Special Issue Article

QJEP

Quarterly Journal of Experimental Psychology
2019, Vol. 72(2) 335–345
© Experimental Psychology Society 2017
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1080/17470218.2017.1339719
qjep.sagepub.com
SAGE

Attention to perceive, to learn and to respond

Geoffrey Hall[1,2] and Gabriel Rodríguez[3]

Abstract

Mackintosh and his collaborators put forward an account of perceptual learning effects based, in part, on learned changes in stimulus salience. In the workshop held to mark Mackintosh's retirement, and published as a special issue of this journal, Hall discussed Mackintosh's theory and proposed his own alternative account. We now want to take the story forward in the light of findings and theoretical perspectives that have emerged since then. Specifically, we will argue that neither Mackintosh nor Hall was correct in his account of the principles that govern how changes in salience occur. Both supposed (in different ways) that such changes depend on the way in which the stimulus (or stimulus element) is predicted by another event. In contrast, theories of attentional learning have stressed the notion that changes in the properties of a stimulus might depend on the way in which it predicts its consequences. These theories have been concerned with attention-for-learning (associability). We now consider how the general principle they both employ might be relevant to the other forms of attention (for perception and for performance) that are, we will argue, critical for the perceptual learning effect.

Keywords

Attention; learning; association; salience; associability; habituation

Received: 12 October 2016; revised: 2 February 2017; accepted: 24 May 2017

N.J. Mackintosh was best known for his contribution to the associative theory of learning. His most cited book on the subject (*The Psychology of Animal Learning*, Mackintosh, 1974) induced one reviewer (Weisman, 1975) to refer to it (or possibly to him) as 'the compleat associationist'. The theoretical approach implicit in that book was made explicit in his next (and next most cited), which was entitled *Conditioning and Associative Learning* (Mackintosh, 1983). Given knowledge only of this background, a newcomer to the field might be a little surprised to learn that much, perhaps the bulk, of Mackintosh's own experimental work was concerned with the role of perceptual processes in learning (and of learning processes in perception) rather than with associative learning directly. But reading a little more widely would soon eliminate the surprise. Thus in a review article published in 1997, summarizing the previous 50 years of work on learning theory, Mackintosh (1997) wrote,

> . . . associative learning theory can explain many aspects of animal behaviour and learning . . . where the elementary theory breaks down the failure is not in the associative analysis, but in how the theory conceptualizes the representation of stimuli. (p. 889)

'Once cognisance is taken of the ways in which stimuli can be represented, associative models are powerful explanatory systems' (Mackintosh, 1997, p. 884).

Mackintosh's first major contribution in this area, work done in collaboration with his mentor Stuart Sutherland, was a theoretical and experimental analysis of how learning processes can modify the attention paid to stimuli, and how changes in attention can control the course of discrimination learning (Sutherland & Mackintosh, 1971). The central idea was that the attention controlled by an aspect of a stimulus display would increase if that aspect gave information about – was a good predictor of – subsequent events. This idea was developed and formalized in an influential paper published a few years later (Mackintosh,

[1]Department of Psychology, University of York, York, UK
[2]School of Psychology, University of New South Wales, Kensington, NSW, Australia
[3]Facultad de Psicología, Universidad del País Vasco (UPV/EHU), Spain

Corresponding author:
Geoffrey Hall, Department of Psychology, University of York, York YO10 5DD, UK.
Email: Geoffrey.hall@york.ac.uk

1975). In this formal theory, each stimulus was given an *associability* parameter (α) that increased when this stimulus predicted an outcome better than other stimuli and decreased when the reverse was true. The associability of a stimulus determined the readiness with which that stimulus would enter into associations. This simple idea proved fruitful and has been long-lived. For example, the special issue of this journal published in 2003, to commemorate Mackintosh's retirement (see also Dickinson & McLaren, 2003), included a demonstration of a transfer effect (later called 'learned predictiveness') in human causal learning that followed directly from the application of Mackintosh's theory (Le Pelley & McLaren, 2003).

The 1975 theory was concerned with the attentional/perceptual changes that occurred as a consequence of reinforced training, with a focus on the discrimination learning procedure, in which different stimuli are associated with different outcomes. But perceptual learning effects can be obtained by, indeed are sometimes uniquely identified as being a consequence of, mere exposure to stimuli (Hall, 1991). The case that has been most studied is that in which appropriately scheduled exposure to a pair of similar stimuli enhances a subject's ability to discriminate between them (or, equivalently, reduces the extent of generalization between them). Mackintosh's later theorizing (developed in association with McLaren; e.g., McLaren, Kaye, & Mackintosh, 1989; McLaren & Mackintosh, 2000) dealt with such effects. In this article, we begin by reviewing this theory and alternatives to it, as they were discussed in the special issue of *QJEP* published in 2003 (see Hall, 2003). We then take the story forward to deal with new experimental findings and theoretical notions that have emerged since then.

The position in 2003

The basic observation that we sought to explain was that animals (or people – see Mitchell & Hall, 2014) are better able to discriminate between similar stimuli when they have previously been exposed to them. For rats, the procedure that has been used routinely involved enhancing the similarity of two flavours (e.g., vanilla and almond) by adding a common taste (e.g., saline) to them. The compounds are then referred to as AX and BX, where A and B are the distinguishing features, and X the features they hold in common. Discrimination between the compounds can be assessed by using one (AX) as the conditioned stimulus (CS) in a flavour-aversion procedure. If the aversion fails to generalize to BX in a subsequent test, we conclude that the subject can discriminate BX from AX. Prior exposure, particularly when it is arranged according to a schedule in which AX and BX are presented in alternation, promotes such discrimination.

Mackintosh and his colleagues attempted to account for this effect in terms of known principles of associative learning. Thus, McLaren et al. (1989) pointed out that, according to standard accounts of conditioning, alternating exposure to AX and BX will establish a set of connections among the various elements of the stimuli. Critically, for present purposes, it will establish inhibitory links between A and B – the presence of A in the AX compound predicts the absence of B and the presence of B predicts the absence of A. These inhibitory links would act to reduce generalization between AX and BX. Without them, the representation of B (activated by the presence of X) would acquire strength during conditioning, and the representation of A (also activated by X) would contribute to performance on test.

There is experimental evidence to confirm aspects of this analysis (e.g., Dwyer & Mackintosh, 2002; Dwyer, Bennett, & Mackintosh, 2001), but Hall (2003) argued that it could not be a complete explanation, as there were procedures in which a perceptual learning effect could be obtained but where this mutual inhibition process could not be responsible. For example, Blair and Hall (2003; see also Hall, Blair, & Artigas, 2006; Rodríguez, Blair, & Hall, 2008) trained a novel stimulus (Y) as a CS and then tested the effects of superimposing the unique element of a preexposed stimulus element on this CS (e.g., they gave BY on test). They demonstrated that B was particularly effective in suppressing the conditioned response (CR) normally controlled by Y. They suggested that the preexposure had enhanced the perceptual effectiveness of the B element with the result that it was better able to interfere with the response governed by the conditioned element, Y. The inhibitory mechanism proposed by McLaren et al. (1989) depends on the use of a procedure in which the AX compound is trained as a CS – the ability of B to inhibit A on test will only be relevant if A has acquired some associative strength. But the result of Blair and Hall shows an effect of intermixed preexposure to AX and BX on a test where this mechanism could not operate.

Hall (2003) suggested, therefore, that an important consequence of alternating exposure to AX and BX was that it enhanced the perceptual effectiveness (or salience) of A and B – or at least that it helped to maintain their salience, given that exposure to the stimuli constitutes an habituation training procedure that might normally be expected to bring about a reduction in effective salience. Subsequent work has generated some experimental evidence to support the proposal that the salience of A and B will be high after alternating exposure to AX and BX. Artigas, Sansa, Blair, Hall, and Prados (2006) gave rats training in which the elements A and B were presented together, and the strength of the association formed between them was assessed by endowing one element with motivational properties and testing the response controlled by the other. It was found that alternating prior exposure to AX and BX allowed subsequent A–B pairings to produce a particularly strong A–B association. Artigas et al. (2006) concluded that preexposure had maintained or enhanced the effective salience of the unique stimulus elements. The broad conclusion reached by Hall (2003) was that here was a potentially critical source of the perceptual learning effect. A learning mechanism that enhances the effective salience specifically of the unique features of a pair of stimuli will, of course, facilitate discrimination between them.

Mechanisms for changing stimulus salience

It now becomes necessary to specify the mechanism responsible for producing changes in stimulus salience. One way of expressing the effect of preexposure is to say that it constitutes an increase in the ability of unique stimulus features to command attention. If subjects attend preferentially to A and B rather than X, then they will be better able to learn a discrimination in which AX signals one outcome and BX another; that is, learning new things about the cues would be enhanced. And a tendency to attend to the unique element when it is superimposed on some other stimulus (as in the experiment by Blair & Hall, 2003, described above) would detract from the ability of that stimulus to evoke its response; that is, performance of a response already acquired would be modified. Given this analysis, it is appropriate to consider the application of theories of attention in associative learning. We begin with the senior example, the forerunner of several similar theories (e.g., George & Pearce, 2012; Le Pelley, 2004; Pearce & Hall, 1980) – that proposed by Mackintosh (1975).

As we mentioned previously, Mackintosh (1975) supposed that the α-value of a stimulus will increase when it is a good predictor of its outcome and decrease when it is not. In preexposure experiments, AX and BX are not followed by outcomes, but we could still apply the theory if we allow a role for within-event learning and consider the extent to which one element of compound stimulus ‘predicts’ the presence of another. In this case, we might say that X is a poor predictor of its partners (appearing sometimes with A and sometimes with B), and so its α-value might decline. On the other hand, a regime of exposure to AX and BX can be seen as consisting of continuous reinforcement of the associations (A-X and B-X) by which the unique features are established as consistent predictors of their associate (X). Accordingly, the α-values of A and B might increase.

This account has the merit of simplicity and arises from a theory developed for, and applicable to, other training situations. But it was not an implication of the theory that Mackintosh developed himself. Perhaps this was because the α parameter was thought of principally as a learning-rate parameter, the associability of the stimulus determining (obviously) how readily it entered into associations.[1] In perceptual learning procedures, the attentional change is one that affects performance, not just the rate of new learning. Whether for this reason or some other, the theory later developed by Mackintosh (McLaren et al., 1989; McLaren & Mackintosh, 2000) made use of a different mechanism, referred to as salience modulation, that explicitly modulates the degree of activation of the node representing a stimulus, and this affects both learning and performance.

In outline, the proposal was as follows. Activation of a representational unit can be produced by external input (the application of its relevant stimulus) and also by internal input derived from activity in other units with which it is associated. Repeated presentation of a stimulus allows its various elements to become linked together and thus increases internal activation. The difference between the internal and external inputs is assumed to determine the magnitude of a ‘boost’ applied to the external input. A novel stimulus lacking associates will receive a substantial boost, a familiar stimulus will not. On the face of things, this mechanism might seem to predict that the effective salience of A and B will necessarily decline during exposure, but the standard exposure procedure, involving alternations of AX and BX, ensures that this is not so. With this procedure, the associations formed between X and A, and between X and B, mean that, on alternate trials, the set of units representing A (or B) is activated associatively, in the absence of the cue itself. According to the learning rule used by the theory, associative activation of units will reduce the strength of a connection between them. Such extinction will reduce the strength of connections among the various elements that constitute stimulus A (or B). Internal inputs will therefore be reduced and thus the salience boost will be restored when the stimulus is next encountered.

This general account has a number of strengths and weaknesses. It is a strength that it directly addresses the important issue of stimulus unitization, long accepted as an important component of perceptual learning (and returning to prominence in some recent accounts; see, for example, Hall, 2008; Mitchell & Hall, 2014). It may also be seen as a strength that it supplies an account of latent inhibition, in that a novel stimulus, being the beneficiary of the salience boost system, will be learned about more readily than a familiar stimulus, the elements of which will be predicted by within-stimulus associations or associations with the context of training. It should be noted, however, that this aspect of the theory can be seen as a strength only by ignoring the evidence indicating that latent inhibition is determined not (or not only) by the extent to which the target stimulus is predicted, but (also) by its past history as a predictor (e.g., Hall & Rodríguez, 2010; Mackintosh, 1975; Pearce & Hall, 1980). And it is, perhaps, a weakness that a critical aspect of the application to perceptual learning depends on adopting a learning rule that is far from secure. That is, it is necessary to assume that simultaneous associative activation of two stimulus elements will weaken an excitatory link between them. But according to Wagner (1981), no learning will occur under these circumstances, and according to others (Dickinson & Burke, 1996; Shevill & Hall, 2004), the link may actually be strengthened.

The account of salience modulation offered by Hall (2003) avoids the problems of the mechanism proposed by McLaren and Mackintosh (2000), but at the expense, it must be admitted, of failing to specify any mechanism at all. Hall's starting point was that the training procedure used to produce perceptual learning can be seen as an instance of habituation consisting, as it does, simply of repeated presentation of a stimulus (or stimuli). In studies of habituation, the dependent variable is usually the magnitude of the unconditioned response (UR) evoked by the stimulus. There are various accounts of the source of the decline in the UR (see, for example, Thompson, 2009) but the consensus is probably that it reflects a reduction in the sensitivity of the node representing the stimulus, or in other words, a decline in the effective salience of the stimulus (see Hall & Rodríguez, 2017). The preexposure phase of a perceptual learning experiment can thus be expected to produce a reduction in the effective salience of the cues. The challenge is to explain why the distinctive features of the cues should be immune to this effect – why exposure to AX and BX should leave A and B with high (or relatively high) salience.

Hall's (2003) answer to this question was to propose that in some circumstances habituation might go into reverse. Specifically, he proposed that different forms of activation of a stimulus node would have different effects. Direct activation, by presentation of the stimulus itself, is, of course, habituation training, and this will result in a loss of sensitivity. Indirect or associative activation of the node (by presentation of an event that has previously been associated with the target stimulus), it was suggested, would have the opposite effect, increasing sensitivity, and restoring or enhancing stimulus salience. The procedure of presenting intermixed trials with AX and BX is one that ensures that the nodes representing A and B will receive repeated associative activation, whereas X will not. Thus, the salience of the common stimulus elements will decline and that of the unique elements will be sustained or enhanced.

Hall's (2003) assumption about reverse habituation seemed to fit the facts (see also Hall, Prados, & Sansa, 2005), but it remains just that – an assumption. No mechanism was specified. Perhaps, this is not surprising in that no mechanism was specified for the basic phenomenon of habituation; again it was simply assumed that that repeated direct activation of a node would reduce its sensitivity. A first step in devising a better-specified theory of the effects of preexposure would be to consider the nature of habituation itself.

Habituation and extinction

Habituation is found in a wide range of procedures (from gill withdrawal in *Aplysia* to the human orienting response [OR]), and it seems likely that a range of mechanisms will be involved. Accordingly, we do not suppose that the account we discuss next will apply to all cases, or even be sole process operating in the experimental procedures of direct interest in this context. At present, it is offered chiefly as an avenue worth exploring. We begin with consideration of seemingly different phenomena – extinction and latent inhibition.

Extinction and latent inhibition

When a CS evokes a CR, standard associative theory attributes this to the ability of the CS to activate the node representing the unconditioned stimulus (US). If the CS is repeatedly presented alone, extinction occurs and the strength of the CR declines. This effect is commonly attributed to an inhibitory process. In the formulation of Konorski (1967; see also Pearce & Hall, 1980), the CS comes to activate a centre ('no-US') that is antagonistic to, and inhibits activation of, the US centre. How, if at all, does the loss of responding evident in extinction relate to the loss of responding that signifies the occurrence of habituation? One possibility is found in the theory proposed by Hall and Rodríguez (2010) as part of an account of latent inhibition.

Hall and Rodríguez (2010) postulated that any novel stimulus will activate the representation of some consequent event, and that this activation is responsible for the UR observed (for stimuli that are powerful enough to evoke some response). Frequently, the UR will be defensive, or for less salient stimuli, investigatory (an OR). The nature of the node activated by the stimulus is taken to be a generic representation of 'an event', and, although it was allowed that even a novel stimulus could, by way of generalization, activate a range of more specific nodes, the one most strongly activated would be that for 'an event'. Given this associative structure, it is natural to assume that the process that operates during extinction of a CR will also operate during nonreinforced presentations of a novel stimulus (Hall & Rodríguez, 2017; see also Lingawi, Westbrook, & Laurent, 2016). That is, the novel stimulus will arouse the expectation of 'an event', but in the absence of any consequence, inhibitory learning will occur to negate this false expectation. In terms of the theory, the stimulus will come to activate a 'no event' representation that inhibits that for 'event'. The stimulus will then be unable to activate the representation responsible for the original UR; that is, habituation will be observed.

Hall and Rodríguez (2010) developed this theory as an account of latent inhibition. They pointed out that, according to the original Pearce-Hall (1980) model, the associability of a stimulus will be high when it is novel, but will decline as it comes to predict its consequence accurately. (It may be noted that this rule for change in associability, α, is quite the opposite of that proposed by Mackintosh, 1975.) Habituation training is a procedure that allows expectation to come to match reality, and accordingly, we can expect that α will decline as a result of nonreinforced exposure to a stimulus. Further learning with this stimulus as a CS will thus proceed slowly (the latent inhibition effect). An implication of this analysis is that habituation training changes not only the representations activated by the stimulus (i.e., reduces the expectation of 'an event'), but it also modifies the properties of the target stimulus itself (produces a decline in α).

Changes in URs

Modification in the properties of the target stimulus can be expected to produce changes in the UR it evokes, a decline in the UR being, of course, the basic behavioural sign of habituation. Studies of various examples of habituation have shown, however, that the decline is not always simple, and that it can differ for different classes of UR.

Some URs appear to reflect the value of the associability of the stimulus. Specifically, when the stimulus is one that evokes an obvious OR, this response has been found to change over the course of repeated presentations in a way that fits with the theory just outlined. Thus, Lovibond (1969) measured the electrodermal component of the OR to a light stimulus. The response declined when the light was presented without a consequence, and it also did so when the light was followed by a tone on all trials. Importantly, however, habituation of the OR was much attenuated when the tone was presented following the light on a random 50% of trials. Thus, the OR declines when the outcome of the stimulus is consistent (either because there is no event or because the outcome is always the same), but the OR is maintained when the subject is uncertain about the outcome of the eliciting stimulus. This aspect of the OR directly tracks the value of α that is expected according to the Pearce-Hall (1980) model, or, for the case in which there is no outcome, the version of the model proposed by Hall and Rodríguez (2010). Experiments with rats (reviewed by Pearce & Hall, 1992) investigating the behavioural OR to the brief illumination of a discrete light have confirmed the generality of these effects.

The functional significance of the OR is widely accepted as being attentional (e.g., Sokolov, 1963; Spinks & Siddle, 1983), controlling changes in information processing. It is no surprise, then, to find that it tracks changes in a parameter (α) that controls 'attention-for-learning'. Equally, it should be no surprise to discover that other URs show different properties. The defensive response (DR) evoked by many novel stimuli seems to function to reduce rather than enhance interaction with the stimulus. After a recent review of the relevant literature, Hall and Rodríguez (2017) concluded that habituation of URs such as the DR obeys different rules from those governing the OR. As with the OR, repeated presentation of the eliciting stimulus on its own produces a decline in the DR. But the effect of presenting the stimulus followed by a consistent consequence is rather different. As we have noted, with this procedure, the OR also shows habituation; in contrast, the evidence reviewed by Hall and Rodríguez showed that the DR tends to be maintained in these circumstances. Thus, this UR does not track associability; rather, it seems to be sensitive to the extent to which its eliciting stimulus is associated with some other event.

That a defensive UR should show this property accords with the interpretation of the effects of the habituation training procedure with which we began this section. Recall that our account of stimulus exposure is in essence a version of extinction in which the initial expectation of an event comes to be negated primarily by acquisition of an opposing association with the 'no event' representation. Prior to training, the stimulus is capable of evoking a response (likely to be defensive, if aversive events are represented among the nodes activated). The response will decline as extinction progresses. But if such extinction is not possible, because the target stimulus is reliably followed by another event, the stimulus will remain effective in evoking a UR. It appears that some aspect of the attention governed by a stimulus declines when the stimulus predicts nothing but is maintained when it predicts a consequence.

Two forms of attention

The developing argument points in the direction of the need to allow (at least) two forms of attention: one concerned with learning and another with performance. There is nothing novel in this proposal. For many years, Holland and his collaborators (e.g., Holland & Gallagher, 1999; Holland & Schiffino, 2016) have made a distinction between attention in learning, the value of which is determined by the predictive validity of the stimulus (prediction error), and attention in action, which is determined by the strength of the stimulus in predicting a consequence. They have provided evidence from studies of the function of the prefrontal and parietal cortex in support of the distinction (e.g., Maddux, Kerfoot, Chatterjee, & Holland, 2007).

We have considered several different ways in which this interpretation might be formalized. For the time being, we will pursue the implications of the following proposal. The formal model of Pearce and Hall (1980; also Hall & Rodríguez, 2010) includes two parameters associated with a CS: associability (α) and salience (S). The first form of attention, attention for learning, follows the rules for change in α put forward in the Pearce-Hall (1980) model. We now propose that S can change too. We will assume that a novel stimulus has a given initial level of salience (determining the attention paid to it at a perceptual level and also its ability to evoke responding). Effective salience is assumed to decline as the extinction process that constitutes habituation occurs, but it is maintained if the target stimulus is associated with some other event. We now explore the implications of these notions in the context of attempting to provide an explanation of perceptual learning effects.

Application to perceptual learning

In what follows, we apply the principles just outlined to the effects of preexposure to the usual stimuli, AX and BX. These principles concern the changes that will occur in S and α as a consequence of what a stimulus predicts. In this form of preexposure the stimuli are, of course, presented without consequences. Earlier in this article, in considering how to apply Mackintosh's 1975, theory to these procedures, we allowed it the freedom of considering within-stimulus associations rather than stimulus-consequence associations. For the time being, however, we will restrict ourselves to a more literal interpretation, considering the changes that occur to A, B and X as a result of intermixed exposure to AX and BX, when each is followed by no other event. The observation of central interest is that generalization between AX and BX is reduced by this procedure, and to a greater extent than when AX and BX are presented in separate blocks of training.

Outline of the model

We take as our starting point the model formalized by Hall and Rodríguez (2010) for nonreinforced stimulus preexposure. According to this, a novel stimulus will evoke the expectation that some event will follow; that is, there is a stimulus–event association that has some initial strength (V_{event}). This expectation is contradicted, in nonreinforced preexposure, by the fact that no event follows the stimulus. Such exposure results in the development of a stimulus–no event association ($V_{\text{no event}}$) that acts to oppose the existing stimulus–event association. Its growth over successive trials is given by the following equation

$$\Delta V_{\text{no event}} = S\,\alpha\,\lambda_{\text{no event}} \tag{1}$$

where S represents the stimulus salience, α is the associability and $\lambda_{\text{no event}}$ represents the inhibitory reinforcer. This equation exactly parallels that used by Pearce and Hall (1980) to describe the formation of CS–no US associations during extinction.

The critical feature of the present account that distinguishes it from that of Pearce and Hall (1980) (and of Hall & Rodríguez, 2010) is that here it is assumed that both the multiplicative factors (S and α) that determine the processing received by the stimulus are variable. In line with the original Pearce-Hall model, the value of α changes according to this equation

$$\alpha^n = \left|\lambda_{\text{event}} - \left(\sum V_{\text{event}} - \sum V_{\text{no event}}\right)\right|^{n-1} \tag{2}$$

where the associability of the stimulus on trial n, α^n, is determined by the absolute value of the discrepancy between λ_{event} (which will be zero during nonreinforced preexposure trials) and the total strength of the expectation that some event was going to occur ($\Sigma V_{\text{event}} - \Sigma V_{\text{no event}}$) as determined by all the stimuli present on trial $n-1$. We now add the further proposal that stimulus salience can also change and will do so according to the associative value of the stimulus. We suggest that salience depends on the ability of the stimulus to activate the expectancy of occurrence of some event. A novel intense stimulus will arouse such an expectation readily, but this will diminish with nonreinforced exposure. We have attempted to capture this notion with the following equation

$$S^n = \left|V_{\text{event}} - V_{\text{no event}}\right| \tag{3}$$

where the salience (S) of a stimulus on a given trial, n, is equated with the net strength with which it activates the expectation that some event is going to occur on that trial ($V_{\text{event}} - V_{\text{no event}}$). Note that the use of the absolute value in this case allows the magnitude of S to be independent of whether the stimulus anticipates the occurrence of some event (i.e., when $V_{\text{event}} - V_{\text{no event}}$ is positive) or its absence (i.e., when $V_{\text{event}} - V_{\text{no event}}$ is negative).

These assumptions are compatible with the notion (as adopted, for example, by the original Pearce-Hall, 1980, model) that the salience of a stimulus is influenced by its intensity. As in Hall and Rodríguez (2010), we assume that an intense stimulus will activate initially a strong expectation that some event will occur, and that it will therefore have a high initial salience. What the current new formalization captures is the notion that initial salience is later modulated by experience. Finally, also in line with the analysis of inhibition offered by Pearce and Hall (1980), we specify that the magnitude of the inhibitory reinforcer depends on the degree to which an event that does not occur was expected, that is

$$\lambda_{\text{no event}} = \sum V_{\text{event}} - \sum V_{\text{no event}} \tag{4}$$

where ΣV represents the summed associative strength of all stimuli present.

Intermixed versus blocked preexposure

We now use these equations to simulate the effects of eight trials of nonreinforced exposure to each of two compound stimuli, AX and BX, with the following parameters. For both A and B, the initial values of S, α and V_{event} were set at 0.4, 0.5 and 0.4, respectively. For X, the initial values of S, α and V_{event} were set at 0.8, 0.5 and 0.8, respectively. For A and B the initial value for net V_{event} of .4 assumed starting values of $V_{\text{event}} = .5$ and of $V_{\text{no event}} = .1$. For X, the initial value for net V_{event} of .8 assumed an initial $V_{\text{event}} = 1$ and initial $V_{\text{no event}} = .2$. Since the intensity of the stimulus is represented in the starting value of S, we decided (in contrast to the original Pearce-Hall 1980 model) to use a common starting value for all stimuli for α. The S values chosen for the various stimulus elements were intended to reflect the case in which the stimuli are difficult to discriminate because the common features shared by AX and BX (X) are relatively salient. For the intermixed condition (INT in Figure 1), the schedule of the stimulus presentations was strictly alternated (AX, BX, AX, . . .); for the blocked condition (BLK in Figure 1), the schedule of presentations consisted of a block of eight AX presentations followed by a block of eight BX presentations.

Figure 1 shows the progressive decline in salience (S) and in associability (α) for A, B and X, across the exposure trials. As can be seen, X suffers a faster and deeper loss of salience in the INT than in the BLK condition. The source of this effect is found on the early trials of exposure, when, in the INT condition, X is presented on alternate trials with a different partner (A or B), but is always presented with the same partner (A, in the particular case we have chosen to simulate) in the BLK condition. During this phase, the partners of X in the INT condition (i.e., A and B) are each presented on only half the trials, whereas the partner of X in the BLK condition is continuously presented on these early trials. This means that A (and B) suffers less habituation in the INT than in the BLK condition. Varying the elements presented from trial to trial ensures a more powerful activation of the expectancy of the occurrence of some event in the INT than in the BLK condition. As a consequence, during the first half of the trials when the alpha of X is still relatively high, the inhibitory reinforcer (equation (4)) is larger in the INT than in the BLK condition. The consequence is that extinction of the X-event association (according to equation (1)) is faster in the INT condition, resulting in faster loss of salience according to equation (3). During the second half of trials, the introduction of a new partner in the BLK condition will restore to some extent the associability of X and the value of the inhibitory reinforcer. However, at this point, X will have a relatively low associability and although extinction of X will develop somewhat faster than in the INT condition, the salience of X will remain lower in the INT than in the BLK condition.

Figure 1 also shows that A will suffer a greater loss of salience in the BLK than in the INT condition. In this case, the critical factor is that the aggregate value of the expectancy of occurrence of some event ($\Sigma V_{\text{event}} - \Sigma V_{\text{no event}}$) on the AX trials will be greater in the BLK than in the INT condition. This will be so because, in the INT condition, the partner of A (X) suffers additional habituation during its presentations with BX, thus reducing its ability to activate the relevant expectancy on the AX trials.

The other side of this coin is seen in the fact that there is no clear difference between the INT and BLK conditions in the salience of B after preexposure. In this case, B in the BLK condition is presented with a partner (X) that has already suffered a considerable amount of habituation (during the prior block of AX trials); X will thus have already lost its ability to increase the expectancy that some event will occur and thus to increase the magnitude of the inhibitory reinforcer. An implication of this analysis – that the BLK condition in which AX trials precede BX trials should differ from that in which the trial types are presented in the reverse order – has been the subject of experimental test. Results have been mixed. Symonds and Hall (1995) found no difference between the two conditions; Espinet, Caramés, and Chamizo (2011), on the other contrary, found that generalization between AX and BX was less after AX–BX training than after BX–AX training. The matter remains unresolved for the time being.

To summarize, the effects shown by this simulation are due to two of our central assumptions: (1) reductions in salience and associability of a stimulus (the source of habituation and of latent inhibition) directly depend on extinction of the initial expectancy that an event will follow the stimulus and (2) the rate to which this extinction occurs depends on the aggregate activation of that expectancy which determines the value of the inhibitory reinforcer (equation (4)) and the value of the stimulus associability (equation (2)). The INT and BLK conditions generate different patterns of activation of the aggregate activation of the expectancy that an event will occur, thus generating different habituation and latent inhibition effects for A, B and X.

It now remains to demonstrate that these changes in the properties of A, B and X mean that discrimination between AX and BX is superior after intermixed than after blocked exposure. We assume that the initial similarity of the stimuli, and thus the degree of generalization between them, will be determined by the proportion of features they hold in common and the salience of those features. Thus, following Pearce (1987), we propose that the similarity between AX and BX, their 'similarity ratio' (${}_{\text{AX}}\text{SIM}_{\text{Bx}}$), will be as follows

$${}_{\text{AX}}\text{SIM}_{\text{BX}} = \frac{S_{\text{X}}}{S_{\text{A}} + S_{\text{X}}} * \frac{S_{\text{X}}}{S_{\text{B}} + S_{\text{X}}}$$

How will this ratio change as a result of exposure to the stimuli? The result of our simulation on this ratio is presented in Figure 2. It shows that preexposure of either type reduces similarity, but that the effect is more marked in the INT than in the BLK condition.

This outcome allows the present model to explain the intermixed-blocked effects found in procedures involving direct test of discrimination, such as the same/different judgements often used with human participants (see Mitchell & Hall, 2014). We must also consider the procedure usually used with animal subjects, which involves a phase of conditioning to AX followed by a generalization test with BX. We ran further simulations for this case, in which differences in α as well as in salience will play a

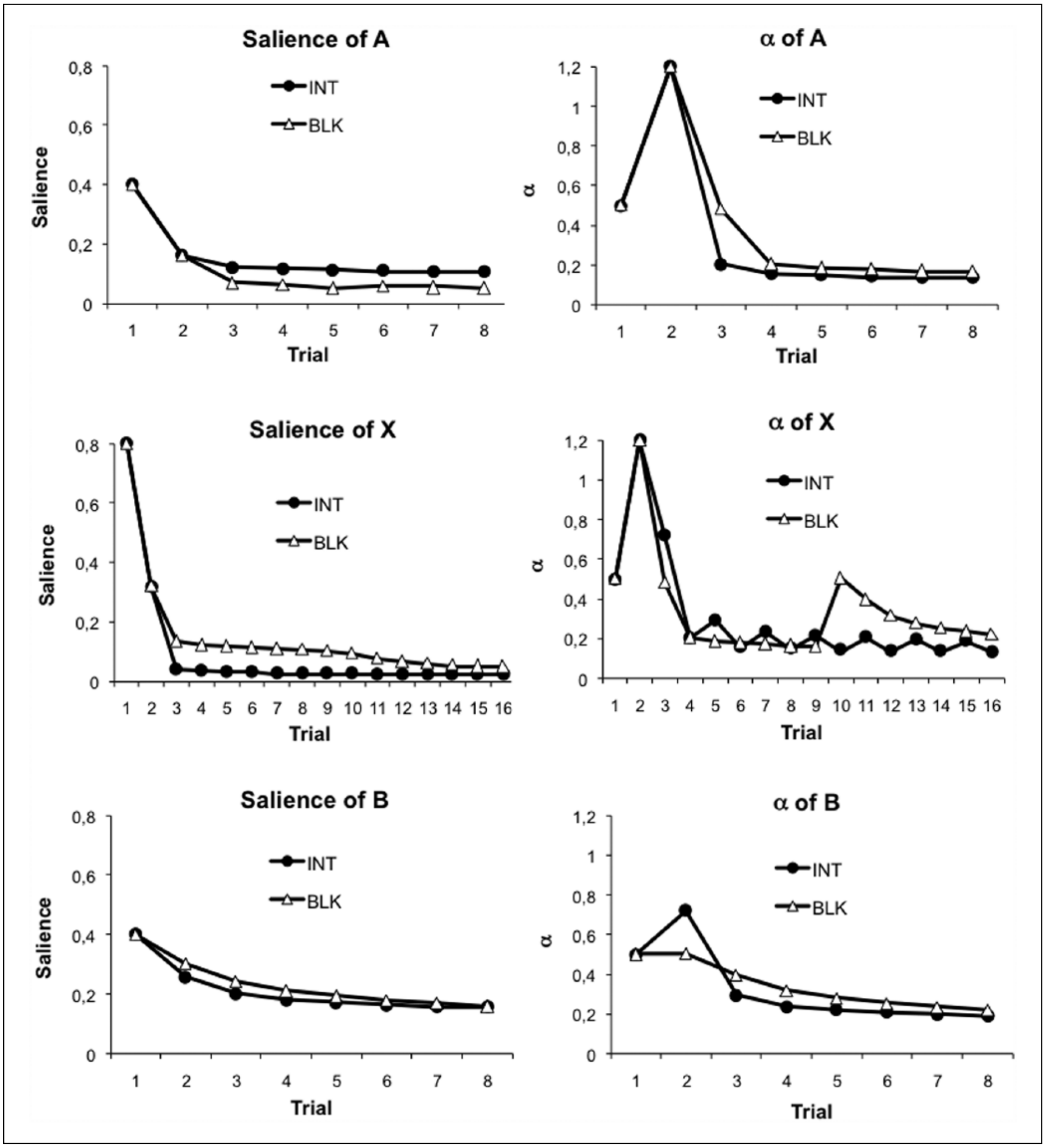

Figure 1. Simulations of the effects of preexposure on the salience and the associability (α) of the elements of the compound stimuli AX and BX, presented according to a blocked (BLK) or an intermixed (INT) schedule.

role. Figure 3 shows changes in the associative strength of A and of X over the course of three conditioning trials given after INT or BLK prior exposure.[2] Since the expectation of some event following A extinguishes more rapidly during preexposure in the BLK than in the INT condition, A starts the conditioning phase evoking a greater expectancy (and thus with greater salience) in the INT than in the BLK condition (c1 in the left panel of the figure). Although in both conditions, the occurrence of the US (i.e., an event) restores the associability of A on the subsequent conditioning trials (c2 and c3), the maintained greater salience of A in the INT condition ensures rapid acquisition. Similar considerations apply to learning about X (right panel of Figure 3). The initial difference in V_{event}, and therefore in salience, on the first trial (c1) will promote the gradual increase in differences between conditions on the subsequent trials. Critically, it is predicted that the common features X (from which direct generalization depends) will acquire less strength in the INT than in the BLK condition; also, that the unique feature A will acquire more strength in the INT than in the BLK condition. We thus anticipate the intermixed-blocked effect in those animal procedures in which generalization of the conditioning to AX is measured.

The simulations just presented used parameter values intended to represent the likely state of affairs (specifically, a salient common, X, element, and less salient distinctive, A and B, elements) that will hold for most studies of perceptual learning. We have, however, explored a range of other starting values. In all cases, the similarity between AX and BX is less after intermixed than after blocked training, although, as may be expected, the size of the effect diminishes as the salience of X is reduced and that of A and B is increased.

Conclusion

The model we have just presented is in its early stages and we do not pretend that, as it currently stands, it provides a satisfactory solution to all instances of perceptual learning, let alone the wide range of procedures used in studies of animal learning, that have been thought to engage attentional learning processes. For example, the original Hall-Rodríguez (2010) model was devised specifically to deal with latent inhibition and did so satisfactorily in terms of changes in α and the development of stimulus–no event associations. We now need to explore the effects on our predictions about latent inhibition of adding changes in salience to the theoretical mix. Again, ours is an example of a two-factor theory of attention allowing, in our case, changes in both associability and salience. Other theories of this general type (which admittedly do not usually address the issue of perceptual learning directly) have been successful in dealing with the positive transfer that can follow training on a reinforced discrimination task (as in the ease with which an intradimensional shift may be learned, or in the phenomenon known as learned predictiveness). A well-worked-out example is that proposed by George and Pearce (2012), which deals neatly with the cases just mentioned. Like our account, it makes use of two principal parameters, one reflecting salience and determined, essentially, by the associative strength of the stimulus and a second reflecting the ease with which the stimulus will enter into associations and determined by rules akin to those proposed in the Pearce-Hall (1980) model.[3] The successes of the model of George and Pearce gives hope that application of our own, rather similar account, will prove no less successful.

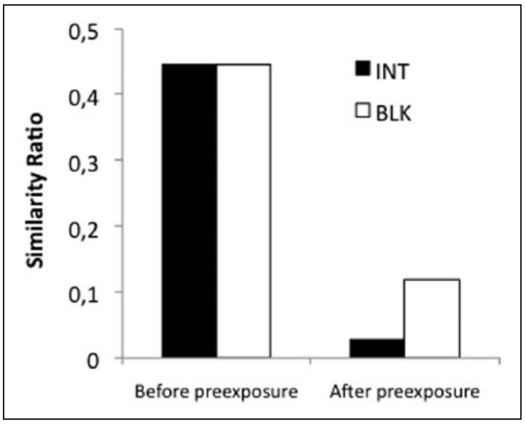

Figure 2. Similarity ratios of for AX and BX before exposure and after either intermixed (INT) or blocked (BLK) exposure.

That is for the future. It is appropriate to conclude by looking back and saying something about the origins of the theories currently being developed. What is clear is that, in spite of their various differences, from each other, and from Mackintosh (1975), they all have their roots in the approach proposed by the latter. Central to this theory was the notion that the properties of a stimulus would change in a way that was determined by the way in which it predicted its consequences. In attempting to deal with perceptual learning, both Hall (2003) and Mackintosh (e.g., McLaren &

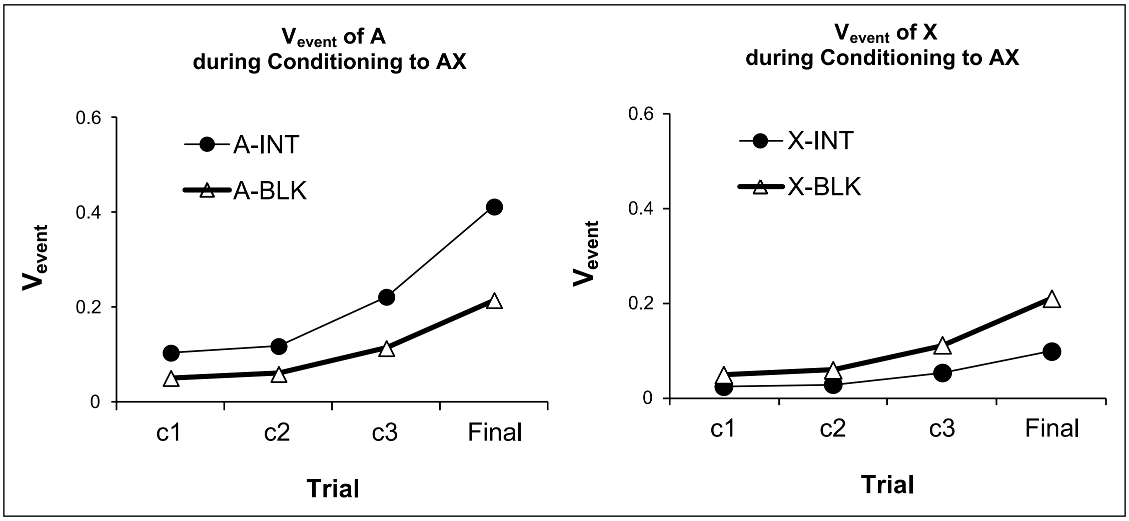

Figure 3. Acquisition of associative strength by the elements of the compound AX during three conditioning trials (c1, c2 and c3) and a Final test trial, after either intermixed (INT) or blocked (BLK) exposure to AX and BX.

Mackintosh, 2000) gave consideration to the notion that other associations – those involved in the target stimulus coming to be predicted by some other event – might be of importance. Such associations are certainly worth our consideration in some contexts (see, for example, Wagner, 1979, 1981), but it is interesting to see how far we can get in the analysis of perceptual learning, without departing from the essence of Mackintosh's (1975) theory.

Should we still be following in his footsteps? We cannot resist quoting William James' (1890) views on an eminent psychologist of an earlier age (Fechner), of whom he said,

> . . . it would be terrible if even such a dear old man as this could saddle our Science forever with his patent whimsies, and . . . compel all future students to plough through the difficulties, not only of his own works, but of the still drier ones written in his refutation . . . (p. 549)

We would plead that, however dry our refutations (or amendments), the original works of Mackintosh are well written and do not present difficulties that must be 'ploughed through'. And we may go on, and quote the rest of James' (1890) view, not in the spirit of mockery to be found in the original, but as a genuine statement of our debt:

> . . . Fechner's [Mackintosh's] critics . . . always feel bound, after smiting his theories hip and thigh and leaving not a stick of them standing, to wind up by saying that nevertheless to him belongs the imperishable glory, of first formulating them and thereby turning psychology [or, at least the analysis of the role of attention in associative learning] into an exact science . . . (p. 549)

Acknowledgements

The authors gratefully acknowledge the comments of M. Le Pelley and the contribution of A.S. Killcross in developing the theory presented here.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by grants from the Spanish Ministerio de Economía y Competitividad (Grant No. PSI2015-64309-P, MINECO/FEDER) and Gobierno Vasco (Grant No. IT-694-13).

Notes

1. Although some later theorists have found it useful to suppose that the value of α might influence performance (e.g. Le, Pelley, Suret, & Beesley, 2009), Mackintosh (1975) noted that there was no good evidence to favour the idea and that 'until such evidence is provided, it would be reasonable to suggest that α may simply be a learning-rate parameter, with no effect on the control of behavior' (p. 294).
2. For the purposes of this simulation, we made the simplifying assumption of treating conditioning as the formation of an association between the stimulus and 'an event' (which equates to the US of motivational significance in animal learning). The 'event' was given a value of 1.
3. There is danger of terminological confusion here. The parameter that we, following Mackintosh (1975), refer to as associability and symbolize by α is referred to by George and Pearce (2012) to as 'conditionability' and is symbolized by σ (an unfortunate choice, to the extent that the Greek 's' might put one in mind of 'salience'). Even more unfortunate is that the symbol they use for salience is (of all things) α. Note also that these same symbols are used (differently) by Le Pelley (2004), with α representing 'attentional associability' and σ representing 'salience associability'.

References

Artigas, A. A., Sansa, J., Blair, C. A. J., Hall, G., & Prados, J. (2006). Enhanced discrimination between flavor stimuli: Roles of salience modulation and inhibition. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*, 173–177.

Blair, C. A. J., & Hall, G. (2003). Perceptual learning in flavor aversion: Evidence for learned changes in stimulus effectiveness. *Journal of Experimental Psychology: Animal Behavior Processes*, *29*, 39–48.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, 49B, 60–80.

Dickinson, A., & McLaren, I. P. L. (2003). *Associative learning and representation: An EPS workshop for N.J.* Mackintosh. Hove; New York, NY: Psychology Press.

Dwyer, D. M., Bennett, C. H., & Mackintosh, N. J. (2001). Evidence for inhibitory associations between the unique elements of two compound flavours. *Quarterly Journal of Experimental Psychology*, 54*B*, 97–107.

Dwyer, D. M., & Mackintosh, N. J. (2002). Alternating exposure to two compound flavors creates inhibitory associations between their unique features. *Animal Learning & Behavior*, *30*, 201–207.

Espinet, A., Caramés, J. N., & Chamizo, V. D. (2011). Order effects after blocked preexposure to two compound flavors. *Behavioural Processes*, *88*, 94–100.

George, D. N., & Pearce, J. M. (2012). A configural theory of attention and associative learning. *Learning & Behavior*, *40*, 241–254.

Hall, G. (1991). *Perceptual and associative learning*. Oxford, UK: Clarendon Press.

Hall, G. (2003). Learned changes in the sensitivity of stimulus representations: Associative and nonassociative mechanisms. *Quarterly Journal of Experimental Psychology*, 56B, 43–55.

Hall, G. (2008). Perceptual learning. In J. Byrne (Editor in chief) & R. Menzel (Vol. Ed.), *Learning and Memory: A Comprehensive Reference: Vol. 1. Learning theory and behavior* (pp. 103–121). Amsterdam, The Netherlands: Elsevier.

Hall, G., Blair, C. A. J., & Artigas, A. A. (2006). Associative activation of stimulus representations restores lost salience: Implications for perceptual learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*, 145–155.

Hall, G., Prados, J., & Sansa, J. (2005). Modulation of the effective salience of a stimulus by direct and associative activation of its representation. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*, 267–276.

Hall, G., & Rodríguez, G. (2010). Associative and nonassociative processes in latent inhibition: An elaboration of the Pearce-Hall model. In R.E. Lubow & I. Weiner (Eds.), *Latent inhibition: Cognition, neuroscience, and applications to schizophrenia* (pp. 114–136). Cambridge, UK: Cambridge University Press.

Hall, G., & Rodríguez, G. (2017). Habituation and conditioning: Salience change in associative learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*, 48–61.

Holland, P. C., & Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in Cognitive Sciences*, *3*, 65–73.

Holland, P. C., & Schiffino, F. L. (2016). Mini-review: Prediction errors, attention and associative learning. *Neurobiology of Learning and Memory*, *131*, 207–215.

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

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

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

Lingawi, N. W., Westbrook, R. F., & Laurent, V. (2016). Extinction and latent inhibition involve a similar form of inhibitory learning that is stored in and retrieved from the infralimbic cortex. *Cerebral Cortex*. Advance online publication. doi:10.1093/cercor/bhw322

Lovibond, S. H. (1969). Habituation of the orienting response to multiple stimulus sequences. *Psychophysiology*, *5*, 435–439.

Mackintosh, N. J. (1974). *The psychology of animal learning*. London, England: Academic Press.

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

Mackintosh, N. J. (1983). *Conditioning and associative learning*. Oxford, UK: Clarendon Press.

Mackintosh, N. J. (1997). Has the wheel turned full circle? Fifty years of learning theory, 1946–1996. *Quarterly Journal of Experimental Psychology*, 50A, 879–898.

Maddux, J.-M., Kerfoot, E. C., Chatterjee, S., & Holland, P. C. (2007). Dissociation of attention in learning and action: Effects of lesions of the amygdala central nucleus, medial prefrontal cortex, and posterior, parietal cortex. *Behavioral Neuroscience*, *121*, 63–79.

McLaren, I. P. L., Kaye, H., & Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 102–130). Oxford, UK: Clarendon Press.

McLaren, I. P. L., & Mackintosh, N. J. (2000). Associative learning and elemental representations. I: A theory and its application to latent inhibition and perceptual learning. *Animal Learning & Behavior*, *26*, 211–246.

Mitchell, C., & Hall, G. (2014). Can theories of animal discrimination explain perceptual learning in humans. *Psychological Bulletin*, *140*, 283–307.

Pearce, J. M. (1987). A model for stimulus generalization in classical conditioning. *Psychological Review*, *94*, 61–73.

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

Pearce, J. M., & Hall, G. (1992). Stimulus significance, conditionability, and the orienting response in rats. In B. A. Campbell, H. Hayne & R. Richardson (Eds.), *Attention and information processing in infants and adults: Perspectives from human and animal research* (pp. 137–160). Hillsdale, NJ: Lawrence Erlbaum.

Rodríguez, G., Blair, C. A. J., & Hall, G. (2008). The role of comparison in perceptual learning: Effects of concurrent exposure to similar stimuli on the perceptual effectiveness of their unique features. *Learning & Behavior*, *36*, 75–81.

Shevill, I., & Hall, G. (2004). Retrospective revaluation effects in the conditioned suppression procedure. *Quarterly Journal of Experimental Psychology*, 57B, 331–347.

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. New York, NY: Macmillan.

Spinks, J. A., & Siddle, D. (1983). The functional significance of the orienting response. In D. Siddle (Ed.), *Orienting and habituation: Perspectives in human research* (pp. 237–314). Chichester, UK: John Wiley.

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

Symonds, M., & Hall, G. (1995). Perceptual learning in flavor aversion conditioning: Roles of stimulus comparison and latent inhibition of common elements. *Learning and Motivation*, *26*, 203–219.

Thompson, R. F. (2009). Habituation: A history. *Neurobiology of Learning and Memory*, *92*, 127–134.

Wagner, A. R. (1979). Habituation and memory. In A. Dickinson & R. A. Boakes (Eds.), *Mechanisms of learning and motivation* (pp. 53–82). Hillsdale, NJ: Lawrence Erlbaum.

Wagner, A. R. (1981). SOP: A model of automatic memory processing. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale, NJ: Lawrence Erlbaum.

Weisman, R. G. (1975). The compleat associationist: A review of N. J. Mackintosh's *The psychology of animal learning. Journal of the Experimental Analysis of Behavior*, *24*, 383–389.