Learn Behav (2012) 40:292–304
DOI 10.3758/s13420-012-0084-4

Modeling attention in associative learning: Two processes or one?

M. E. Le Pelley · Mark Haselgrove · Guillem R. Esber

Abstract Certain studies of associative learning show that attention is more substantial to cues that have a history of being predictive of an outcome than to cues that are irrelevant. At the same time, other studies show that attention is more substantial to cues whose outcomes are uncertain than to cues whose outcomes are predictable. This has led to the suggestion of there being two kinds of attention in associative learning: one based upon a mechanism that allocates attention to a cue on the basis of its predictiveness, the other based upon a mechanism that allocates attention to a cue on the basis of its prediction error (e.g., Le Pelley, *Quarterly Journal of Experimental Psychology*, *57B*, 193–243, 2004). As an alternative, it has been demonstrated that the effects of both predictiveness and uncertainty can be accounted for with only one kind of attention: one that emphasizes the role of prediction (Esber & Haselgrove, *Proceedings of the Royal Society B*, *278*, 2553–2561, 2011). Here, we consider the alternative: whether the effects of predictiveness and uncertainty can be reconciled with a model of learning that emphasizes the role of prediction error (Pearce, Kaye, & Hall, 1982). Simulations of this model reveal that, in many cases, it too is able to account for the influence of predictiveness and uncertainty in associative learning.

M. E. Le Pelley (✉)
School of Psychology, University of New South Wales,
Sydney, NSW 2052, Australia
e-mail: m.lepelley@unsw.edu.au

M. Haselgrove
University of Nottingham,
Nottingham, UK

G. R. Esber
National Institute on Drug Abuse,
Bethesda, MD, USA

Keywords Associative learning · Attention · Classical conditioning · Acquisition · Stimulus preexposure

This article concerns attentional models of associative learning. These models share the basic assumption that that there are two aspects to learning: Organisms learn (1) how much attention to pay to a cue and (2) an association between the representation of that cue and the outcome that follows it. Most important, these two learning processes do not proceed independently, but interactively. That is, learning about the outcome that follows a cue influences the amount of attention that is paid to the cue. And the amount of attention that is paid to a cue, in turn, influences the rate of learning of associations involving that cue.

Attentional theories of associative learning specify that the effectiveness of a cue (attention) is not merely a fixed function of the intrinsic salience of that cue (e.g., its magnitude, intensity, color, etc.) but is also influenced by learning. As a result of experience of the relationships between cues and outcomes, organisms learn to allocate more attention to some cues than to others. Evidence for the influence of such learned attentional processes is now well-established in both humans and nonhuman animals (for reviews, see Le Pelley, 2004; Mitchell & Le Pelley, 2010). In the field of conditioning and discrimination learning, most of these studies take the same general form. In stage 1, animals experience a particular relationship between a cue (or cues) and an outcome (or outcomes). The crucial question is how this experience influences the rate of future learning about that cue in a subsequent conditioning stage (stage 2). If the treatment in stage 1 results in an increase in attention to the cue, we would expect rapid learning about that cue in stage 2 (relative to a suitable control); if the stage 1 treatment results in a decrement in attention, we would expect retarded learning in stage 2.

The big question is exactly how attention and associative learning interact. Two theoretical perspectives have dominated this issue. The first is associated most strongly with Mackintosh's (1975) theory (but see also Kruschke, 2001; Lovejoy, 1968; Sutherland & Mackintosh, 1971; Zeaman & House, 1963) and suggests that a cue maintains attention to the extent that it is the best available predictor of the outcome with which it is paired.

Formally, the degree to which the outcome of the current trial is predicted by cue X is given by the discrepancy between the magnitude of the outcome (λ) and the extent to which cue X predicts that outcome–that is, the associative strength of X (V_X). This discrepancy is the absolute value of the error term $|\lambda - V_X|$. Hence, following each trial, attention to each presented cue X (denoted α_X) is updated according to the following rules:

$$\begin{aligned} \Delta\alpha_X > 0 \text{ if } \left|\lambda - V_X^{\text{before}}\right| < \left|\lambda - V_Y^{\text{before}}\right| \\ \Delta\alpha_X < 0 \text{ if } \left|\lambda - V_X^{\text{before}}\right| \geq \left|\lambda - V_Y^{\text{before}}\right|, \end{aligned} \quad (1)$$

where V_Y is the associative strength of all cues other than X present on that trial. The *before* superscripts indicate that it is the associative strength of the cues before they are updated on trial n that is used when updating α.

Consider a *blocking* design (see Kamin, 1969) in which animals first experience pairings of cue A with an outcome (denoted A+). On later trials, a compound of A with another stimulus is paired with the same outcome (AB+). Evidence suggests that B undergoes a reduction in attention as a result of this treatment (Mackintosh, 1978; Mackintosh & Turner, 1971). This finding is consistent with Mackintosh's theory. Pretraining of A ensures that, on AB+ trials, B is a poorer predictor of the outcome than is A (i.e. $|\lambda - V_B| > |\lambda - V_A|$). Hence, the theory correctly predicts that attention to the *blocked* cue B will decline.

Mackintosh's approach makes intuitive sense; it seems sensible to devote processing resources to cues that have predicted things in the past, since these cues are likely to be useful in predicting things in the future. In a testament to the danger of relying on intuition, however, the second theoretical perspective (which is, in a sense, opposite to Mackintosh's) also sounds intuitively plausible. Pearce and Hall (1980) argued that it makes little sense to devote learning resources to cues whose consequences are already well-established—that is, cues that predict well the outcome with which they are paired. Instead, Pearce and Hall suggested that resources should be devoted to cues to the extent that the outcome with which they are paired is surprising, so that animals will learn more rapidly about the true significance of those cues.

The extent to which the outcome is predicted is given by the absolute value of the summed error term $|\lambda - \Sigma V|$—that is, the discrepancy between the outcome's actual magnitude and the magnitude predicted by all presented stimuli combined. The Pearce–Hall model states that, following each trial n, attention to cue X is updated such that on the trial $n + 1$ we have:

$$\alpha^{n+1} = \left|\lambda^n - \sum V^{n,\text{before}}\right| \quad (2)$$

In fact, Pearce, Kaye, and Hall (1982) noted that Equation 2 makes some rather odd, and clearly incorrect, predictions as a consequence of calculating α solely on the basis of the events of the preceding trial. Consequently, they suggested a refinement to Equation 2:

$$\alpha^{n+1} = \gamma\left|\lambda^n - \sum V^{n,\text{before}}\right| + (1-\gamma)\alpha^n \quad (0 < \gamma \leq 1), \quad (3)$$

where γ is a free parameter that determines how much α is influenced by events on the immediately preceding trial. If $\gamma \approx 1$, α is determined almost solely by the immediately preceding trial, with earlier trials having little effect. Conversely, if $\gamma \approx 0$, α is determined largely by earlier trials, with the immediately preceding trial having little effect. Since Pearce et al.'s (1982) model (hereafter, the PKH model) is a straightforward improvement on the original Pearce–Hall theory, it is the model that we focus on below as being representative of its class.

Consider again an A+, AB+ blocking study. Pretraining of A ensures that the outcome on AB+ trials is unsurprising, since $|\lambda - \Sigma V| = |\lambda - (V_A + V_B)|$ is small. Hence, Equation 3 anticipates that attention to A and (critically) B will be relatively low. So the Pearce–Hall model also explains the empirical finding that a blocked cue undergoes a decline in attention.

Bizarrely then, given that the Mackintosh theory (in which attention is maintained by cues that are consistently followed by the same outcome) and PKH theory (in which attention declines for cues consistently followed by the same outcome) are essentially opposites, both anticipate the observed decline in attention to blocked cues. Other phenomena of learning, however, can discriminate between these theories. The problem is that certain effects in the experimental literature seem to provide unique support to Mackintosh's theory, while others uniquely support the PKH theory. This has led some authors to suggest that multiple attentional processes might contribute to the overall attention paid to a cue (George & Pearce, 2012; Le Pelley, 2004, 2010; Pearce & Mackintosh, 2010).

A recent example highlighting this issue is provided by Haselgrove, Esber, Pearce, and Jones (2010). These studies are particularly interesting because they aimed to provide a contrast between different attentional mechanisms within a single experimental paradigm, by investigating the impact of slight changes in the training design. We shall consider these studies in some detail here, since they form the basis for much of the discussion in this article.

In Experiment 1, Haselgrove et al. (2010) gave rats stage 1 training in which a single cue was presented on each trial.

Cues A and B (the *predictive* cues) were consistently paired with food (A+, B+), while cues X and Y (the *uncertain* cues) were partially reinforced, being followed by food on only half of the trials on which they were presented (X+/−, Y+/−). Rats were then trained on novel discriminations using these cues during stage 2. This provided an index of attention to the cues following stage 1, on the basis of the fundamental assumption of attentional theories of learning that the greater the attention to a cue, the faster it will be learned about. In this test phase, one group was trained with an AY+, BY− discrimination, the solution of which relies on learning about cues A and B. The second group was instead trained with the discrimination AY+, AX−, the solution of which relies on learning about cues X and Y. Haselgrove et al. found that the AY+, AX− discrimination was learnt significantly faster than the AY+, BY− discrimination, indicating that stage 1 training had produced greater attention to uncertain cues X and Y than to predictive cues A and B. That is, rats learned faster about cues when they had a history of being followed by surprising, rather than expected, outcomes. We label this an *uncertainty effect* (see also Hall & Pearce, 1982; Swan & Pearce, 1988; Wilson, Boumphrey, & Pearce, 1992).

Haselgrove et al.'s (2010) Experiment 2 extended this uncertainty effect. Here, stage 1 training was with AB+, XY+/−, such that the consistently and partially reinforced cues were combined into compounds. During stage 2, all rats were presented with an AY+, AX−, BY− discrimination. Rats acquired the discrimination between AY and AX more rapidly than between AY and BY, indicating once again that stage 1 training had produced greater attention to the uncertain cues X and Y than to the predictive cues A and B.

These uncertainty effects are clearly consistent with the PKH model, since this model's guiding principle is that attention will be greater to cues followed by surprising outcomes. However, in Experiments 3 and 4, Haselgrove et al. (2010) showed that a slight change in stage 1 produced very different findings. In Experiment 3, stage 1 training was with AX+, BY+, X−, Y−. Thus, once again, predictive cues A and B were consistently reinforced, while uncertain cues X and Y were partially reinforced.[1] The stage 2 discrimination was as for Experiment 2. Contrary to Experiment 2, however, rats now acquired the discrimination between AY and BY more rapidly than between AY and AX, indicating that Experiment 3's pretraining had produced greater attention to predictive cues A and B than to uncertain cues X and Y.

Stage 1 training in Experiment 4 was with AX+, BY+, A+, B+, X−, Y−, equating exposure to each of the cues (in Experiment 3, cues X and Y were presented more frequently than A and B). Experiment 4 used a different stage 2 procedure to assess the resulting attention to these cues, using them as discriminative stimuli in an instrumental conditioning task. During trials with AY, performance of response 1 (R1) was reinforced with food, but response 2 (R2) was not. During trials with AX and BY, R2 was reinforced, but R1 was not. Haselgrove et al. (2010) found that animals learned the subdiscrimination between AY and BY faster than that between AY and AX. This indicates that stage 1 training in Experiment 4, like that in Experiment 3, resulted in greater attention to A and B than to X and Y.

In summary, while Experiments 1 and 2 revealed uncertainty effects with greater attention to partially reinforced than consistently reinforced cues, Experiments 3 and 4 found *predictiveness effects* in which attention was greater to predictive than to uncertain cues (see also Baker & Mackintosh, 1979; George & Pearce, 1999; Le Pelley & McLaren, 2003; Le Pelley, Suret, & Beesley, 2009; Mackintosh & Little, 1969). The predictiveness effects observed in Haselgrove et al.'s (2010) Experiments 3 and 4 seem to run counter to the ethos of the PKH model. They fit well, however, with Mackintosh's theory. In both experiments, A and B are better predictors of the outcome on AX+ and BY+ trials than are X and Y, and hence, according to Equation 1, attention to A and B will increase, while attention to X and Y will decline. The problem is that Mackintosh's theory cannot explain the uncertainty effect observed in Experiments 1 and 2.

Two kinds of attention?

It would seem, then, that while the results of Haselgrove et al.'s (2010) individual experiments could potentially be explained by either the Mackintosh or PKH model in isolation, neither model can provide a full account of the whole set of data. This led Haselgrove et al. to suggest that their data supported a *hybrid model* in which both Mackintosh and PKH mechanisms contribute to determining the overall attention paid to a cue. The issue then becomes one of specifying how these mechanisms interact and, so, explaining why, under certain conditions, the Mackintosh mechanism dominates (giving predictiveness effects), while under other conditions, the Pearce–Hall mechanism dominates (giving uncertainty effects).

One solution is offered by the hybrid model of Le Pelley (2004, 2010), who noted that Mackintosh's theory determines attention by comparing the *relative* predictiveness of

[1] Given AX+, BY+, X−, Y− training, the outcome following cues X and Y is uncertain when these cues are considered individually (since half of the presentations of X are reinforced and half are nonreinforced). Hence, for consistency with the descriptions of other experiments, we continue to refer to X and Y as "uncertain" cues here. Note, however, that when cues are considered in compound, the outcome is perfectly predictable on each trial (all trials with AX are reinforced, and all trials with X alone are nonreinforced), and hence, this discrimination is soluble, unlike X+/− training, which cannot be solved perfectly since it is impossible to predict the outcome on a given trial with X.

different cues (Equation 1 compares the error term for X with that for other cues Y), while in PKH, attention to a cue is determined by the *absolute* predictiveness of the compound containing that cue (Equation 3 has no comparison of different error terms). So the two theories rely on different properties of a cue—its relative versus its absolute predictiveness—and this distinction makes a clear prediction as to when each mechanism will dominate. Specifically, the Mackintosh mechanism should dominate when predictiveness is established during a pretraining phase involving multiple simultaneously presented stimuli, some of which are more predictive than others, since under these circumstances a comparison of the cues' relative predictiveness will produce a differential change in α. The AX+ and BY+ trials in stage 1 of Haselgrove et al.'s (2010 Experiments 3 and 4 allow exactly this kind of comparison. For example, on AX+ trials, a comparison of the error terms for A and X reveals that A is the better predictor of reinforcement (since X undergoes extinction on X− trials), and hence, the Mackintosh mechanism will favor A over X. Since the Mackintosh mechanism dominates the PKH mechanism in Le Pelley's model, this means that the model correctly anticipates a predictiveness effect (greater attention to A and B than to X and Y) following this training.

Conversely, Le Pelley's (2004, 2010) hybrid approach anticipates that the PKH mechanism will overcome the dominance of the Mackintosh mechanism when two cues do not differ in their relative predictiveness but do differ in terms of absolute predictiveness. This is the case in Haselgrove et al.'s (2010) Experiment 1. On each stage 1 trial, only a single cue is presented, and hence, the presented cue is the best predictor of the current outcome on each trial, such that the Mackintosh mechanism will not differentiate between these cues. However, the cues differ in their absolute predictiveness (A and B have greater absolute predictiveness than do the partially reinforced X and Y), so the PKH mechanism will result in greater attention to the cues with lower absolute predictiveness, X and Y, giving the uncertainty effect observed empirically (see Le Pelley, 2010, for simulations). A similar argument applies to Haselgrove et al.'s Experiment 2, in which cues presented in compound did not differ in their relative predictiveness (e.g., for the XY compound, both X and Y were equally predictive), but compounds AB and XY did differ in their absolute predictiveness.

One kind of attention?

The preceding discussion implies that two classes of behavioral phenomena pervade the learning and attention literature: (1) *predictiveness effects*, in which organisms learn to attend to cues that are the best available predictor of an outcome, and (2) *uncertainty effects*, in which organisms learn to attend to cues that are followed by surprising, rather than expected, outcomes. Furthermore, it has been suggested that two different kinds of attentional mechanism are required to account for these seemingly contradictory effects (Haselgrove et al., 2010; Le Pelley, 2004, 2010; Pearce & Mackintosh; 2010). Recently, this suggestion has been challenged.

Esber and Haselgrove (2011) described a theory of learning that explained predictiveness and uncertainty effects with a single attentional mechanism. In this model, the attention that a cue captures is equal to its salience (α). Following the spirit of Mackintosh's theory, Esber and Haselgrove suggested that a cue acquires salience as a consequence of becoming a predictor of outcomes. The acquired salience of a cue (ε) is a function of the *sum of its associations* with outcomes, regardless of their motivational sign. Thus, a cue associated with both aversive and appetitive consequences might capture more attention than a cue associated with just an appetitive or just an aversive outcome. Esber and Haselgrove employed a variant of the delta rule proposed by Rescorla and Wagner (1972) to update the associative strength of each cue. According to this rule, training of the form AX+, X−, for example, results in A acquiring more associative strength than X. It follows that the acquired salience of the predictive cue A would be higher than that of cue X.

However, these principles alone do not permit the model to predict that a partially reinforced cue can acquire more salience than a consistently reinforced cue (Haselgrove et al., 2010, Experiments 1 and 2). This is because, according to the Rescorla–Wagner model, nonreinforcement produces a weakening of the cue–outcome association. Consequently, each nonreinforced trial of a partial reinforcement schedule ensures that the associative strength of the cue is lower than that of a consistently reinforced cue. It follows that the acquired salience of a partially reinforced cue would be lower than that of a consistently reinforced cue.

However, it has been suggested (Konorski, 1967) that nonreinforcement can be seen as a motivational event in the same way as reinforcement; for example, the omission of an expected aversive or appetitive event might evoke states of relief and disappointment, respectively. With this approach, excitatory associations are modeled as links between a representation of a cue and an outcome (an unconditioned stimulus, US); inhibitory associations are modeled as links between a representation of the cue and a "no-US" representation ($\overline{\text{US}}$). It is assumed that there is an inhibitory relationship between US and $\overline{\text{US}}$ representations such that, if both are activated simultaneously, $\overline{\text{US}}$ activity inhibits activity in the US representation. Hence greater activation of $\overline{\text{US}}$ will tend to produce a reduction in conditioned responding. A given cue might be

followed by the US on some trials and by its absence on others and so, according to Konorski, might develop both excitatory and inhibitory associations. If V_X represents the strength of the excitatory association from cue X to the US, and $\overline{V}_X$ represents the strength of the association to $\overline{\text{US}}$, then conditioned responding to X is proportional to V_X^{net}, where

$$V_X^{\text{net}} = V_{\text{X}} - \overline{V}_{\text{X}} \tag{4}$$

Esber and Haselgrove (2011) adopted the principles of Rescorla and Wagner (1972) and of Konorski (1967) in their model, such that the effect of nonreinforcement was twofold. First, it strengthens the cue → $\overline{\text{US}}$ association, and second, it weakens the cue→US association. The effect of reinforcement was similarly twofold: It strengthens the cue→US association and weakens any cue → $\overline{\text{US}}$ association.

With these assumptions, Esber and Haselgrove's (2011) model can explain, with only one kind of attentional mechanism, both predictiveness and uncertainty effects. Consider training with AX+, X− trials (cf. Haselgrove et al., 2010, Experiment 3). On AX + trials, A→US and X→US associations develop. However, the X→US association is weakened on X− trials; and, because X never gains a significant association with the US, its ability to enter into an X → $\overline{\text{US}}$ association on X− trials will be similarly limited. The X → $\overline{\text{US}}$ association will also be weakened on AX + trials. Consequently, the X→US and X → $\overline{\text{US}}$ associations will be weak, and so too will be the acquired salience of X. In contrast, A will form a strong association with the US, and therefore, the acquired salience of A will be higher than that of X, consistent with the empirical findings of Haselgrove et al.'s Experiments 3 and 4.

Consider now A+, X+/− training (cf. Haselgrove et al., 2010, Experiment 1). As above, the partially reinforced X is sometimes paired with an outcome and sometimes not. However, because there is no better predictor of the outcome and no-outcome on the reinforced and nonreinforced trials, respectively, X can enter into a reasonably strong association with both of these events. In contrast, the consistently reinforced cue A will be associated only with the outcome representation. The parameterization used by Esber and Haselgrove (2011) ensures that, under these circumstances, the influence on acquired salience of X's two associations outweighs that of A's one association. Consequently, it follows that the acquired salience of the partially reinforced cue X will, ultimately, be higher than that of the consistently reinforced cue A. This is consistent with the empirical findings of Haselgrove et al.'s Experiments 1 and 2.[2]

[2] The description of the Esber–Haselgrove model presented here considers only the mechanism that supports increments in cue salience. In keeping with Wagner (1978), Esber and Haselgrove (2011) suggested that the salience of cues declines to the extent that they themselves are predicted (for example, by contextual stimuli).

The PKH model

The preceding discussion has emphasized the role that prediction can have on variations in the processing of stimulus representations. In keeping with the spirit of the Mackintosh model, Esber and Haselgrove (2011) suggested that attention to a cue increases when it is a good predictor of outcomes. By conceiving of a partially reinforced cue as a cue that is reasonably predictive of two outcomes, their model reconciles the influence of predictiveness and uncertainty on stimulus salience with only one kind of attentional mechanism. But the alternative possibility must also be considered: that rather than *prediction* being the sole basis of variations in stimulus attention, perhaps *prediction error* is. In addressing this possibility, it is natural to turn back to the PKH model, since this is an archetypal model in which attention is determined by a single process based on prediction error. In fact, we shall see that this approach can explain rather more than it has previously been given credit for.

The PKH model follows Konorski's (1967) approach, described earlier, wherein inhibition is modeled as the formation of a CS → $\overline{\text{US}}$ association, with Equation 4 giving the net associative strength of a cue, V^{net}. If several cues are presented simultaneously on a trial, the overall expectancy of the US given the presence of these cues is calculated by summing V^{net} across all presented cues. If λ represents the magnitude of the outcome that occurs on this trial, overall error is given by

$$R = \lambda - \sum V^{\text{net}}. \tag{5}$$

If R is positive, this is a trial that will support excitatory learning. Thus, when $R > 0$, the change in the excitatory associative strength of X (ΔV_{X}) is given by

$$\Delta V_{\text{X}} = \alpha_{\text{X}}\beta_{\text{E}}\lambda. \tag{6}$$

If instead $R < 0$, this is a trial that supports inhibitory learning. On such trials, the change in the associative strength of the X – $\overline{\text{US}}$ inhibitory association is

$$\Delta\overline{V}_{\text{X}} = \alpha_{\text{X}}\beta_{\text{I}}|R|. \tag{7}$$

β_{E} and β_{I} are learning-rate parameters for excitatory and inhibitory learning, respectively.

Following each trial, α of each of the presented cues is updated according to Equation 3, although the error term is based on V^{net} rather than V.

The parameters used in the all of the simulations reported below were λ (reinforced trial) = 1, λ (nonreinforced trial) = 0, $\beta_{\text{E}} = \beta_{\text{I}} = .01$, and $\gamma = .01$. The latter value ensures that changes in α occur gradually and smoothly, with the most recent trial exerting only a small influence on α. The starting value of α was .5, as used by Pearce et al. (1982). In all

simulations, the numbers of training trials experienced by the model matched exactly the number experienced by rats in the corresponding empirical study. Simulations were originally run by the first author in Microsoft Visual Basic 6 and then independently verified by the third author in MATLAB. An executable version (PKH_executable.zip) and Visual Basic source code (PKH_code.zip) are available at http://www2.psy.unsw.edu.au/Users/MLepelley/mike.html.

Haselgrove et al. (2010), Experiment 1 Each stage 1 block featured (in random order) two A + trials, two B+ trials, and one trial each of X+, X−, Y+ and Y− . There were 120 such blocks during pretraining. During stage 2, simulations of group AY-AX had AY+ and AX− trials; group AY-BY had AY+ and BY− trials. In each case, the model experienced 192 trials of each in type in random order, with the constraint that no more than two successive trials of the same type could occur.

Figure 1 shows simulation data for this study, using the PKH model. These data, and those in all simulations reported below, are the average of 20 runs of the model, with each run representing a different animal. The only source of variation in each run of the simulation is the trial order during training, which is randomly determined for each run. In fact, this variation has very little effect on the model's predictions. For example, all points in Fig. 1 are subject to standard error of less than .0007, and similar standard errors apply to all other simulations presented in this article. Consequently any ordinal differences represent reliable and significant predictions made by the model.

Figure 1a shows changes in net associative strength (V^{net}) of the cues during stage 1. Unsurprisingly, V^{net} increases rapidly for the consistently reinforced cues A and B and more slowly for the partially reinforced X and Y. Figure 1b shows the corresponding changes in α. Since X is reinforced ($\lambda = 1$) on 50% of the trials and nonreinforced ($\lambda = 0$) on the other 50%, the mean value of λ on each trial with X is .5. Hence, as long as $0 \leq V_X^{net} \leq 1$, the mean value of the error term in Equation 3 will be .5, and so α_X will remain at .5 throughout training (likewise for cue Y).

In contrast, $\lambda = 1$ for all trials with cues A and B. At the start of stage 1, when $V_A^{net} = 0$, the error term for cue A will be (1−0) = 1, and hence, α_A will, at first, rise above its starting value and above $\alpha_{X,Y}$ (likewise for cue B). As V_A^{net} increases, however, the error term on A + trials will decrease, tending eventually to zero since V_A^{net} tends to 1 and, hence, the error term in Equation 3 tends to (1−1) = 0. Consequently, as training continues α_A begins to fall and eventually drops below $\alpha_{X,Y}$, tending to zero.[3] Notably, α_A "lags behind" V_A^{net};

[3] Interestingly, this cross-over in the attention paid to X /Y and A/B with training is also predicted to occur by the Esber and Haselgrove (2011) model. To the best of our knowledge, this prediction remains to be tested.

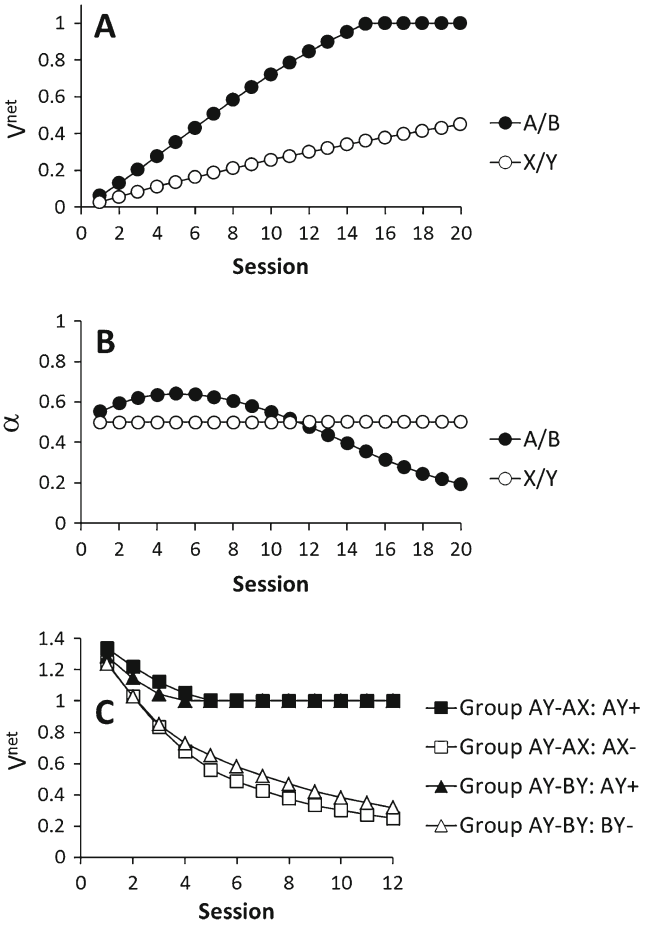

Fig. 1 Results of a simulation of Haselgrove, Esber, Pearce, and Jones's (2010) Experiment 1 using the PKH model. Data shown are the average of 16 runs of the model, with each run representing a different animal. For details of the simulated procedure, see the text. **a** Net associative strength (V^{net}) of the cues presented during stage 1; "A/B" refers to averaged data for cues A and B, which were equivalent during training, and "X/Y" refers to averaged data for cues X and Y, which were also equivalent. **b** α of the cues during stage 1. **c** V^{net} for the compounds presented during stage 2, shown separately for Group AY-AX and Group AY-BY (the two groups were treated identically during stage 1; hence, separate data are not shown in panels a and b)

in Fig. 1, $V_A^{net} \approx 1$ by session 15, whereas α_A is still significantly greater than zero at this point. This is because the low value of γ used in these simulations slows changes in α, relative to changes in V^{net}.

The consequence of all this is that α_X and α_Y are greater than α_A and α_B following stage 1 training. As was noted earlier, this will ensure that learning of an AX+, AY− discrimination proceeds faster than learning of AY+, BY−, since the former relies on learning about the strongly attended X and Y, while the latter relies on learning about the weakly attended A and B. This is confirmed in Fig. 1c, which shows V^{net} for the compounds during stage 2 training of Groups AY-AX and AY-BY. The difference in V^{net} between reinforced and nonreinforced compounds is, at all time points, greater in Group AY-AX than in Group AY-BY;

consistent with the empirical findings, discrimination learning is more rapid in Group AY-AX. The simulated difference is small, but (as noted above) reliable.

Haselgrove et al. (2010), Experiment 2 Each stage 1 block featured (in random order) two AB+ trials, one XY+ trial, and one XY− trial[4]; there were 96 such blocks. In stage 2, the model experienced two AY+ trials, one AX− trial, and one BY− trial in each block, for 64 blocks.

Figure 2a shows changes in V^{net} of the compounds experienced during stage 1, and Fig. 2b shows corresponding changes in α. For the same reasons as for Experiment 1, α_X and α_Y end stage 1 higher than α_A and α_B. This produces more rapid learning of the discrimination between AY+ and AX− during stage 2 (which relies on learning about the strongly attended X and Y) than between AY+ and BY− (which relies on learning about the weakly attended A and B), as is shown in Fig. 2c. Once again, the PKH model explains the empirical data.

Haselgrove et al. (2010), Experiment 3 Each of 96 stage 1 blocks featured AX+ , BY+ , X− , and Y− trials in random order. Stage 2 was as for Experiment 2. Figure 3a shows changes in V^{net} of the cues and compounds experienced during stage 1, and Fig. 3b shows corresponding changes in α. In contrast to Experiment 2, in Experiment 3 the associative strength of the reinforced compounds (AX and BY) approaches asymptote only slowly, because one element of this compound (X or Y) undergoes continual extinction (on X− and Y− trials). Consequently, the error on reinforced compound trials remains at a relatively high value for longer than in Experiment 2, and so the decline in α_A and α_B across stage 1 is less pronounced. Of course, X and Y are also paired with the (relatively) surprising reinforcement on compound trials which will—initially at least—tend to produce increments in α_X and α_Y. However, these increments are offset by the influence of intervening X− and Y− trials: Since X and Y begin with $V^{net} = 0$ and the associative strength of these cues never rises far, the error term on nonreinforced trials remains low, and hence, α_X and α_Y decline on these trials throughout training. The overall effect is that, relative to their starting values, α_X and α_Y will decline more in total than will α_A and α_B across stage 1, such that $\alpha_A, \alpha_B > \alpha_X, \alpha_Y$ at the end of training. This, in turn, will fuel more rapid learning of the discrimination between AY+ and BY− than between AY+ and AX− during stage 2 (Fig. 3c). In other words, according to the PKH model—and consistent with Haselgrove et al.'s (2010) empirical data—the small difference

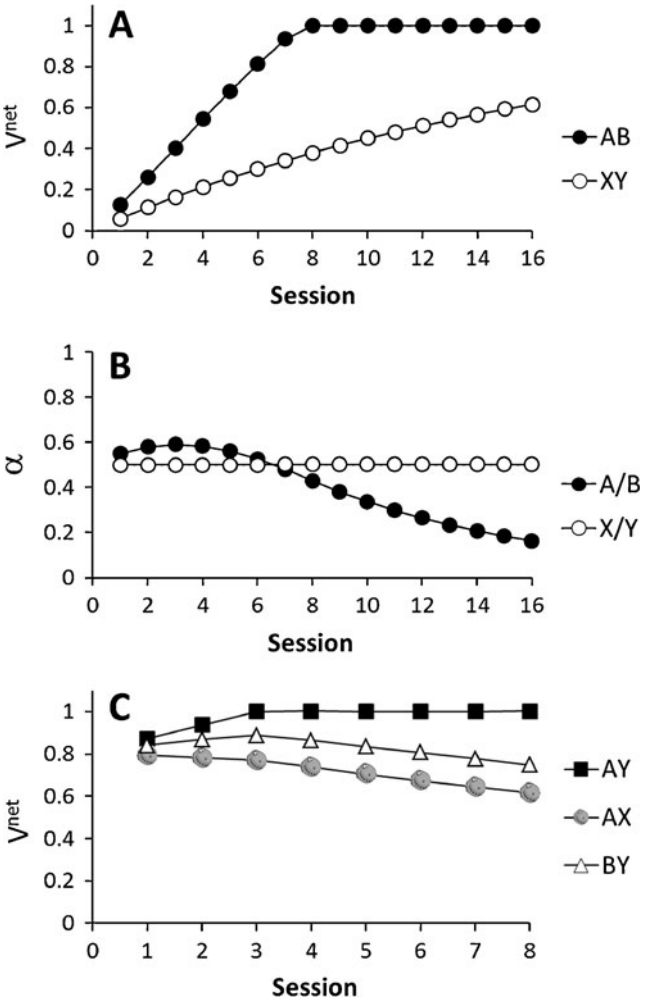

Fig. 2 Results of a simulation of Haselgrove, Esber, Pearce, and Jones's (2010) Experiment 2 using the PKH model. For details of the simulated procedure, see the text. **a** Net associative strength (V^{net}) of the compounds presented during stage 1. **b** α of the cues during stage 1, averaged across equivalent cues. **c** V^{net} for the compounds presented during stage 2

in training between Experiments 2 and 3 can give rise to a different pattern of attention following this training and, hence, a different pattern of discrimination learning in stage 2. (Note that the success of the PKH model with regard to Experiment 3 relies on learning not having reached asymptote in stage 1; we return to this issue in the General Discussion).

Haselgrove et al. (2010), Experiment 4 Recall that Haselgrove et al.'s (2010) Experiment 4 used a test discrimination based on an instrumental rather than a Pavlovian task. PKH, as a model of Pavlovian conditioning, does not allow us to simulate the instrumental test procedure used in this experiment. However, the crucial finding of Experiment 4 (tested using the instrumental procedure) was that the Pavlovian training in stage 1 led to greater attention to A and B than to X and Y, and the PKH model is successful in accounting for this finding.

[4] This renders the simulation equivalent to Haselgrove et al.'s (2010) Group 12-12. Simulations (not reported here) under the conditions of Haselgrove et al.'s Group 8-16 yielded near-identical results.

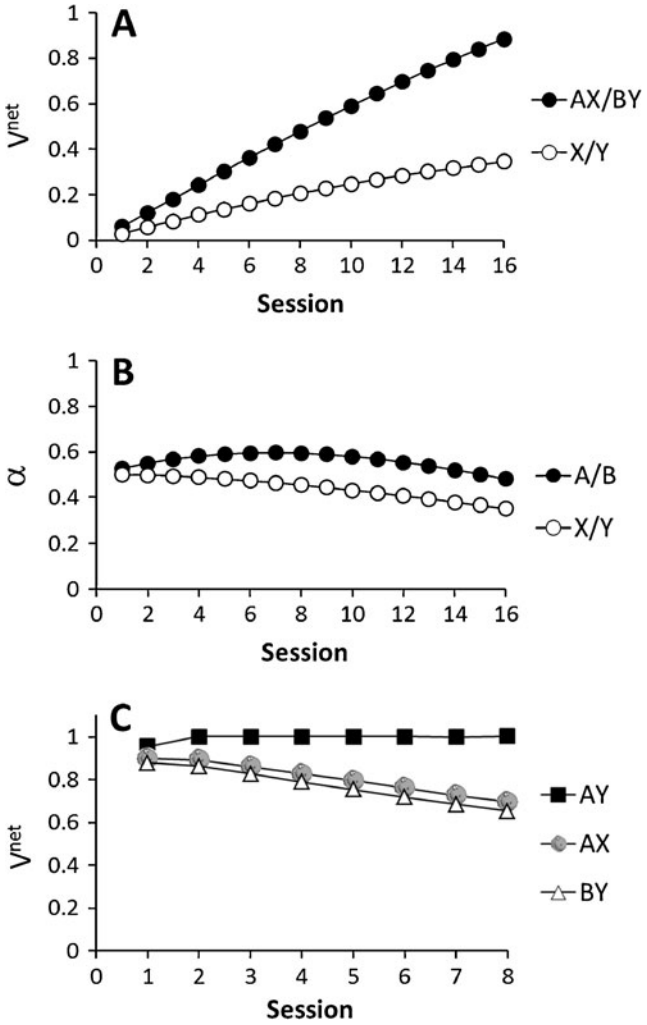

Fig. 3 Results of a simulation of Haselgrove, Esber, Pearce, and Jones's (2010) Experiment 3 using the PKH model. For details of the simulated procedure, see the text. **a** Net associative strength (V^{net}) of the cues and compounds presented during stage 1, averaged across equivalent cues and compounds. **b** α of the cues during stage 1. **c** V^{net} for the compounds presented during stage 2

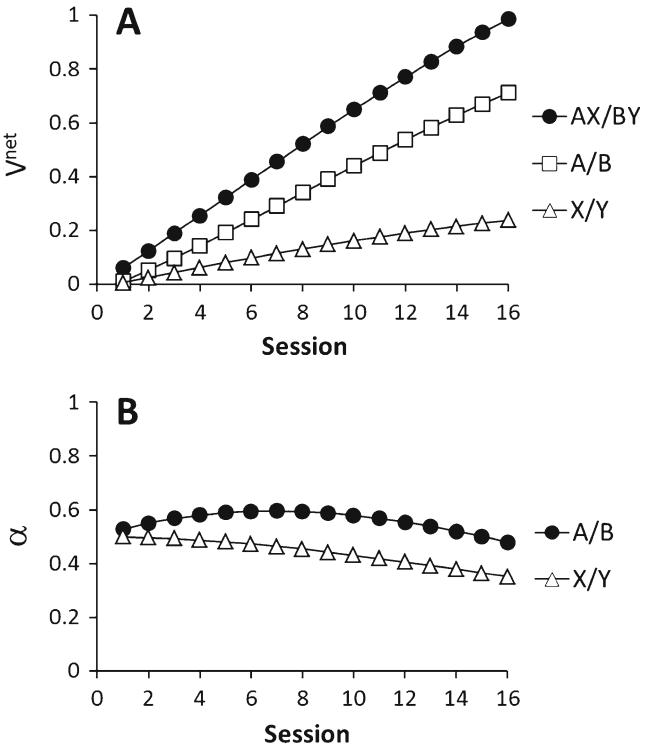

Fig. 4 Results of a simulation of Haselgrove, Esber, Pearce, and Jones's (2010) Experiment 4 using the PKH model. For details of the simulated procedure, see the text. **a** Net associative strength (V^{net}) of the cues and compounds presented during stage 1, averaged across equivalent cues and compounds. **b** α of the cues during stage 1

In the simulation, each of 64 stage 1 blocks featured AX+, BY+, A+, B+, X−, and Y− trials in random order. Figure 4a shows changes in V^{net} of these cues and compounds across training, and Fig. 4b shows corresponding changes in α. For reasons that are very similar to those for Experiment 3, cues α_X and α_Y will decline more in total than will α_A and α_B across stage 1, such that $\alpha_A, \alpha_B > \alpha_X, \alpha_Y$ at the end of training. Hence, the PKH model can account for greater attention to A and B than to X and Y following stage 1 of Experiment 4, such that the former cues would then act as more effective discriminative stimuli in the instrumental conditioning task used by Haselgrove et al. (2010) in their test procedure.

In summary, the PKH model can account for both the predictiveness and uncertainty effects observed empirically by Haselgrove et al. (2010), using only a single attentional process in which attention to a cue is determined by prediction error. It is not surprising that the PKH model can account for uncertainty effects, since these are the kinds of effects that it was originally formulated to explain. What is more surprising is its ability to explain the predictiveness effects observed in Haselgrove et al.'s Experiments 3 and 4. In essence, the model is successful because Haselgrove et al.'s studies compared attention to a consistently reinforced cue with attention to a partially reinforced cue. Since all cues begin with zero associative strength, reinforcement on early trials will be highly surprising, and hence, cues that are more frequently paired with reinforcement will undergo greater increments in attention. While attention will eventually begin to decline once learning renders the outcome on reinforced trials less surprising, if changes in associative strength and changes in attention are sufficiently gradual and the number of training trials is sufficiently small, the model can predict that, overall, a continuously reinforced cue will end training with higher attention than will a partially reinforced cue.

Dopson, Esber, and Pearce (2010) This analysis suggests another recently reported predictiveness effect that the PKH model might be able to address. Dopson, Esber, and Pearce (2010) trained pigeons in stage 1 with a set of discriminations, such as AX+, CX−, and BW+, DW−, in which one cue of

each compound was relevant (A and B consistently signal reinforcement, C and D consistently signal nonreinforcement), while the other cue was irrelevant (W and X are reinforced on half of trials and nonreinforced on the other half); Table 1 shows the full design. Stage 2 then involved training with AW+, AX−, BW−. Dopson, Esber, and Pearce found that the discrimination between AW+ and BW− (which relies on learning about A and B, which were consistently reinforced in stage 1) was learned significantly more rapidly than the discrimination between AW+ and AX− (which relies on learning about W and X, which were partially reinforced in stage 1).

Figure 5 shows results of a simulation of Dopson, Esber, and Pearce's (2010) study using the PKH model. Since learning was rather slow in this study, all rate parameters were reduced; the simulation uses β_E = .003, β_I = .003, and γ = .003. For the same reasons as for Haselgrove et al.'s (2010) Experiments 3 and 4 (described above), the consistently reinforced cues A and B end stage 1 with higher α values than the partially reinforced cues X and Y (Fig. 5b). And this, in turn, fuels more rapid learning of the discrimination between AW+ and BW− than between AW+ and AX− during stage 2 (Fig. 5c); the effect is small but systematic.

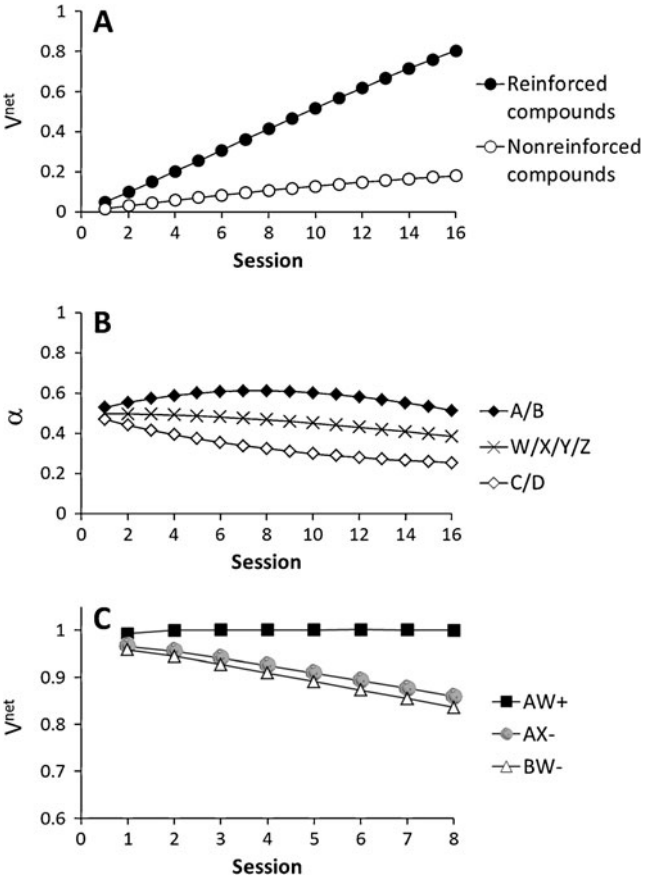

Fig. 5 Results of a simulation of Dopson, Esber, and Pearce's (2010) study using the PKH model. For details of the simulated procedure, see the text. **a** Net associative strength (V^{net}) of the compounds presented during stage 1, averaged across equivalent reinforced compounds and nonreinforced compounds. **b** α of the cues during stage 1, averaged across equivalent cues. **c** V^{net} for the compounds presented during stage 2

General discussion

Several authors have noted that there now exists reliable evidence for both uncertainty and predictiveness effects in studies of conditioning and discrimination learning (see Le Pelley, 2004, 2010; Pearce, George, & Redhead, 1998; Pearce & Mackintosh, 2010). The question then becomes the following: How should these findings be reconciled in an attentional theory of associative learning?

One option is to posit two attentional processes, one fueled by predictiveness (cf. Mackintosh, 1975), and the other fueled by prediction error (cf. Pearce & Hall, 1980). The challenge then is to specify how the components of this hybrid model interact such that the appropriate mechanism dominates under a given set of circumstances; various solutions have been offered (George & Pearce, 2012; Le Pelley, 2004, 2010; Pearce et al., 1998; Pearce & Mackintosh, 2010).

A second approach, taken by Esber and Haselgrove (2011), is to develop a model that emphasizes a single attentional process based on predictiveness (cf. Mackintosh, 1975), but in such a way that the model can also account for uncertainty effects. Esber and Haselgrove achieved this by considering an uncertain cue as a stimulus that is predictive of two or more different outcomes. For example, a partially reinforced cue predicts two different outcomes (presence and absence of reinforcement) relatively weakly, while a consistently reinforced cue predicts one outcome (presence of reinforcement) strongly. It is therefore possible for the total predictiveness of a partially reinforced cue to be either greater or less than that of a consistently reinforced cue. Since it is total predictiveness that determines attention in Esber and Haselgrove's model, the model therefore has the capacity to account for both predictiveness and uncertainty effects.

The present article describes a third potential solution to the problem of reconciling predictiveness and uncertainty effects. Like Esber and Haselgrove's (2011) theory, the PKH model has only a single attentional process, but in this case, it is fueled by prediction error. Clearly, this model is well-suited to explaining uncertainty effects, but in the present article, we have demonstrated that, rather surprisingly, it can

Table 1 Design of experiment by Dopson, Esber, and Pearce (2010)

Stage 1		Stage 2
AX+ CX−	AW+ CW−	
BW+ DW−	BX+ DX−	AW+ AX− BW−
AZ+ CZ−	AY+ CY−	
BY+ DY−	BZ+ DZ−	

also account for certain predictiveness effects that have previously been thought to lie beyond it.

It is true that some of the "successes" of the PKH model with regard to predictiveness effects rely on relatively small differences in associative strengths (see, e.g., Figs. 3 and 5), while alternative models anticipate larger differences (see, e.g., the simulations using Le Pelley's, 2004, model reported by Le Pelley, 2010). However, the important contribution of the present article is to demonstrate that the PKH model *can* account for the ordinal findings of these studies, despite strong claims in the literature that it cannot (Haselgrove et al., 2010; Pearce & Mackintosh, 2010). It is also worth noting that we are simulating associative strength, and not behavior. It remains unspecified exactly how associative strengths translate into conditioned responding (see Rescorla, 2000, for a discussion of this issue). Consequently, it is at least possible that a small difference in associative strengths might, through a nonlinearity in scaling, translate to a relatively large difference in behavior. Thus, it is ordinal differences that are critical in these simulations, and in each case, the ordinal differences anticipated by PKH are in the direction observed empirically.

One objection that may be raised against PKH's account of predictiveness effects is the rather extreme choice of parameters that must be imposed for the size of these effects to even merit discussion (e.g., using $\gamma = .01$ makes changes in attention markedly slow). It should be noted, however, that the making of constrained parametric assumptions is a characteristic shared with many other attentional theories of learning (see Esber & Haselgrove, 2011, Supplemental Materials; Kutlu & Schmajuk, 2012; Le Pelley, 2004).

A more serious criticism of this account becomes apparent when considering Fig. 6e, which depicts PKH's predicted changes in attention to four cues over the course of

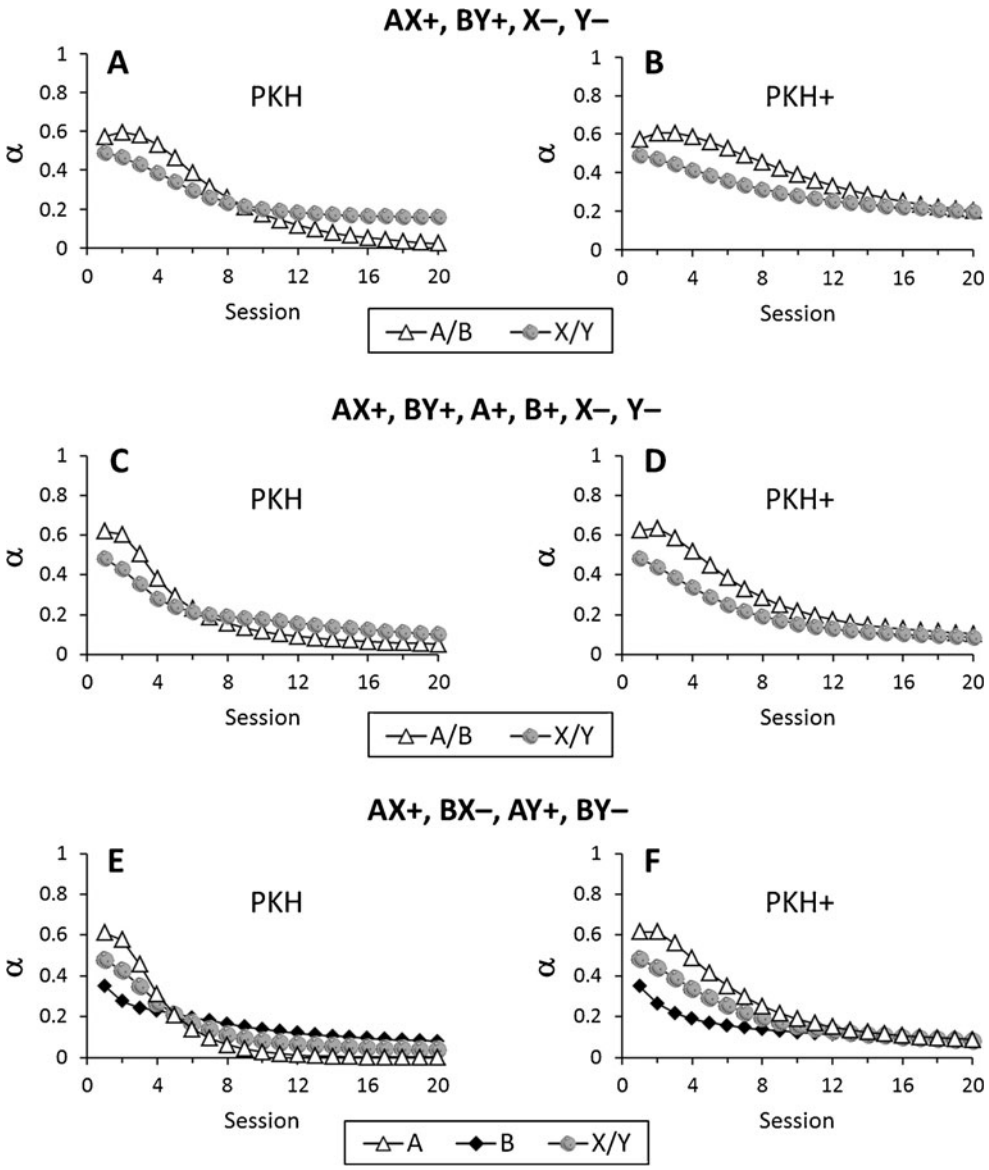

Fig. 6 Values of α from simulations of extended training using the PKH and PKH+ models. **a, b** Training with an AX+, BY−, X−, Y− discrimination as used by Haselgrove, Esber, Pearce, and Jones (2010), Experiment 3. **c, d** Training with an AX+, BY−, A+, B+, X−, Y− discrimination as used by Haselgrove, Esber, Pearce, and Jones (2010), Experiment 4. **e, f** Training with an AX+, BX−, AY+, BY− discrimination as used by Dopson, Esber, and Pearce (2010). In all cases, training was for 20 sessions, with each session containing 20 occurrences of each different trial type. Hence, training in these simulations was more extensive than in the corresponding empirical studies, in order to demonstrate changes in the predictions made by these models over extended training. Parameters used for the two models were identical and were the same as those used for previous simulations (Figs. 1, 2, 3, 4 and 5)

training with an AX+ BX− AY+ BY− discrimination. In this discrimination, A is the best predictor of the outcome, while X and Y are irrelevant stimuli. As can be observed, PKH initially predicts, along with all predictiveness-based models, that A will command greater attention than will X or Y. As training progresses, however, the levels of attention paid to A and X/Y cross over, such that at asymptote, the irrelevant X/Y will be better attended than the predictive A. This prediction is no accident. If the simulations are conducted over an extended number of trials, equivalent cross-overs between A and X are observed in Experiments 3 and 4 by Haselgrove et al. (2010) (Fig. 6a, and c). To the best of our knowledge, this prediction of the PKH model remains empirically untested and so may potentially be correct. Notably, however, this prediction of a cross-over in associability with extended training can be attenuated and, in some cases, eliminated by slightly modifying the PKH model's equation for increments in the CS–US association. This modified version of the model, which we call PKH+, makes correct ordinal predictions for the experiments described earlier (just as PKH does), but these predictions tend to be more robust and less parameter dependent than those made by PKH, although of a similar order of magnitude.

The critical modification in the PKH+ model consists of allowing prediction error to determine not only the amount of processing undergone by the CS, but also that undergone by the US. Parallel influences of prediction error on the effectiveness of CS *and* US representations are commonly assumed by theories of attention in learning (Esber & Haselgrove, 2011; Le Pelley, 2004; Pearce & Mackintosh, 2010). Less well appreciated is the fact that such an assumption already appears, albeit incompletely, in PKH. On trials in which the US is overexpected, the model uses prediction error to modulate processing of both cue and no-US representations: The R term in Equation 7 is a summed prediction error. It seems theoretically arbitrary to assume that, while prediction error modulates US processing when the US is overexpected, it does not when the US is underexpected. This asymmetry is even harder to justify on empirical grounds when evidence points to the diminution of the response to a US that is predicted by a CS, relative to a US that is not predicted (e.g., Donegan, 1981). Furthermore, work by Holland and colleagues has shown that blocking occurs even when the mechanism for reducing CS processing is not available, suggesting variations in US processing contribute to this effect, which occurs when the US is generally underexpected (Baxter, Gallagher, & Holland, 1999).

In this light, it seems reasonable to suggest a more symmetrical version of PKH in which prediction errors resulting from US underexpectation can also modulate the effectiveness of both the CS and US representations. This PKH+ model is identical to PKH, except that Equation 6 is replaced with

$$\Delta V_X = \alpha_X \beta_E R. \quad (8)$$

Figure 6b, d, and e show simulation data with the PKH+ model for Haselgrove et al.'s (2010) Experiments 3 and 4 and the discrimination used by Dopson, Esber, and Pearce (2010). It is clear that this simple modification enhances the predictiveness effects anticipated by PKH and lessens their volatility. The attentional advantage of the predictive cue A over the uncertain cue X across all three discriminations is superior and lasts longer in training to that predicted by original PKH. Equally important, further simulations have confirmed that this amendment does not compromise the PKH+ model's ability to predict the uncertainty effects observed in Haselgrove et al.'s Experiments 1 and 2; the pattern of simulated results for these studies is virtually unchanged.

Moreover, PKH+ is less parametrically constrained than is PKH. Further simulations have confirmed that it can capture the results of the empirical studies described in this article even if γ is relatively high (e.g., $\gamma = .8$), allowing α to change rapidly, since it is strongly influenced by the events of the previous trial. PKH+ can also simulate Dopson, Esber, and Pearce's (2010) data with higher learning rates than can be used with PKH (e.g., $\beta_E = \beta_I = .01$ rather than .003). The executable file and source code available at http://www2.psy.unsw.edu.au/Users/MLepelley/mike.html allow the user to run simulations using either the PKH or the PKH+ model.

Despite its relative success, one fundamental prediction of PKH+ remains: Predictiveness effects should be transient. More specifically, in discriminations of the kind shown in Fig. 6b, d, and e, it predicts that attention to all cues will tend to converge and ultimately decline. Since other models fail to make this prediction (e.g., Le Pelley, 2004) or even predict that attentional differences should increase as learning reaches asymptote (e.g., Esber & Haselgrove, 2011), investigating the fate of predictiveness effects over extended training is an obvious starting point for assessing the merits of the current analysis.

But do all predictiveness phenomena fall out of the PKH model (or PKH+) equally easily? One observation that may pose a challenge for this framework is that cues for the absence of reinforcement seem to command greater attention than do irrelevant cues. After training pigeons with a set of discriminations like those in Table 1, Dopson, Williams, Esber, and Pearce (2010) showed that C and D were better-attended than X or Y in a subsequent test phase similar to that reported by Dopson, Esber, and Pearce (2010). The simulations in Figs. 5b and 6e and f reveal that neither PKH nor PKH+ anticipates this result. Interestingly, the Esber–Haselgrove model similarly fails to anticipate greater attention to C and D than to X and Y following this training, at least with the parameters that Esber and Haselgrove (2011) used in their article. In contrast, the dual-process hybrid attentional model

offered by Le Pelley (2004; see also Pearce & Mackintosh, 2010) is able to account for Dopson, Williams, et al.'s findings. Since C and D are better predictors of nonreinforcement than are X and Y during stage 1, the Mackintosh mechanism in this model ensures that attention to C and D remains greater than to X and Y (see Le Pelley, 2010, for details).

Note, however, that it remains to be determined whether the results of Dopson, Williams, et al. (2010) reflect more than the differential instrumental reinforcement of orienting responses. The results of Dopson, Esber, and Pearce (2010), for example, can be explained by assuming that pigeons are simply more likely to look at A and B than at X and Y during testing because any orienting responses made (initially by chance) toward A and B during training would be consistently rewarded and, thus, more likely to be repeated in the future (see Pearce, Esber, George, & Haselgrove, 2008). Of course, any rewarded orienting response that is made *toward* A and B on AX+ and BY+ trials is identical in form to the "avoidance" response made *away* from X and Y on the same trials. Consequently, it is conceivable that pigeons in the experiment reported by Dopson, Williams, et al. acquired a tendency to look away from X and Y at test and, therefore, were retarded in learning the component of the test discrimination that was based upon these stimuli.

It should also be noted that, in these studies, cues A–D all belonged to the same dimension (e.g., color), while cues W–Z belonged to another (e.g., orientation). We know from studies of the *intradimensional–extradimensional shift effect* (e.g., George & Pearce, 1999; Mackintosh & Little, 1969) that attention generalizes between cues from the same dimension. Recall that the PKH and PKH+ models (and the Esber–Haselgrove model) predict greater attention to the consistently reinforced cues A and B than to the irrelevant cues X and Y in the training used by Dopson, Williams, et al. (2010). If some of the attention commanded by cues A and B generalized to cues C and D (the consistently nonreinforced cues), this could, in theory at least, allow these models to account for the greater attention to C and D than to X and Y observed on test.

Before one fully accepts the implications of the results of Dopson, Williams, et al. (2010) for theories of attention and learning, these matters must be resolved. In contrast, the findings of Haselgrove et al. (2010) are less prone to criticism in these terms: These experiments used only diffuse auditory stimuli that are less susceptible to selective orienting responses, and all cues belonged to the same stimulus dimension. These studies therefore provide a more reliable assessment of attentional processes and, hence, a more solid test-bed on which to compare the predictions made by different attentional models.

These speculations aside, PKH's analysis of predictiveness effects fills in a void in the theoretical space that was previously thought unviable. This leaves us with three possible solutions to the problem of explaining the effects of predictiveness and uncertainty on associative learning: (1) hybrid (multiple-process) attentional theories, (2) a single-process theory based on predictiveness (e.g., Esber & Haselgrove, 2011), or (3) a single-process theory based on uncertainty, such as PKH or PKH+ (see also Kutlu & Schmajuk, 2012). It would seem, then, that currently available empirical evidence may not allow us to answer the question posed in the title of this article (two attentional processes or one?) as definitively as proponents of multiple-process theories have previously argued (Le Pelley, 2004, 2010; Pearce & Mackintosh, 2010). Consequently, developing new experimental procedures in order to identify circumstances in which these different approaches diverge in their predictions will provide a crucial test-bed for future theoretical endeavors.

Author note An executable file and source code (Microsoft Visual Basic 6) for simulations using the PKH and PKH+ models are available at http://www2.psy.unsw.edu.au/Users/MLepelley/mike.html.

References

Baker, A. G., & Mackintosh, N. J. (1979). Preexposure to the CS alone, or CS and US uncorrelated: Latent inhibition, blocking by context or learned irrelevance? *Learning and Motivation, 10,* 278–294.

Baxter, M. G., Gallagher, M., & Holland, P. C. (1999). Blocking can occur without losses in attention in rats with selective removal of hippocampal cholinergic input. *Behavioral Neuroscience, 113,* 881–890.

Donegan, N. H. (1981). Priming-produced facilitation or diminution of responding to a Pavlovian unconditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes, 7,* 295–312.

Dopson, J. C., Esber, G. R., & Pearce, J. M. (2010a). Differences in the associability of relevant and irrelevant stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 36,* 258–267.

Dopson, J. C., Williams, N. A., Esber, G. R., & Pearce, J. M. (2010b). Stimuli that signal the absence of reinforcement are paid more attention than are irrelevant stimuli. *Learning & Behavior, 38,* 337–347.

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, 278,* 2553–2561.

George, D. N., & Pearce, J. M. (1999). Acquired distinctiveness is controlled by stimulus relevance not correlation with reward. *Journal of Experimental Psychology: Animal Behavior Processes, 25,* 363–373.

George, D. N., & Pearce, J. M. (2012). A configural theory of attention and associative learning. *Learning & Behavior*. doi:10.3758/s13420-012-0078-2

Hall, G., & Pearce, J. M. (1982). Restoring the associability of a preexposed CS by a surprising event. *Quarterly Journal of Experimental Psychology, 34B,* 127–140.

Haselgrove, M., Esber, G. R., Pearce, J. M., & Jones, P. M. (2010). Two kinds of attention in Pavlovian conditioning: Evidence for a hybrid model of learning. *Journal of Experimental Psychology: Animal Behavior Processes, 36,* 456–470.

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

Konorski, J. (1967). *Integrative activity of the brain*. Chicago, IL: University of Chicago Press.

Kruschke, J. K. (2001). Towards a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45,* 812–863.

Kutlu, M. G., & Schmajuk, N. A. (2012). *One kind of attention in Pavlovian conditioning*. Manuscript submitted for publication.

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

Le Pelley, M. E. (2010). The hybrid modeling approach to conditioning. In N. A. Schmajuk (Ed.), *Computational models of conditioning* (pp. 71–107). Cambridge: Cambridge University Press.

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

Le Pelley, M. E., Suret, M. B., & Beesley, T. (2009). Learned predictiveness effects in humans: A function of learning, performance, or both? *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 312–327.

Lovejoy, E. (1968). *Attention in discrimination learning*. San Francisco, CA: Holden-Day.

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

Mackintosh, N. J. (1978). Cognitive or associative theories of conditioning: Implications of an analysis of blocking. In H. Fowler, W. K. Honig, & S. H. Pulse (Eds.), *Cognitive processes in animal behavior* (pp. 155–175). Hillsdale, NJ: Erlbaum.

Mackintosh, N. J., & Little, L. (1969). Intradimensional and extradimensional shift learning by pigeons. *Psychonomic Science, 14,* 5–6.

Mackintosh, N. J., & Turner, C. (1971). Blocking as a function of novelty of CS and predictability of UCS. *Quarterly Journal of Experimental Psychology, 23,* 359–366.

Mitchell, C. J., & Le Pelley, M. E. (Eds.). (2010). *Attention and associative learning: From brain to behaviour*. Oxford: 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,* 188–199.

Pearce, J. M., George, D. N., & Redhead, E. S. (1998). The role of attention in the solution of conditional discriminations. In N. A. Schmajuk & P. C. Holland (Eds.), *Occasion setting: Associative learning and cognition in animals* (pp. 249–275). Washington, DC: American Psychological Association.

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

Pearce, J. M., Kaye, H., & Hall, G. (1982). Predictive accuracy and stimulus associability: Development of a model for Pavlovian conditioning. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Acquisition, vol 3* (pp. 241–255). Cambridge, MA: Ballinger.

Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 11–40). Oxford: Oxford University Press.

Rescorla, R. A. (2000). Associative changes in excitors and inhibitors differ when they are conditioned in compound. *Journal of Experimental Psychology: Animal Behavior Processes, 26,* 428–438.

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

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

Swan, J. A., & Pearce, J. M. (1988). The orienting response as an index of stimulus associability in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 4,* 292–301.

Wagner, A. R. (1978). Expectancies and the priming of STM. In S. H. Hulse, H. Fowler & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 177-209). Hillsdale, NJ: Erlbaum.

Wilson, P. N., Boumphrey, P., & Pearce, J. M. (1992). Restoration of the orienting response to a light by a change in its predictive accuracy. *Quarterly Journal of Experimental Psychology, 44B,* 17–36.

Zeaman, D., & House, B. J. (1963). The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency: Psychological theory and research* (pp. 378–418). New York: McGraw-Hill.