1037/xan0000150>
- Van Hamme, L. J., Kao, S. F., & Wasserman, E. A. (1993). Judging interevent relations: From cause to effect and from effect to cause. *Memory & Cognition*, 21(6), 802–808. <https://doi.org/10.3758/BF03202747>
- Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, 412(6842), 43–48. <https://doi.org/10.1038/35083500>
- Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of experimental psychology*, 76(2 p1), 171–180. <https://doi.org/10.1037/h0025414>

- Wagner, A.R. (1981). SOP: a model of automatic memory processing in animal behavior. In: Spear, N.E., Miller, R.R. (Eds.), *Information Processing in Animals: Memory Mechanisms*. Erlbaum, Hillsdale, NJ, pp. 5–47.
- Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, 121(2), 222–236. <https://doi.org/10.1037/0096-3445.121.2.222>
- Willner, J. A. (1978). Blocking of a taste aversion by prior pairings of exteroceptive stimuli with illness. *Learning & Motivation*, 9, 125–140. [https://doi.org/10.1016/0023-9690\(78\)90016-4](https://doi.org/10.1016/0023-9690(78)90016-4)