2019-01

Uncertainty and blocking in human causal learning

Jones, Peter

10.1037/xan0000185
Journal of Experimental Psychology: Animal Learning and Cognition
American Psychological Association

Uncertainty and blocking in human causal learning

Peter M. Jones, Tara Zaksaite, and Chris J. Mitchell

School of Psychology, Plymouth University, Plymouth, UK

Running head: Blocking and uncertainty

Author note:

The experiments reported here were conducted as part of Tara Zaksaite's PhD. Raw data have been archived using APA's repository hosted by the Center for Open Science at osf.io/ynje9. Correspondence concerning this article should be addressed to Peter M. Jones, School of Psychology, Plymouth University, Plymouth, PL4 8AA, United Kingdom. Email: peter.m.jones@plymouth.ac.uk

The blocking phenomenon is one of the most enduring issues in the study of learning. Numerous explanations have been proposed, which fall into two main categories. An associative analysis states that, following A+/AX+ training, cue A prevents an associative link from forming between X and the outcome. In contrast, an inferential explanation is that A+/AX+ training does not permit an inference that X causes the outcome. More specifically, the trials on which X is presented (AX+) are often argued to be uninformative with respect to the causal status of X because the outcome would have resulted on AX trials whether X was causal or not. If participants are uncertain about X, their ratings on test might be particularly sensitive to the overall base rate of the outcome. That is, a blocked cue, about which one is uncertain, should be rated as a more likely cause when most cues lead to the outcome than when most cues do not. This hypothesis was supported in two experiments. Experiment 1 used an overshadowing control and Experiment 2 used an uncorrelated control (to demonstrate a redundancy effect). Variations in the ratings of the blocked cue as a result of manipulating the outcome base rate can be explained if participants are uncertain about the status of the blocked cue. Experiment 3 showed that participants are uncertain about blocked cues by using a direct self-report measure of certainty. These data are consistent with the inferential account, but are more challenging for the associative analysis.

Keywords: associative learning, blocking, redundancy effect, base rate, uncertainty

Introduction

In many causal learning experiments, participants are presented with multiple potential causes of an outcome. The job of the participant is to work out which events are most likely to cause the outcome to occur. For example, in the allergist task participants assume the role of a doctor and are shown a series of meals eaten by a fictional patient. Each meal consists of one or more foods, followed on some occasions by an allergic reaction, with participants required to decide which foods cause this reaction. Aitken, Larkin, and Dickinson (2001) used this paradigm to demonstrate cue competition; that is, they found that learning about a given food was dependent to some extent on the status of the other foods that were present in the same meal. More specifically, Aitken et al. demonstrated blocking. In their experiments, a single food A was paired with the allergic reaction (A+, where the + denotes the presence of the reaction). On subsequent trials, a compound of A and another food, X, was also paired with the allergic reaction (AX+). Participants were then asked to rate the likelihood of the allergic reaction following consumption of each individual food. Ratings for X were lower here than in a control condition in which A had not been paired separately with the outcome. Informally, we say that A 'blocked' X.

This resembles the blocking effect observed in non-human animals (e.g. Kamin, 1969) and is consistent with the model of learning proposed by Rescorla and Wagner (1972). In this model cues compete to become associated with the outcome because the prediction error that governs learning takes into account all cues that are present. Following A+ training, the outcome is predicted by A and is therefore less able to support learning about X on AX+ trials. However, while this model can readily explain blocking, it is less successful at explaining the fact that blocking is often far from complete, particularly in human learning experiments (Lovibond, Been, Mitchell, Bouton, & Frohardt, 2003). In its simplest form,

Rescorla and Wagner's model predicts not just that learning about X will be restricted to some extent, but that X will be virtually unable to become associated with the outcome as a result of learning about A.

One way of solving this problem is to assume that each cue shares a common element that becomes associated with the outcome in the same way as any other feature. This approach has been used to reconcile Rescorla and Wagner's (1972) model with other findings that, at first glance, do not seem to be compatible with it (e.g. Haselgrove, 2010). In the case of blocking, this assumption allows the model to predict a substantial amount of associative strength for the blocked cue. This is because the common element competes with the distinctive features of A on A+ trials, and becomes associated with the outcome as a result. Since the common element is also present when X is tested alone, the association between the common element and the outcome supports an expectation that the outcome will occur.

This explanation relies on the assumption that blocking occurs because the blocked cue has only a weak association with the outcome. Low probability ratings on test are a reflection of the weakness of this link. But does reduced learning necessarily imply a weak association? A quite different interpretation has been suggested by Cheng and her colleagues (Cheng, 1997; Cheng & Holyoak, 1995; Waldmann & Holyoak, 1992). She proposed that, rather than restricting the extent to which the blocked cue X becomes associated with the outcome, the blocking cue A prevents participants from making a valid inference about the causal status of X. Their argument is that, since participants only encounter cue X during AX+ trials, and since A caused the allergic reaction by itself on A+ trials, there are two possibilities with respect to the causal status of X. Firstly, the reaction that occurred on AX+ trials may have been due solely to A and not to X. Secondly, both A and X might be causes of the outcome. The potential effect of X alone is therefore unknown; participants do not

have enough information about this cue to conclude that the allergic reaction will occur. This theory also accounts for the incomplete nature of blocking, since participants also lack evidence that X is *not* a cause of the allergic reaction, and should therefore give intermediate probability ratings for X during test.

The above logic, in which the causal status of X is ambiguous, applies well to the usual causal scenario in which the allergic reaction either occurs or does not occur on each trial. However, under other conditions, it is possible for a valid inference to be drawn about cue X following A+ and AX+ training. Lovibond et al. (2003; see also De Houwer, Beckers & Glautier, 2002) achieved this by allowing allergic reactions of different strengths to occur. As well as the standard allergic reaction, the fictional patient sometimes suffered from a severe reaction (denoted "++"). For one group of participants, an initial phase of training demonstrated that this severe reaction followed the consumption of two foods (IJ++) that individually caused the normal allergic reaction (I+ & J+). This rule can be termed additivity, because the magnitude of the reactions to individual foods I and J 'added together' when they were consumed in the same meal to produce a strong allergic reaction. If participants extracted this rule and applied it to subsequent A+/AX+ training, ambiguity about the causal status of X should have been reduced. This is because if both A and X were causes of the normal allergic reaction, their joint consumption should have resulted in the severe allergic reaction (AX++), which it did not. Cue X could not therefore have been a cause of allergic reaction by itself. Lovibond et al. demonstrated that probability ratings for X were indeed lower (and the blocking effect was larger) for participants who had received additivity training, when compared to a group of participants for whom allergic reactions were non-additive.

To return to the standard blocking effect in the absence of additivity training, if the status of X is ambiguous, then participants are required to guess the status of X based on whatever evidence is available. One line of evidence that might be relevant is the extent to which, in general, cues are followed by the outcome – the base rate of the outcome. In the usual causal learning scenario, participants might ask themselves, "To what extent do other foods usually produce an allergic reaction in this patient?" If the base rate of the outcome is high (the patient is allergic to most foods), then an ambiguous cue about which one has no information (the blocked cue X) is more likely to be followed by the outcome than when the base rate is low. If this is true, changing the overall base rate of the outcome should change the ratings given for the blocked cue, and in turn the size of the blocking effect. Some support for the idea that base rates influence predictions when cues are uncertain, albeit from a somewhat different paradigm, comes from Kahneman and Tversky (1973). They asked participants to predict whether a fictional person was an engineer or a lawyer, either when there were more engineers than lawyers in the population, or when lawyers were more prevalent. Some of the participants were also given a short vignette that was designed to be representative of either engineers or lawyers. Kahneman and Tversky found that when participants were given a vignette, judgments of the probability that the fictional person was an engineer or a lawyer were not affected by the base rates. However, when they were not given a vignette, and hence had no information to use other than the base rates, probability judgments followed the base rates closely. It follows that this effect might be observed for blocked cues, about which no other information is available.

The notion that the base rate of the outcome might be related to the size of the blocking effect has already been suggested by Livesey, Lee, and Shone (2013). They pointed out that, while the base rate of the outcome should affect ratings for a blocked cue X, this effect should be smaller for the standard "overshadowing" control cue (e.g. Y from the

trained compound BY+). This is because, while participants will be uncertain about which of the overshadowing cues B and Y cause the outcome, they know that at least one of these cues must be a cause. It follows that, even when the base rate for the outcome is low, the average probability of B or Y causing the outcome cannot be lower than 0.5. At low base rates, blocking will be substantial because ratings for the blocked cue will be very low, whereas high outcome rates will result in high ratings for both blocked and overshadowing control cues and consequently less blocking. However, this prediction assumes that participants will use conditional probabilities to infer the likelihood of the outcome occurring for each cue, which (as Livesey et al. found) they may not do.

Although the inferential account summarised above predicts an effect of base rate on the size of blocking, a modified version of Rescorla and Wagner's (1972) model that incorporates a common element can also explain the effect in the following way. When the base rate of the outcome is high, the common element will have many opportunities to become associated with the outcome and will acquire a large amount of associative strength. Since the common element will compete with the distinctive features of each cue, those distinctive features will acquire a small amount of associative strength and the differences between cues will be correspondingly small. Conversely, when the base rate of the outcome is low, the common element will gain less associative strength and the distinctive features will gain more, allowing for larger differences between cues. As a consequence, the blocking effect (the difference in probability ratings between the control and blocked cues) should be smaller when the base rate is high than when it is low.

Given that both inferential and associative accounts predict less blocking when the base rate is high than when it is low, we were somewhat surprised that we could not find a published demonstration of this effect. Experiment 1 in the current paper tested this

prediction by training participants on a version of the allergist task containing blocking and control cues. Additional trials were added with alternative cues in order to manipulate the base rate of the outcome. It was predicted that reducing the outcome rate would result in a larger blocking effect. More specifically, low base rate will reduce probability ratings more for the blocked cue X than for the overshadowing controls B and Y.

Experiment 1

We used a typical allergist task in which participants were required to learn which foods caused an allergic reaction in a fictitious patient. The design of the experiment is shown in Table 1. Two groups of participants were given blocking training comprising A+ and AX+ trials. In addition, the BY+ trials served as an overshadowing control. The remaining 'filler' trials were included so that the overall proportion of trials on which the outcome occurred could differ between the two groups. For participants in the 75% group, stomach ache occurred for six of the nine 'filler' trial types. As a result, stomach ache occurred on 9 trials out of 12, or 75% of all trials. For participants in the 25% group, no trials other than A+, AX+, and BY+ included the stomach ache outcome. As a result, stomach ache occurred on only 3 of the 12 types of trials, or 25%. For both groups, 'filler' trials included both single foods and two-item compounds. Participants were therefore unable to learn a response rule based on the number of items present. Following this training stage, participants were asked to rate the likelihood of stomach ache for A, B, C, X, Y & Z. If participants base their ratings of the blocked cue on the overall outcome rate, then participants in the 75% group should give higher ratings for X than participants in the 25% group. Following the arguments summarised in the Introduction, we predicted that any effect of outcome rate on ratings for the overshadowing controls B and Y should be smaller.

Consequently, a more substantial blocking effect was predicted for the 25% group than for the 75% group.

Method

Participants. Forty Psychology undergraduate students (20 per group) at Plymouth University took part in this experiment. They were aged 18-53 (M = 22.28, SD = 6.71) and ten were male. They received course credit for their participation.

Materials. Participants were tested individually in cubicles at Plymouth University. The experiment was presented on a 22-inch desktop computer with a 1280 x 1024 screen resolution. The experiment was designed, presented, and responses recorded, using E-prime 2.0 software (Psychology Software Tools, PA, US). Pictures of 19 foods served as cues in the experiment: apple, banana, broccoli, cabbage, cherries, corn, grapefruit, grapes, kiwi, mango, orange, peach, pear, pepper, pineapple, pomegranate, pumpkin, strawberries, and watermelon. Each food was presented against a white square, measuring 300 x 300 pixels. The foods were randomly assigned to each type of cue (A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, X, Y, Z) for each participant. The outcomes were stomach ache, signified by text and a sad face on a red background, and no stomach ache, indicated by text and a happy face on a green background. Cues and outcomes were presented on a black background with white text. Participants responded using the mouse.

Procedure. The instructions for the learning task were adapted from Uengoer, Lotz and Pearce's (2013) causal learning task, and were presented on the screen as follows:

This study is concerned with the question of how people learn about relationships between different events. In the present case, you should learn whether the consumption of certain foods leads to stomach ache or not.

Imagine that you are a medical doctor. One of your patients often suffers from stomach ache after meals. To discover the foods the patient reacts to, your patient eats specific foods and observes whether stomach ache occurs or not.

The results of these tests are shown to you on the screen one after the other. You will always be told what your patient has eaten. Sometimes he has only consumed a single kind of food, and other times he has consumed two different foods. Please look at the foods carefully.

Thereafter you will be asked to predict whether the patient suffers from stomach ache. For this prediction, please click on the appropriate response button. After you have made your prediction, you will be informed whether your patient actually suffered from stomach ache.

Use this feedback to find out what causes the stomach ache your patient is suffering from. Obviously at first you will have to guess because you do not know anything about your patient, but eventually you will learn which foods lead to stomach ache in this patient and you will be able to make correct predictions.

For all of your answers, accuracy rather than speed is essential. Please do not take any notes during the experiment.

If you have any questions, please ask them now. If you do not have any questions, please start the experiment by clicking the mouse.

The training phase consisted of 24 blocks of trials. Each of the 12 trial types was presented once per block in a random order. The order of trials within each block was random, except for the constraint that the first trial of a block could not be of the same type as the last trial of the preceding block. Each trial started with the presentation of either one or two images of foods at the top half of the screen, below the phrase “The patient ate the following food(s)”. For trials with two images, one was located on the left and one on the

right. The left-right allocation of positions for pairs of images was balanced, with each of the two possible arrangements occurring once in each sequential pair of blocks. The sentence "Which reaction do you expect?" was presented below the images. Participants responded by clicking one of two response buttons placed at the bottom of the screen. The left-hand button was labelled "No stomach ache", and the right-hand button was labelled "Stomach ache". As soon as the participant responded, the response buttons and the sentence above them were replaced by a statement and picture showing the outcome of the trial, while the images of the cues and the sentence "The patient ate the following food(s):", remained. When the outcome was stomach ache, the statement was "The patient has stomach ache" and the picture of the sad face was shown. When the outcome was no stomach ache, the statement was "The patient has no stomach ache", and the picture of the happy face was shown. This feedback display remained on the screen for 3000 ms, followed by a 500-ms blank screen after which the next trial began. After 12 blocks of trials, participants were given the option to have a short break.

At test, participants were asked to provide outcome probability ratings for the cues of interest (A, B, X, Y) as well as two cues presented in compound but with no outcome (C, Z).They were given the following instructions:

Now, your task is to judge the probability with which specific foods cause stomach ache in your patient. For this purpose, single foods will be shown to you on the screen.

In this part, you will receive no feedback about the actual reaction of the patient. Use all the information that you have collected up to this time.

The test stage then began. On each trial, the sentence "What is the probability that the food causes stomach ache?" was shown above a single food image. Participants responded by clicking on an 11-point rating scale ranging from 0 (*Certainly not*) to 10 (*Very certain*). The rating scale was located in the lower half of the screen, oriented horizontally. Each cue was

presented twice in a random order, except for the constraint that no cue could be presented twice in succession. For each participant, average probability ratings were calculated for the two presentations of each cue.

For all experiments reported here, training data were divided into a series of successive epochs, with each epoch containing two trials of each type. For all figures and analyses, data were collapsed for equivalent trials (e.g. test trials with C and Z in Experiment 1) by calculating the participant-level mean. Estimates of effect size for each Analysis of Variance (ANOVA) are given as partial eta squared, estimates of effect size for paired *t*-tests are given as Cohen's d_{av}, and estimates of effect size for between-subjects t-tests are given as Cohen's d_s (as recommended by Lakens, 2013). Bayesian *t*-tests were used to evaluate the strength of support for the null hypothesis where appropriate, using a Cauchy prior with a width of .707. The resulting Bayes factors (B_{01}) indicate the level of support for the null and alternative hypotheses. Values higher than 3 can be regarded as support for the null hypothesis, whereas values lower than 1/3 can be regarded as support for the alternative hypothesis.

Results

Figure 1 shows the proportion of trials on which participants predicted stomach ache. Training proceeded smoothly, with participants in the 75% group predicting the correct outcome on 98% of trials during the final epoch, and participants in the 25% group achieving 99% accuracy. To ensure that there were no differences between the groups that might have carried over to the test phase, we conducted a two-way Trial Type x Group ANOVA using ratings from the final epoch for A+, AX+, BY+, and CZ-. This revealed a significant effect of trial type, $F(3, 114) = 674.97$, $p < .001$, $\eta^2_p = .95$, no effect of group, $F(1, 38) = 3.04$, $p = .089$, $\eta^2_p = .07$, and no interaction between these variables, $F < 1$.

Ratings from the test stage are shown in Figure 2. Participants in the 75% group appear to have rated X as a more likely cause of stomach ache than participants in the 25% group, but ratings for other cues were similar in the two groups. A two-way ANOVA comparing ratings for all four cue types in the two groups demonstrated a significant effect of cue type, $F(3, 114) = 226.77$, $p < .001$, $\eta^2_p = .87$, no effect of group, $F(1, 38) = 1.40$, $p < .245$, $\eta^2_p = .04$, and a significant interaction, $F(3, 114) = 3.08$, $p = .030$, $\eta^2_p = .08$. To compare ratings for the blocked and control cues, a separate two-way ANOVA was conducted using only the ratings for B/Y and X in the two groups. This revealed a significant effect of cue type, $F(1, 38) = 10.23$, $p = .003$, $\eta^2_p = .21$, no effect of group, $F(1, 38) = 2.35$, $p = .134$, $\eta^2_p = .06$, and a significant interaction between cue type and group, $F(1, 38) = 4.25$, $p = .046$, $\eta^2_p = .10$. Ratings for X were higher for the 75% group than for the 25% group, $t(38) = 2.09$, $p = .044$, $d_s = .66$, but ratings for B/Y were equivalent in the two groups, $t < 1$, $B_{01} = 3.21$. Ratings for B/Y were higher than for X in the 25% group, $t(19) = 3.18$, $p = .005$, $d_{av} = .90$, demonstrating blocking. There was no significant difference in the ratings for B/Y and X for the 75% group, $t(19) = 1.01$, $p = .324$, $d_s = .21$, although a Bayesian t-test suggested that the evidence in favour of the null result was insufficient, $B_{01} = 2.74$.

Discussion

The results demonstrate that ratings of the blocked cue, X, are dependent on the overall outcome base rate, in accordance with our predictions. This is consistent with an inferential account, according to which the causal status of the blocked cue is ambiguous, and so participants are forced to seek additional evidence when asked to make a probability rating of this cue on test. One source of additional evidence is the extent to which, in general, foods cause an allergic reaction in this patient – the outcome base rate. Therefore, in the absence of any other information, participants tend to assume (at least to some extent) that the blocked cue X is like other cues; it is more likely to be causal when most other foods are causal (the

75% group) than it is when most other foods are safe (the 25% group). At first glance it might appear puzzling that the outcome base rate had no impact on the overshadowing control cues, B and Y. Participants could not be certain of the outcome of B or Y alone, so it seems reasonable to suppose that the chance of both of these cues causing an allergic reaction should have been higher if the patient was allergic to most foods (the 75% group) than if most foods were safe (the 25% group). However, as Livesey et al. (2013) point out, participants' estimates of the probability of the outcome for B and Y should have been affected by the outcome rate to a lesser extent than for X. This is because the causal status of X is completely unknown, whereas participants know that at least one of the overshadowing cues B and Y must be causal. This restricts the range of possibilities, at least somewhat, for the overshadowing cues. Perhaps the procedure used here was sensitive enough to detect the effect of outcome base rate on the blocked cue X, but not the (presumably smaller) effect on the overshadowing controls B and Y.

Our results are also consistent with Rescorla and Wagner's (1972) model, provided we assume that there is a stimulus element that is common to every cue. As with the inferential account, this associative account predicts an effect of the base rate on B and Y, as well as on X. Although we did not see this effect, Rescorla and Wagner's model can be reconciled with our data in a similar way to the inferential account, i.e. by assuming that any between-groups difference for B and Y was not evident in the observed probability ratings because it was smaller than the difference for X.

Another phenomenon in which a blocked cue plays an important role is the "redundancy effect", recently described by Pearce and colleagues (Jones & Pearce, 2015; Pearce, Dopson, Haselgrove, & Esber, 2012; Uengoer, Lotz, & Pearce, 2013). In the redundancy effect, a blocked cue (A+/AX+) is compared not to an overshadowing control,

but to an uncorrelated cue Y, where training is given with two compounds BY+/CY-. Like X, cue Y is informationally redundant because it provides no information about the occurrence of the outcome that is not provided by its companion cues. However, when participants are subsequently asked to give outcome probability ratings for individual cues, they give higher ratings for X than for Y. This effect is often viewed as contrary to Rescorla and Wagner's (1972) model. However, as Vogel and Wagner (2017) have pointed out, the addition of a common element allows Rescorla and Wagner's model to predict the redundancy effect. To understand this idea, it is perhaps instructive to start by considering why, in the absence of the common element, the redundancy effect is not compatible with their model. The simplest version of the model predicts that X will gain little associative strength because it is blocked by A. In contrast, Y is predicted to become associated with the outcome on BY+ trials (cue B is not trained alone, unlike cue A). The associative strength gained by cue Y on BY+ trials will result in a negative prediction error on CY- trials. As a consequence, C should become an inhibitor for the outcome and will protect Y from undergoing complete extinction. The model therefore predicts that Y will maintain some association with the outcome, and will be judged as a more likely cause of the outcome than X, which is the opposite result to that observed. Vogel and Wagner provided simulations that show that the addition of a common element to each trial type results in a reversal of the associative strengths of X and Y (X comes to be rated as higher than Y on test – as observed). This happens for the same reason that this modification predicts weaker blocking; the common element gains associative strength on A+ trials that subsequently mediates a higher expectation of the outcome for X alone.

Similarly, according to the inferential account the lack of information about the causal status of X provides a possible explanation for the redundancy effect. If participants do not know whether the patient will suffer an allergic reaction following consumption of X, then a

common sense approach suggests that they should rate X as a more likely cause of the reaction than Y, because Y was presented in the absence of any outcome on CY- trials. Additionally, in past demonstrations of the redundancy effect, participants may have reason to suppose that the blocked cue is causal. In Uengoer et al.'s (2013) experiments, for example, most training trials ended with the allergic reaction (71% of trials in Experiment 1, 75% in Experiment 2, and 67% in Experiment 3). This is similar to the outcome base rate in the 75% group in Experiment 1 and would, therefore, be expected to promote high probability ratings of cue X on test. It follows that lowering the base rate of the outcome should reduce the size of the redundancy effect, although it seems likely that the redundancy effect would persist due to ongoing uncertainty about the status of X.

It is noteworthy that Vogel and Wagner (2017) also provided simulations demonstrating a profound effect of changing the outcome base rate on the redundancy effect. They showed that decreasing the base rate should reverse the effect, as probability ratings for X fall below those for Y. Hence, we should expect to see higher probability ratings for X than for Y when the outcome base rate is high, but lower ratings for X than for Y when the base rate is low. This is in contrast to the inferential account, which predicts an effect of the outcome base rate on ratings for X, but a redundancy effect in either case. Experiment 2 was designed to test these predictions.

Experiment 2

The purpose of Experiment 2 was twofold. Firstly, we aimed to confirm that varying the overall outcome base rate will affect ratings for a blocked cue, as seen in Experiment 1. Secondly, we sought to determine whether uncertainty about the blocked cue might, at least in part, be responsible for the redundancy effect. The design of the experiment is shown in the bottom half of Table 1. As in Experiment 1, a between-subjects design was used in which

two groups of participants received a single stage of training. For both groups, 12 types of trial were intermixed during the training stage. Four of these trial types were A+, AX+, BY+, and CY-. Cue X served as a blocked cue and Y served as an uncorrelated cue. The remaining 'filler' trials were the same as for Experiment 1, so that participants in the 75% and 25% groups saw the stomach ache on 75% and 25% of trials respectively. Following this training stage, participants were asked to rate the likelihood of stomach ache for A, B, C, X, and Y. As in Experiment 1, it was predicted that ratings for X would be higher in the 75% group than the 25% group. If this effect is a result of a lack of any other information on which to base an assessment of X, then any similar effect for Y should be smaller because participants have evidence that Y is non-causal. The inferential account therefore predicts that the redundancy effect (X>Y) should be larger in the 75% group than the 25% group. The associative account based on Rescorla and Wagner's (1972) model, however, makes a different prediction. We conducted simulations of Experiment 2, with the addition of a common element to all trials, using the parameters provided by Vogel and Wagner (2017). The results of these simulations are shown in Figure 3. They confirmed the prediction that the redundancy effect should be found for the 75% group, but the reverse pattern should be observed in the 25% group. This reversal is due to a marked effect on X of variations in the outcome base rate.

Method

Participants. Participants were 58 (7 male) Psychology undergraduate students at Plymouth University, aged 18-50 (M = 22.07, SD = 7), who received course credit for participation. There were 29 participants in each group.

Materials. The materials and procedure were the same as Experiment 1 except with regard to the number and identity of the stimuli used. The stimuli were 18 images of foods. The foods

were: apple, banana, broccoli, cabbage, cherries, corn, grapefruit, grapes, kiwi, mango, orange, peach, pepper, pineapple, pomegranate, pumpkin, strawberries, and watermelon. They were randomly assigned to each cue (A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, X, Y) for each participant.

Procedure. The procedure was the same as that used in Experiment 1, apart from two changes to the design of the experiment. Firstly, the trial types used in the training stage were those shown in Table 1. Secondly, at test, participants were asked to provide probability ratings for cues A, B, C, X, and Y.

Results

Figure 4 shows the proportion of trials on which participants predicted stomach ache during the training stage. Training proceeded smoothly, with participants learning quickly to predict the occurrence of the stomach ache outcome. During the final epoch participants in the 75% group correctly predicted the outcome on 97% of trials, and participants in the 25% group correctly predicted the outcome on 98% of trials. As in Experiment 1, we wanted to confirm that there were no differences between the groups that might have influenced performance on the test. A two-way Trial Type x Group ANOVA was conducted using ratings from the final epoch for A+, AX+, BY+, and CY-. This revealed a significant effect of trial type, $F(3, 168) = 599.11$, $p < .001$, $\eta^2_p = .92$; no effect of group, $F < 1$; and no interaction between these variables, $F(3, 168) = 1.00$, $p < .393$, $\eta^2_p = .02$.

Figure 5 shows ratings from the test stage. Casual inspection of this figure suggests participants in the 75% group rated X as a more likely cause of the outcome than did participants in the 25% group, but that the two groups gave similar ratings for other cues. A two-way ANOVA comparing ratings for the two groups and for all cues revealed a significant effect of cue, $F(4, 224) = 182.06$, $p < .001$, $\eta^2_p = .77$; no effect of group, $F < 1$;

and a significant interaction, $F(4, 224) = 4.72$, $p = .001$, $\eta^2_p = .08$. To examine the ratings for X and Y, a separate ANOVA was conducted for just these cues. This demonstrated that ratings for X were higher than for Y, $F(1, 56) = 73.95$, $p < .001$, $\eta^2_p = .57$, that the effect of group failed to reach significance, $F(1, 56) = 3.77$, $p = .057$, $\eta^2_p = .06$, and that there was a significant Cue x Group interaction, $F(1, 56) = 5.43$, $p = .023$, $\eta^2_p = .09$. Ratings for X were higher for the 75% group than for the 25% group, $t(56) = 2.85$, $p = .006$, $d_s = .75$, but ratings for Y were equivalent in the two groups, $t < 1$, $B_{01} = 3.65$. Ratings for X were higher than for Y in both the 75% group, $t(28) = 9.12$, $p < .001$, $d_{av} = 1.84$, and the 25% group, $t(28) = 3.92$, $p = .001$, $d_{av} = .86$. Puzzlingly, participants in the 25% group rated B as a more likely cause of the outcome than those in the 75% group, $t(56) = 2.19$, $p = .033$, $d_s = .58$. This effect was not predicted by either of the theories discussed earlier, and is in the opposite direction to any effect we might expect if participants used the outcome base rate to estimate the likelihood of the outcome. It should be noted that we have not replicated this effect and that, although statistically significant, it is an unexpected effect and would not survive a correction for multiple exploratory comparisons.

Discussion

The results of Experiment 2 are a good match for the predictions of the inferential account. Participants rated X as being a more likely cause of the outcome when the outcome rate was high than when the outcome rate was low. This is consistent with the idea that AX+ trials provided little information with respect to the causal status of X, and so participants were required to seek other information on which to base their ratings of this cue. The redundancy effect was larger in the 75% group than the 25% group, although a redundancy effect was still observed in the 25% group. This is the first time that the redundancy effect has been demonstrated when the outcome occurred on a minority of training trials. There are several reasons why participants in the 25% group might have persisted in rating X as a more

likely cause of the outcome than Y, despite the low outcome rate. One possibility is that participants might have inferred that there was a 25% chance of the outcome occurring for X, but that the probability of the outcome for Y was even lower. Alternatively, our manipulation might have been only partially effective because some participants did not use outcome rate to judge the likely status of X, perhaps because participants had some knowledge about the relationship between X and the outcome. For instance, participants might have remembered that each presentation of AX was followed by the outcome, and concluded that the likelihood of the outcome occurring for X was higher than the overall outcome rate. Our data do not allow us to decide between these possibilities, but, like the data from Experiment 1, they support the idea that participants were at least somewhat uncertain about the status of X, and therefore incorporated outcome rate in their judgments.

By contrast, the results of Experiment 2 are only partially consistent with the predictions of Rescorla and Wagner's (1972) model. As the simulations shown in Figure 3 demonstrate, the model only predicts a redundancy effect when the outcome base rate is high. When the base rate is low, probability ratings for Y should be higher than those for X. Additionally, the model predicts a marked effect of changing the base rate not just for X, but for all cues. Our data, however, show an effect of base rate for X only. This is consistent with the inferential account because there is presumably enough information available about other cues for participants to make their probability estimates without incorporating base rates.

Experiment 3

In Experiments 1 and 2, participants' probability ratings for blocked cues were determined partly by the base rate of the outcome. The inferential account of these experiments suggests that this could reflect a lack of certainty about X, but we do not yet have direct evidence for this lack of certainty. To test this idea, in addition to the probability

ratings for each cue, self-reported confidence in those ratings was also collected. The use of multiple test tasks to asses different aspects of participants' knowledge has been used to uncover the psychological processes involved in associative learning in the past. For example, studies have examined cue-outcome memory (e.g., Mitchell, Lovibond, Minard & Lavis, 2006), the ability to categorise cues with outcomes (e.g., Mitchell, Livesey & Lovibond, 2007) and, most relevant to the current study, participants' confidence in their knowledge about the cue-outcome relationship (e.g., Vandorpe, De Houwer & Beckers, 2005). This approach has not yet, however, been used in the context of the redundancy effect.

Experiment 3 was designed primarily to test the idea that participants lack confidence in their probability ratings for blocked cues. This time, as well as probability ratings on test, they were also asked to give their confidence in each rating. For comparison, an overshadowing control (as used in Experiment 1) and an uncorrelated cue (as used in Experiment 2) were included. The full design of Experiment 3 is shown in Table 2. The blocked cue X (A+/AX+) was compared to an uncorrelated cue Y (BY+/CY-) and overshadowed cues P and Q (PQ+). If A+/AX+ training is an effective treatment for producing blocking, then X should be rated as a less likely cause of stomach ache than P and Q. As in Experiment 2, the redundancy effect would be evidenced by higher ratings of X than Y on test. The D- and EF- filler trials were included to prevent participants from developing a strong tendency to predict stomach ache on each trial, or to learn a rule that single foods always cause stomach ache. We expected probability ratings to be lower for the blocked cue than for the overshadowed control (blocking), but higher than for the uncorrelated cue (the redundancy effect). According to the inferential account, differences in confidence ratings for these cues should also be evident. Since the effect of base rate was evident for a blocked cue but not an overshadowed cue (Experiment 1) or an uncorrelated cue (Experiment 2), confidence ratings should be lower for X than for P/Q or Y. However, confidence ratings for

P/Q might also be restricted because, while participants were expected to learn that the outcome will occur when the patient consumed PQ, PQ+ trials do not include any information about which of P and Q is the cause. Confidence ratings for P/Q were therefore expected to be equivalent to or higher than confidence ratings for X.

The inferential account also predicts a relationship between probability and confidence ratings. If the intermediate probability ratings usually given to a blocked cue reflect uncertainty about whether or not it is a cause of the outcome, then participants who are least confident should give ratings that are closest to the middle of the rating scale. The same pattern should be evident for other cues.

Method

Participants. Twenty-one Psychology undergraduate students at Plymouth University took part in this experiment in return for course credit. They were 18-39 years old (M = 20.71, SD = 4.37) and seven were male.

Materials. Ten images served as cues: apple, banana, cherries, grapes, kiwi, mango, orange, pineapple, strawberry, and watermelon. The foods were randomly assigned to each type of cue (A, B, C, D, E, F, P, Q, X, Y) for each participant.

Procedure. During the training stage, participants were presented with 12 blocks of trials, in which each of the different trial types (A+, AX+, BY+, CY-, D-, EF-, PQ+) appeared once.

The procedure was the same as in Experiment 1 with one addition to the test trials. Once the participant had provided a probability rating for the food presented on test, a second rating scale appeared together with the sentence “How confident are you that this rating is

accurate?" This scale ranged from 0 (*Not at all*) to 10 (*Extremely*). For each participant, average causal and confidence ratings were calculated for the two presentations of each cue.

Results

The proportion of trials on which participants predicted stomach ache for each trial type during the training stage is shown in Figure 6. Training proceeded smoothly, with all participants learning quickly to predict the outcome of each trial. During the final epoch, the outcome was predicted correctly on 99% of trials.

Data from the test stage are shown in Figure 7. The left panel of Figure 7 shows the mean ratings of the likelihood that each type of cue (A, B, C, D, E/F, P/Q, X, and Y) would be followed by stomach ache. A one-way ANOVA demonstrated a significant effect of cue type, $F(7, 140) = 132.09$, $p < .001$, $\eta^2_p = .87$. Paired comparisons indicated that ratings for P/Q were higher than for X, $t(20) = 2.67$, $p = .015$, $d_{av} = .58$, demonstrating blocking. The redundancy effect was also observed: ratings for X were higher than for Y, $t(20) = 4.61$, $p < .001$, $d_{av} = 1.00$.

The right panel of Figure 7 shows mean confidence ratings for each cue type. Although participants' ratings of their confidence was generally high, these ratings differed between cue types, $F(7, 140) = 11.78$, $p < .001$, $\eta^2_p = .37$. Of most interest were the ratings for the blocked cue X, the overshadowed cues P/Q and the uncorrelated cue Y. A paired comparison indicated that participants were less confident for X than for Y, $t(20) = 3.18$, $p < .001$, $d_{av} = .69$, but that confidence ratings for P/Q and X did not differ, $t < 1$, $B_{01} = 3.26$. Confidence ratings were also lower for P/Q than for E/F, $t(20) = 2.84$, $p = .01$, $d_{av} = .62$. In other words, participants were less certain about the causal status of one cue from a two-item compound if that compound had been predictive of stomach ache than they were if the compound had been predictive of no stomach ache. We will return to this point in the General

Discussion, but for now it should be noted that there appears to be a difference in how participants learn about the presence and absence of stomach ache.

Figure 8 shows individual participants' probability and confidence ratings for P, Q, X, and Y. As predicted by the inferential account, confidence ratings were lowest when participants gave intermediate probability ratings. To test whether this relationship was statistically significant, each probability rating was converted to a decisiveness rating, based on the magnitude of the difference between the rating and 5. Using this method, probability ratings of 5 corresponded to a decisiveness rating of 0, whereas probability ratings of 0 or 10 produced decisiveness ratings of 5. Spearman's rank correlations were then performed for each cue, to find out whether the decisiveness ratings predicted confidence ratings. We found that this relationship was statistically significant for each of the 10 cues; smallest $\rho = .575$, largest $p = .006$.

Discussion

The first key finding from Experiment 3, with regard to the confidence measure, is that ratings of the blocked cue X were equivalent to those of the overshadowing control P/Q. In fact the same result was observed by Vandorpe et al. (2005) in their Experiment 2, no-information condition. The absence of any difference between these cues is unsurprising given that there is not sufficient information about any of these cues for participants to be sure of whether or not they cause the outcome. However, it is interesting in light of the fact that a blocked cue, but not overshadowing control cues, was affected by the manipulation of outcome base rate in Experiment 1. We had expected that cues about which the participants were uncertain would be most affected by the base rate manipulation in Experiments 1 and 2. This is exactly the pattern we see with regard to the redundancy effect. Hence, participants were more certain about their ratings of the uncorrelated cue Y than the blocked cue X in

Experiment 3, and Y was less affected than X by the base rate manipulation in Experiment 2. Finally, we found that participants gave the lowest confidence ratings for those items to which they gave intermediate probability judgments. This is consistent with the idea that probability judgments are based at least in part on participants' certainty about whether or not each cue causes the outcome, and supports the inferential account of blocking as resulting from a lack of certainty about the status of the blocked cue.

General Discussion

The current experiments suggest that participants are uncertain about the causal status of blocked cues, and that probability ratings for blocked cues are influenced by the overall number of trials on which the outcome is presented. When the overall outcome rate was high, ratings of a blocked cue were higher than when the overall outcome rate was low. The consequence of this variation was that in Experiment 1, the higher outcome rate led to a reduced blocking effect; ratings of the blocked cue X were no lower than to the overshadowing cues P and Q. Also, in Experiment 2, the higher outcome rate was associated with a larger redundancy effect; the blocked cue X was higher than the uncorrelated cue Y regardless of the outcome rate, but this difference was greatest when the outcome rate was high.

These results are difficult to reconcile with Rescorla and Wagner's (1972) model of learning. Although the addition of a common element to each cue allows the model to predict the effect of base rate variation on blocking seen in Experiment 1, it also predicts that the redundancy effect should be reversed when the base rate is low. This is not what we observed in Experiment 2. It should be noted that the failure of Rescorla and Wagner's model to account for our data does not necessarily imply the failure of the broader associative approach; alternative associative models might be a better fit for our data. For instance,

models that explain blocking as a consequence of a decline in the amount of attention paid to blocked cues (e.g. Mackintosh, 1975; Pearce & Hall, 1980) more easily predict the redundancy effect because they allow learning about the blocked cue before any changes in attention take place. However, there are two problems with this approach. The first is that it is not easy to see why the effect of varying the outcome base rate in the present experiments should only apply to the blocked cue, if participants have learned to ignore it. The second problem is that previous attempts to explain the redundancy effect as a consequence of changes in attention have failed to uncover either differences in overt attention for X and Y (Jones & Zaksaite, 2018), or differences in the rate of subsequent learning about those cues (Uengoer, Dwyer, Koenig, & Pearce, 2017).

The results presented here are, however, consistent with a different kind of associative theory that attributes blocking not to a deficit in learning, but to a performance effect. The extended comparator hypothesis (Denniston, Savasatano, & Miller, 2001) states that participants should learn about each cue individually, but that probability ratings should be based on a comparison of the cue in question and other cues with which it is associated. For instance, participants should learn adequately that X is associated with the outcome, but their test ratings should be a consequence not just of this direct associative strength, but also a comparison with A, since A and X occurred together during the first phase of the experiment. Because A is a very good predictor of the outcome, the comparison of A and X should decrease probability ratings for X. This approach not only predicts blocking, but can also predict an effect of varying the outcome base rate that will have different effects for each cue. As well as undergoing comparison with other explicit cues, each cue might be compared to the experimental context. Since the base rate determines the extent to which the context predicts the outcome, it will also determine how the context moderates probability ratings for the cues. The extent to which the context acts as a comparator for each cue will also be

influenced by the strength of the association between the context and the cue, which will be stronger for A than for other cues because A was presented alone on A+ trials. The effects of varying the base rate on each cue are therefore complex, but different effects for different cues are possible and the present results might be accommodated. The problem with this approach is that, while it accounts for the present data adequately, other attempts to evaluate the comparator hypothesis as an explanation for redundancy-effect experiments have proved more challenging for the theory. The comparator hypothesis can explain the redundancy effect itself very simply, because X is consistently followed by the outcome and Y is not. The theory has more trouble accounting for the fact that probability ratings for the predictive cue B are consistently higher than those for X, because both are always followed by the outcome. The theory can only explain this pattern of results because of the relative associative strengths of the comparator cues (A for X, and Y for B), but consequently predicts the abolition of the redundancy effect if the status of the comparator cues changes. Zaksaite and Jones (2017), by contrast, found that probability ratings were higher for B than for X even after successful A-/Y+ re-training.

The current results are easier to explain by assuming that blocking occurs because there is insufficient evidence of a causal relationship between the blocked cue and the outcome. Lovibond et al. (2003) pointed out that, logically, cue X in a blocking design (A+/AX+) is ambiguous. It is not surprising, therefore, that participants were uncertain about cue X (Experiment 3) and based their assessment of this cue on the overall likelihood that any given cue would produce the outcome (Experiments 1 and 2). In both Experiments 1 and 2, it was only the blocked cue X that tracked the overall outcome rate. The effect of the base rate on ratings for X can be thought of as generalisation between cues (Pavlov, 1927). Our proposal is that the outcome rate only generalised to X because this was the only cue about which participants were sufficiently uncertain. With respect to the redundancy effect, this

hypothesis received support from the results of Experiment 3, in which participants indicated that they were more confident in their ratings for an uncorrelated cue Y (trained BY+/CY-) than for X. Confidence in ratings of the blocked cue X did not, however, differ from those of the overshadowed cues P/Q.

The pattern of data with respect to the comparison of blocked cues and overshadowing cues across Experiments 1 and 3, present something of a conundrum. In Experiment 3, confidence in the ratings of the overshadowing cues was no higher than for the blocked cue. In light of this, why was the blocked cue more affected by the base rate of the outcome than the overshadowing cues in Experiment 1? One possibility is that participants only use base rates when there is little or no other information available, even if they are somewhat uncertain. This is consistent with the findings of Kahneman and Tversky's (1973) probability judgment experiment, discussed earlier. In Experiment 1 here, participants could have inferred that at least one of the overshadowing cues must have been causal, and prioritised this information over the base rate. Alternatively, perhaps the confidence rating test in Experiment 3 was simply not sensitive enough to pick up differences between the overshadowing and blocked cues. Objectively, participants do know more about the overshadowing cues than they do about the blocked cue; they know that at least one of the overshadowing cues is causal. It is for this reason that we see an effect of blocking in Experiment 1 (when the outcome base rate was low) and in Experiment 3. A more sensitive forced choice test, in which participants are asked whether they are more certain about their ratings of an overshadowed or a blocked cue might reveal greater certainty about the overshadowing cue.

The findings with respect to the redundancy effect are more straightforward. Participants were less certain about the blocked cue X than they were about the uncorrelated

cue Y (Experiment 3), and the less certain blocked cue X was more affected by the manipulation of the outcome base rate than was cue Y (Experiment 2). In fact, participants were reasonably certain that cue Y did not cause the stomach ache. It is worth considering why participants should be so certain that Y was non-causal. Confidence that Y did not cause the outcome was presumably a result of the omission of the outcome on CY- trials. It appears that participants assumed that, if Y had caused the outcome, then the outcome would have occurred every time Y was consumed. This inference would very simply explain the redundancy effect results seen here. However, low ratings of Y are not a necessary result of uncorrelated (BY+/CY-) training. An alternative possibility is that there was no stomach ache on CY- trials because C prevented its occurrence. The model proposed by Rescorla and Wagner (1972) predicts greater associative strength for Y than for X for exactly this reason (and because it predicts that cue X will accrue very little associative strength). According to their model, the acquisition of inhibitory associative strength by C protects Y from extinction on CY- trials and allows it to retain a modest amount of excitatory associative strength.

For cue Y to gain associative strength in the uncorrelated design, participants must believe that foods (e.g. cue C) are capable of preventing stomach ache. There is, however, little reason to suppose that people will do this. Real-world experience of food poisoning is likely to contain many more examples of foods causing stomach ache than preventing stomach ache. Although there are numerous demonstrations of conditioned inhibition using the allergist task (e.g. Larkin, Aitken, & Dickinson, 1998; Melchers, Lachnit, & Shanks, 2004), these experiments differ from the present experiments in two important ways. Firstly, demonstrations of inhibition typically include instructions that require participants to consider which cues might *prevent* the allergic reaction from occurring. The present experiments contained no such instructions, and it therefore seems plausible that participants simply failed to consider that some foods might prevent stomach ache.

The second issue is that experiments designed to elicit conditioned inhibition usually contain A+/AX- training or similar, where X becomes an inhibitor of the outcome. In this case, inhibition is necessary; since the outcome follows A on A+ trials, its omission on AX- trials must be due somehow to the presence of X. In contrast, BY+/CY- training can be resolved without inhibition, by simply assuming that B is the only cue that results in the occurrence of the outcome. Conditioned inhibition might be less likely to develop when an alternative and simpler causal structure is available, e.g. one in which B is causal, but C and Y are not. In Experiment 2, then, participants may have been confident that Y was not a cause of stomach ache because they did not believe that C was preventative. Experiment 3 provides some evidence for this view. Participants were more confident in their ratings for E/F (from EF-) than P/Q (from PQ+). Why should this be? In both cases, these cues had been presented in compound and followed by a given outcome. However, this finding is easy to explain if we assume that participants did not think that cues could prevent stomach ache. In this case they should have been certain that neither E nor F could be a cause of stomach ache, given that no stomach ache occurred on EF- trials. The alternative, that one cue was causal and the other preventative, would not have occurred to them. Confidence for P/Q was presumably lower because on PQ+ trials no information was available to tell participants whether the stomach ache was caused by P, or by Q, or by both foods. If this analysis is correct, then it suggests that the redundancy effect is caused not just by uncertainty about X, but also by inflated confidence about Y because of the choice of scenario and instructions. Future examinations of the redundancy effect could test this idea, firstly by establishing whether participants can be trained that foods prevent stomach ache in this version of the allergist task, and subsequently by measuring the impact of this training on the redundancy effect. It is possible that the redundancy effect will be reduced under conditions that promote conditioned inhibition.

Another potential topic of future research is the exact nature of the relationship between confidence and probability ratings. According to the inferential account, intermediate probability ratings for blocked cues may be the result of inadequate information and low confidence. However, the results of Experiment 3 do not conclusively demonstrate that participants give medium probability ratings in order to reflect their uncertainty. An alternative, which is consistent with the associative approach, is that confidence ratings are derived from probability estimates. In other words, participants might first gain the knowledge that the probability of the outcome occurring for the blocked cue is neither very high nor very low, and subsequently interpret this probability as indicating a lack of certainty. One way of unpicking this relationship might be to combine confidence ratings with a base rate manipulation, as used in Experiments 1 and 2. If confidence ratings are derived from probability estimates, we would expect any effect of base rate variation on probability ratings to be reflected in corresponding changes to confidence ratings. Alternatively, if confidence ratings are one of the determinants of probability ratings, it should be possible to increase probability ratings by increasing the base rate of the outcome, without increasing participants' confidence that those probability ratings are accurate.

Since we have argued that our results are more consistent with an inferential account of blocking than an associative account, it is worth clarifying what we see as the crucial difference between these two approaches. After all, several commentators (Mitchell, De Houwer, & Lovibond, 2009; Witnauer, Urcelay, & Miller, 2009) have pointed out that complex associative networks can produce outputs that resemble inferential reasoning. Given that inferences must reside somewhere in the brain, and that connectionist networks are model brains (Clark, 1990), it might be argued that there is little real distinction between the inferential and associative approaches. It is not our intention, therefore, to dismiss the idea that associations might be responsible for the blocking phenomenon at some level. Our data

are only problematic for a specific kind of associative theory that explains blocking by proposing that a symbolic representation of the blocked cue fails to become connected to a representation of the outcome. Our data suggest that blocking is the result of uncertainty, rather than a weak connection between mental representations.

In conclusion, the current experiments suggest that participants recognise the ambiguous status of a blocked cue in standard causal learning tasks. Our findings can be explained if the mid-range probability ratings of the blocked cue on test reflect participants' lack of evidence, and hence uncertainty, as to whether there is a relationship between the blocked cue and the outcome. The consequence of this uncertainty is that ratings of the blocked cue will be labile and susceptible to influences beyond the training trials on which that cue was presented. In this case, the degree to which any given cue was likely to produce an allergic reaction – the outcome base rate – influenced ratings of the blocked cue, but no other cue that was tested. Under these circumstances, the outcome base rate will have, and was observed to have, a profound impact on the size of the blocking and redundancy effects observed.

References

Aitken, M. R. F., Larkin, M. J. W., & Dickinson, A. (2000). Super-learning of causal judgements. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology*, *53*, 59-81. doi: 10.1080/027249900392995

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367-405. doi: 10.1037//0033-295X.104.2.367

Cheng, P. W., & Holyoak, K. J. (1995). Complex Adaptive Systems as Intuitive Statisticians: Causality, Contingency, and Prediction. In J. A. Meyer & H. Roitblat (Eds.), *Comparative approaches to cognition* (pp. 271–302). Cambridge: MIT Press.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 2-5. doi: 10.20982/tqmp.01.1.p042

De Houwer, J., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology*, *55*, 965-985. doi: 10.1080/02724980143000578

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

Haselgrove, M. (2010). Reasoning rats or associative animals? A common-element analysis of the effects of additive and subadditive pretraining on blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*, 296-306. doi: 10.1037/a0016603

Jones, P. M. (2018, June 22). Raw data for Jones, Mitchell, and Zaksaite (2018) Uncertainty and blocking in human causal learning. Retrieved from osf.io/ynje9

Jones, P. M., Pearce, J. M. (2015). The fate of redundant cues: Further analysis of the redundancy effect. *Learning and Behavior, 43*, 72-82. doi: 10.3758/s13420-014-0162-x

Jones, P. M., & Zaksaite, T. (2018). The redundancy effect in human causal learning: no evidence for changes in selective attention. *Quarterly Journal of Experimental Psychology,* Advance online publication. doi: 10.1080/17470218.2017.1350868

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251. doi: 10.1037/h0034747

Kamin, L. J. (1969). Selective attention and conditioning. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning* (pp. 42-64). Halifax, Nova Scotia: Dalhousie University Press.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*. doi: 10.3389/fpsyg.2013.00863

Larkin, M. J. W., Aitken, M. R. F., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 1331-1352. doi: 10.1037/0278-7393.24.6.1331

Livesey, E. J., Lee, J. C., & Shone, L. T. (2013). The relationship between blocking and inference in causal learning. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 2920–2925).

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

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

Melchers, K. G., Lachnit, H., & Shanks, D. R. (2004). Within-compound associations in retrospective revaluation and in direct learning: A challenge for comparator theory. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology, 57B*, 25-53. doi: 10.1080/02724990344000042

Mitchell, C. J., Livesey, E. & Lovibond, P.F. (2007). A dissociation between causal judgment and the ease with which a cause is categorized with its effect. *Quarterly Journal of Experimental Psychology,* 60, 400-417. doi: 10.1080/17470210601002512

Mitchell, C. J., Lovibond, P. F., Minard, E., & Lavis, Y. (2006). Forward blocking in human learning sometimes reflects the failure to encode a cue-outcome relationship. *Quarterly Journal of Experimental Psychology*, 59, 830-844. doi: 10.1080/17470210500242847

Pavlov, I. P. (1927). *Conditioned reflexes* (G. V. Anrep, Trans.). Oxford: Oxford University Press.

Pearce, J. M., Dopson, J. C., Haselgrove, M., & Esber, G. R. (2012). The fate of redundant cues during blocking and a simple discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 38*, 167-179. doi: 10.1037/a0027662

Pearce, J. M., & Hall, G. H. (1980). A model for Pavolvian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*, 532-552. doi: 10.1037/0033-295X.87.6.532

Rescorla, R. A., and 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 theory and research* (pp. 64-99). New York, NY: Appleton-Century-Crofts.

Uengoer, M., Dwyer, D. M., Koenig, S., & Pearce, J. M. (2017). A test for a difference in the associability of blocked and uninformative cues in human predictive learning. *Quarterly Journal of Experimental Psychology*. Advanced online publication. doi: 10.1080/17470218.2017.1345957

Uengoer, M., Lotz, A., & Pearce, J. M. (2013). The fate of redundant cues in human predictive learning. *Journal of Experimental Psychology: Animal Behavior Processes, 39*, 323-333. doi: 10.1037/a0034073

Vandorpe, S., De Houwer, J., & Beckers, T. (2005). Further evidence for the role of inferential reasoning in forward blocking. Memory & Cognition, 33, 1047-1056. doi: 10.3758/BF03193212

Vogel, E. H., & Wagner, A. R. (2017). A theoretical note in interpretation of the "Redundancy effect" in associative learning. *Journal of Experimental Psychology: Animal Learning and Cognition, 43*(1), 119–125. doi: 10.1037/xan0000123

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. doi: 10.1037/0096-3445.121.2.222

Zaksaite, T., & Jones, P. M. (2017). The redundancy effect in human causal learning: Evidence against a Comparator Theory explanation. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 3640-3645).

Table 1 – Experiments 1 and 2

Group	Training	Test
75% (Experiment 1)	A+ AX+ BY+ CZ- D+ E+ F- GH+ IJ+ KL+ MN+ OP-	A B C X Y Z
25% (Experiment 1)	A+ AX+ BY+ CZ- D- E- F- GH- IJ- KL- MN- OP-	A B C X Y Z
75% (Experiment 2)	A+ AX+ BY+ CY- D+ E+ F- GH+ IJ+ KL+ MN+ OP-	A B C X Y
25% (Experiment 2)	A+ AX+ BY+ CY- D- E- F- GH- IJ- KL- MN- OP-	A B C X Y

Table 2 – Experiment 3

Training	Test
A+ AX+ BY+ CY- D- EF- PQ+	A B C D E F P Q X Y

Figure 1. The mean proportion of trials on which participants predicted stomach ache during the Training stage of Experiment 1, for the 75% group (left panel) and the 25% group (right panel). Error bars represent the standard error of the mean (SEM).

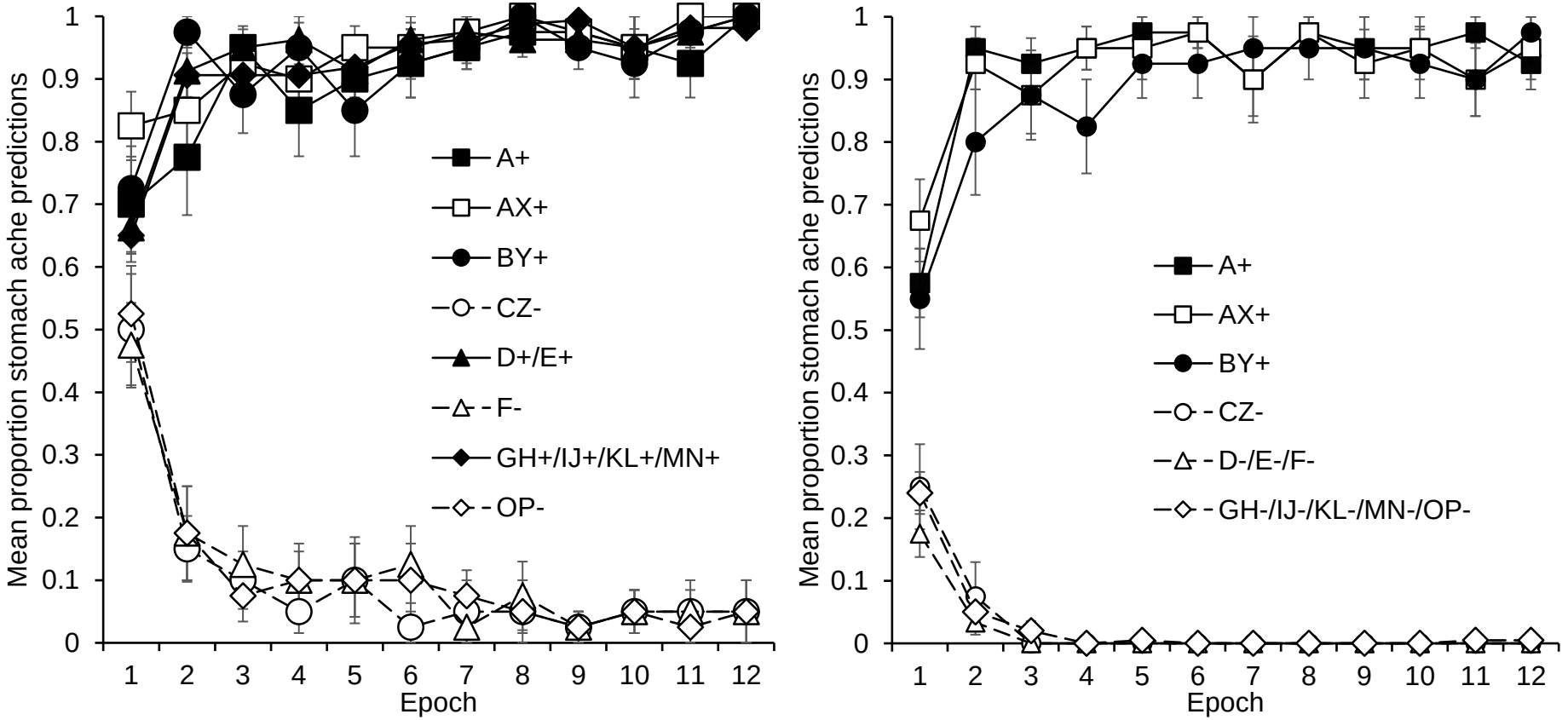

Figure 2. Mean ratings of the probability of stomach ache for each type of cue in the Test stage of Experiment 1. Error bars represent the SEM.

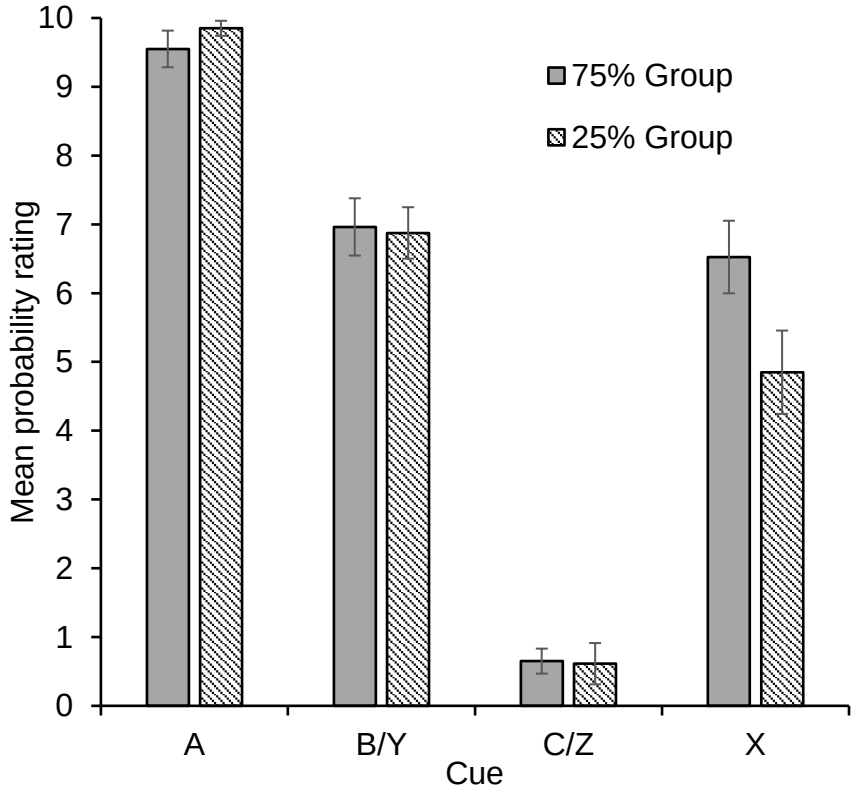

Figure 3. Simulations of Experiment 2 (select cues only; simulations for the 75% group are shown in the left panel, while simulations for the 25% group are shown in the right panel). These simulations were produced using Vogel and Wagner's (2017) modified version of Rescorla and Wagner's (1972) model. The model parameters were the same as those used by Vogel and Wagner: all cues were assigned equal salience, the ratio of β+:β- was 2, and λ was set to 1.

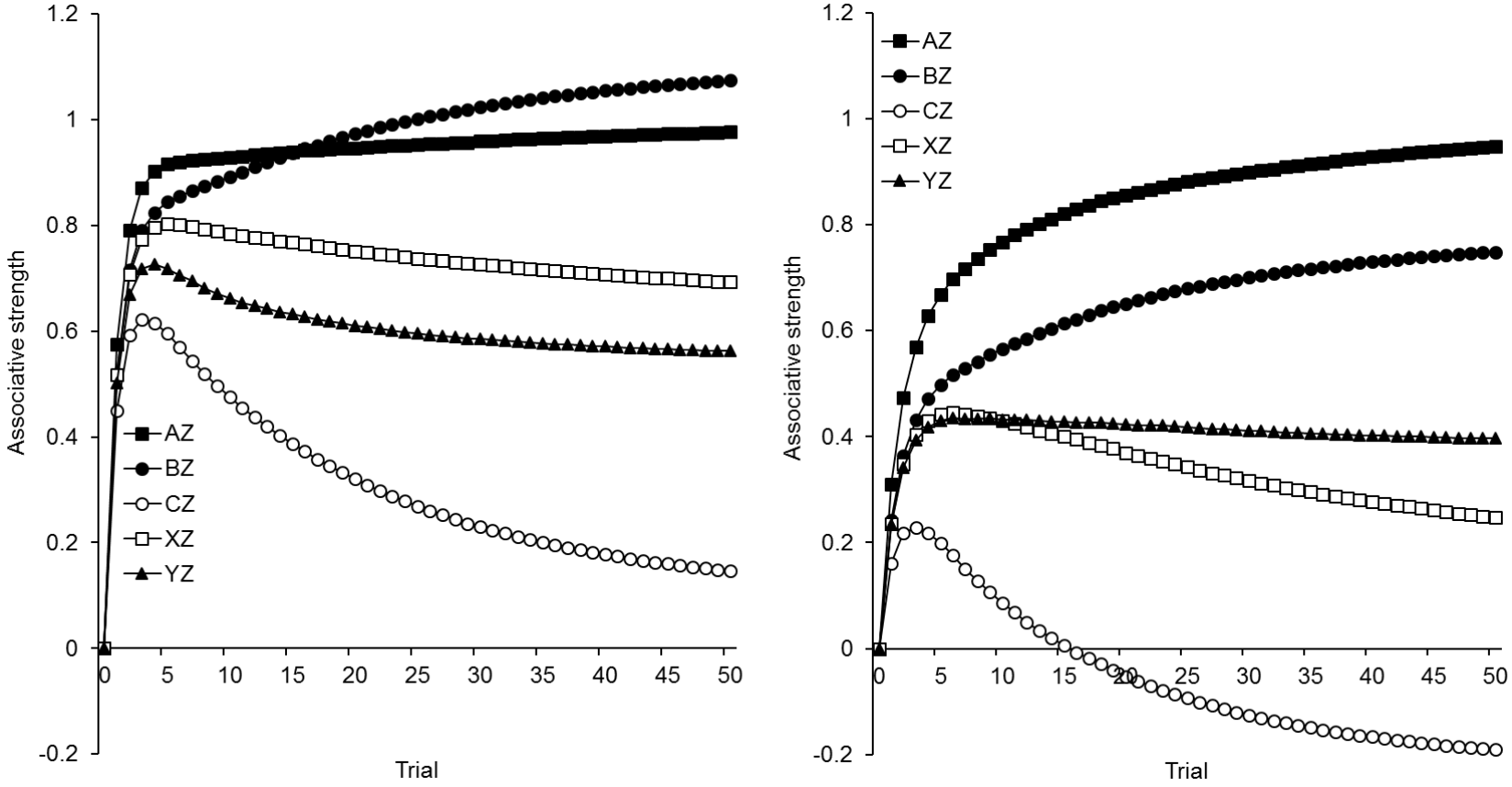

Figure 4. The mean proportion of trials on which participants predicted stomach ache during the Training stage of Experiment 2, for the 75% group (left panel) and the 25% group (right panel). Error bars represent SEM.

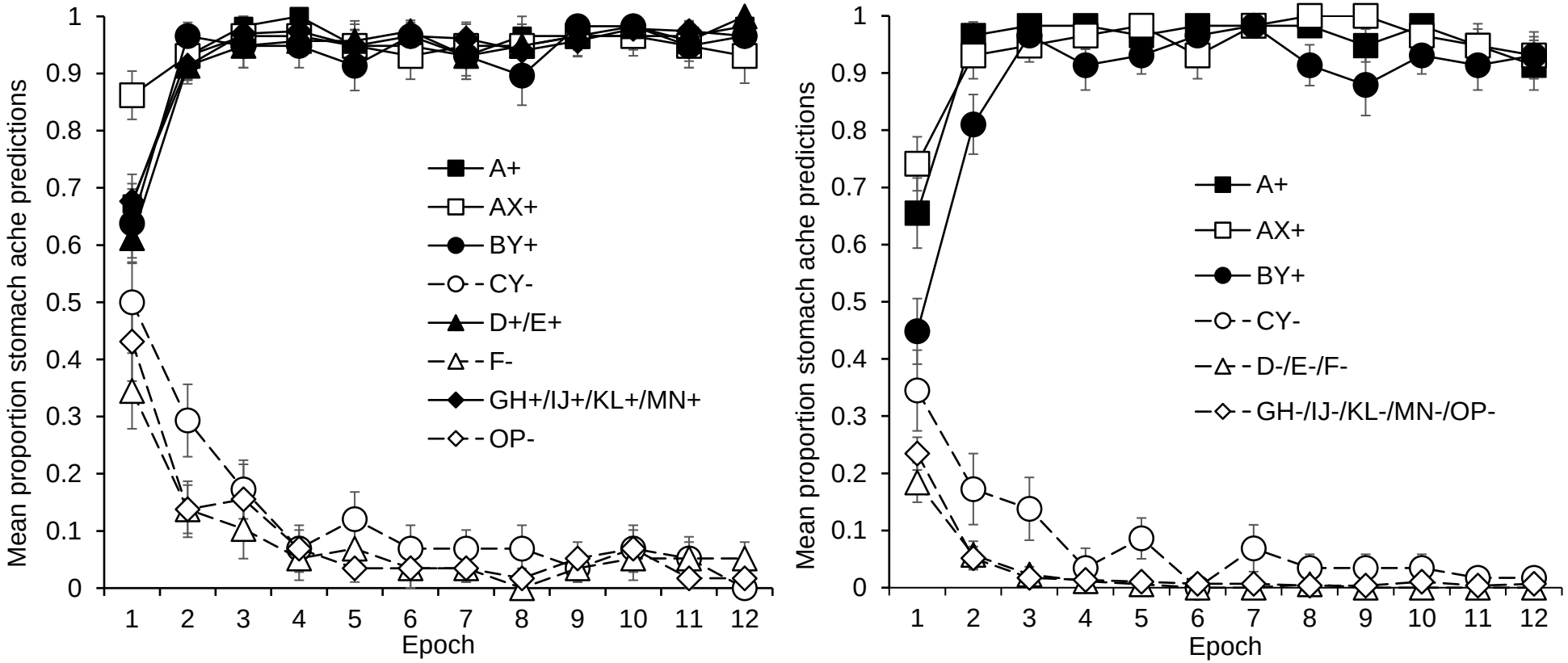

Figure 5. Mean probability ratings for each cue during the Test stage of Experiment 2. Error bars represent SEM.

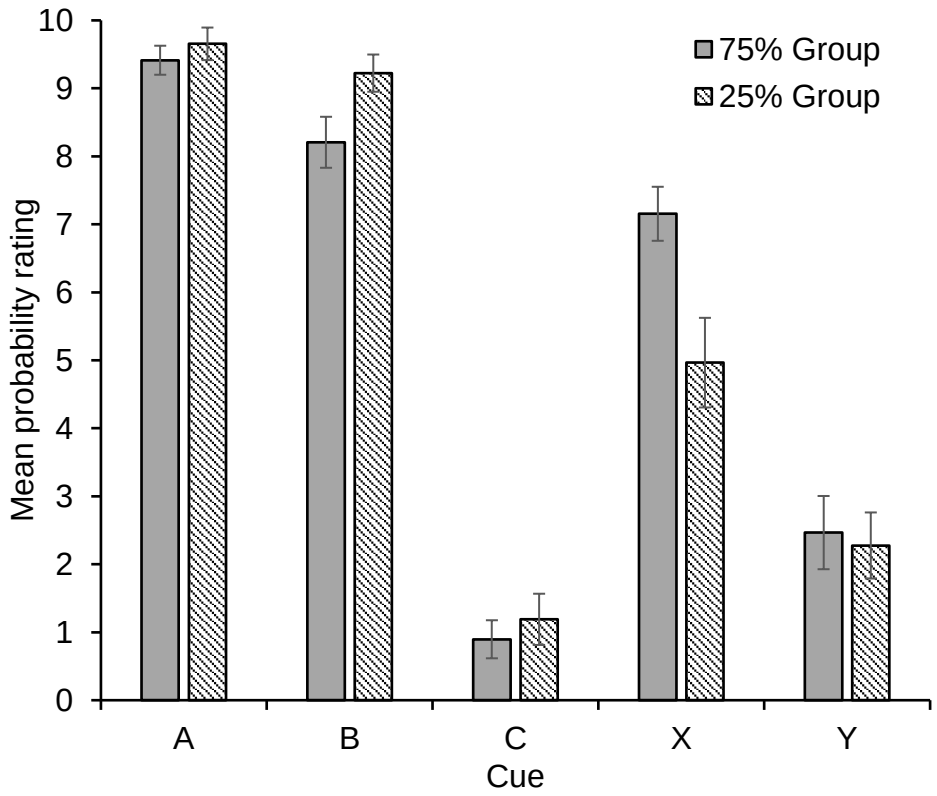

Figure 6. The mean proportion of trials on which participants predicted stomach ache during the Training stage of Experiment 3. Error bars represent the standard error of the mean, adjusted for within-subjects comparisons according to the method described by Cousineau (2005).

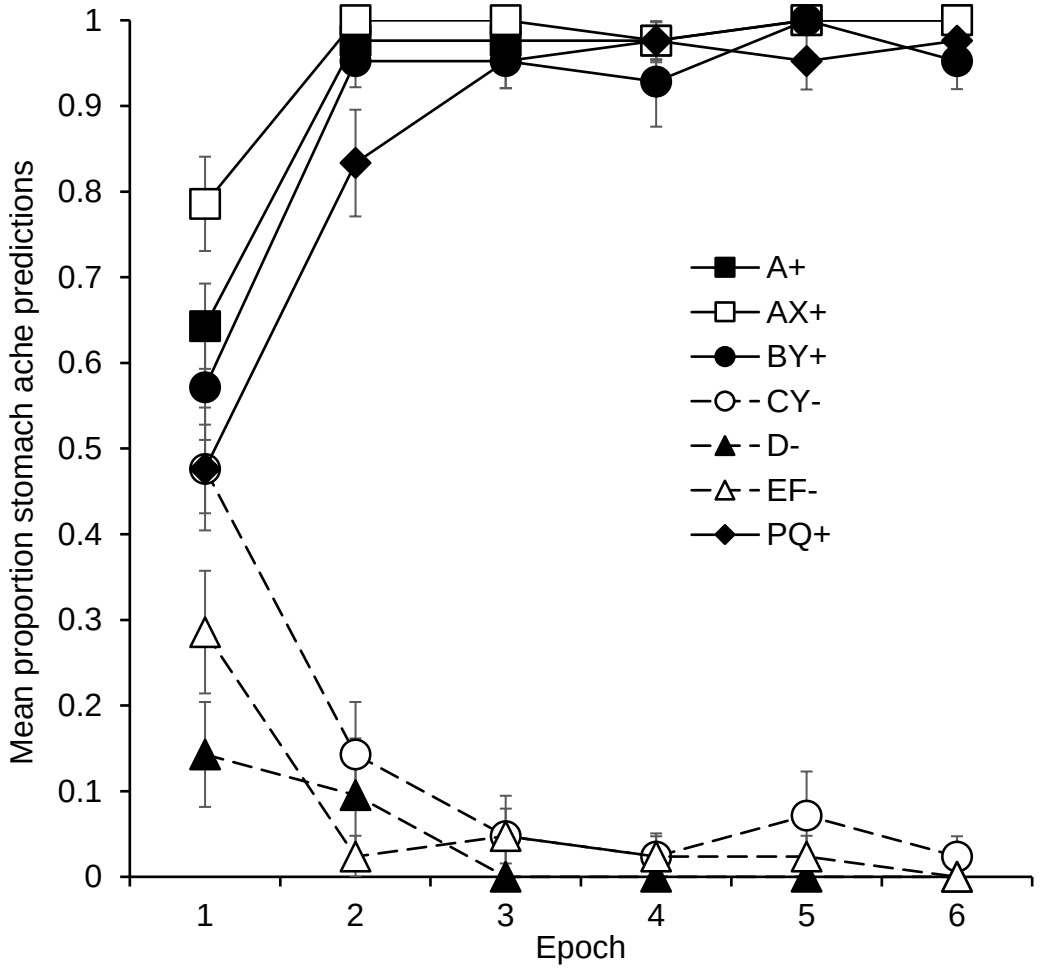

Figure 7. Mean ratings from the Test stage of Experiment 3. Probability ratings are shown in the left panel, and confidence ratings are shown in the right panel. Error bars represent the standard error of the mean, adjusted for within-subjects comparisons according to the method described by Cousineau (2005).

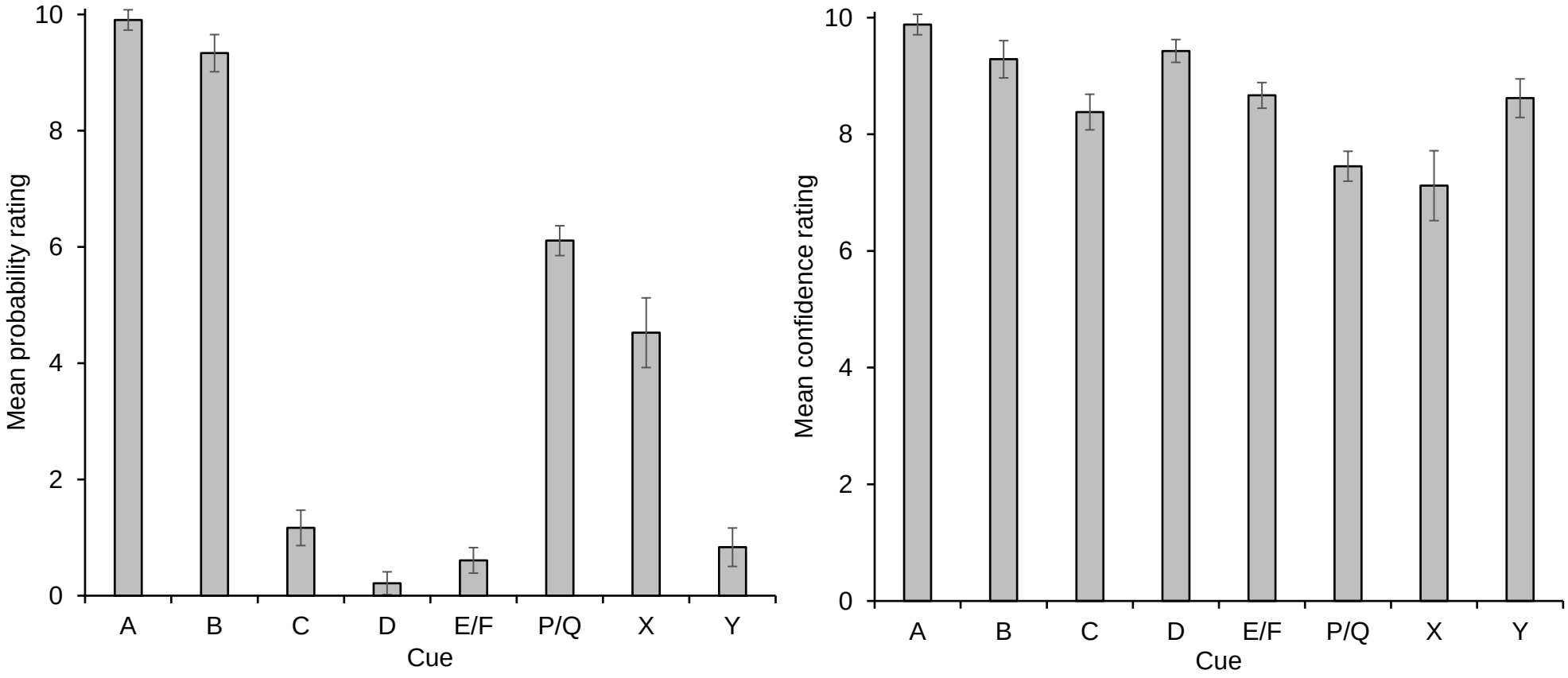

Figure 8. Probability ratings and confidence ratings for each participant in Experiment 3, for P (upper left), Q (upper right), X (lower left), and Y (lower right).