

# Attention as an acquisition and performance variable (AAPV)

Paul Craddock · Ralph R. Miller

Published online: 8 January 2014  
© Psychonomic Society, Inc. 2014

**Abstract** This article introduces a new model of Pavlovian conditioning, attention as an acquisition and performance variable (AAPV), which, like several other so-called attentional models, emphasizes the role of variation of cue salience, together with associative strength, in accounting for conditioning phenomena. AAPV is primarily (but not exclusively) a performance-focused model in that it assumes not only that both the saliences and associative strengths of cue representations change during acquisition, but also that they are both influential at the time of test in determining responding. Different weights are given to the representations' associative strengths according to the representations' respective saliences *at test*. The model also treats the representation of a stimulus that is directly activated by presentation of that stimulus as distinct from the representation of the same stimulus that is activated by presenting a companion of the stimulus. Additionally, extinction is viewed as resulting from a decrease in the salience of the cue's representation, rather than a decrease in associative strength. Simulations of several Pavlovian phenomena are presented in order to illustrate the model and assess its robustness.

**Keywords** Pavlovian conditioning · Attention · Cue salience · Learning theories

P. Craddock  
University of Lille, Lille, France

R. R. Miller  
State University of New York at Binghamton, Binghamton, USA

P. Craddock (✉)  
Université de Lille, Nord de France, Domaine universitaire du  
« Pont de Bois », rue du Barreau, BP 60149, 59653 Villeneuve  
d'Ascq Cedex, France  
e-mail: paul.craddock@univ-lille3.fr

The last third of the 20th century saw an immense growth of interest in the mechanisms underlying so-called *cue competition* phenomena, such as overshadowing (Pavlov, 1927) and the blocking effect (Kamin, 1968). These phenomena were first interpreted as reflecting differences in the degree of learning achieved during acquisition (e.g., Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972). That is, the conditioned response (CR) to a target cue (X) was believed to be a direct reflection of the X–unconditioned-stimulus (US) association that the organism had learned or failed to learn. For example, the Rescorla–Wagner model assumes that the weak response elicited by X at test in forward blocking (i.e., A–US followed by AX–US, where A is the competing cue and X the target cue) results from a failure to acquire a strong X–US association because of A's previously established associative strength with the US. This kind of interpretation has been described as a *what-you-see-is-what-you-get* (WYSIWYG) model. Yet the WYSIWYG assumption was challenged by results showing that the CR to X can often be reversed by simply deflating the association between A and the US (i.e., *retrospective revaluation*). For example, retrospective revaluation is observed when extinction of the competitor A after overshadowing results in enhancement of the CR to X (e.g., Kaufman & Bolles, 1981; Matzel, Schachtman & Miller, 1985; but see Holland, 1999). Critically, in retrospective revaluation, increases in the CR to X are observed without additional X–US pairings.

Miller and Matzel's (1988) comparator hypothesis was the first model designed to account for retrospective revaluation of the target cue's excitatory status. It assumes that cue competition effects are not determined by differences in the strength of the acquired X–US association. Rather than assuming that cues compete for a limited resource of associative strength during acquisition (e.g., Rescorla & Wagner, 1972), the comparator hypothesis states that cues compete at test, during which the associative strength between X and the US is

compared with the competitor's associative strength with the US. This comparison of associative strengths results in the modulation of the CR to the target CS, with strong CRs when the competitor's associative strength with the US is weak as compared with the X–US association and weak CRs when the competitor's associative strength with the CR is strong as compared with the X–US association. Subsequently, Van Hamme and Wasserman (1994), as well as Dickinson and Burke (1996), proposed acquisition-focused WYSIWYG models that can account for retrospective revaluation by allowing the revaluation of a target's associative strength even when the target is absent, provided that an associate of the target is presented. The latter two models succeeded in explaining retrospective revaluation, as well as many traditional phenomena, while relying on X's absolute (as opposed to relative) associative strength to control responding to X at test. In contrast, Stout and Miller's (2007) SOCR, the mathematical implementation of Miller and Matzel's model, demonstrated that referring to relative associative strength rather than absolute associative strength accounts for several newer phenomena in addition to traditional phenomena and retrospective revaluation.

The model described in the present article was designed to explain numerous well known Pavlovian phenomena, as well as retrospective revaluation, by assuming that the comparison that occurs at test is not between different associative strengths (as the comparator hypothesis proposes), but between the saliences of the various stimulus representations that are activated at test. The central idea is that, if the representation of a cue that is activated during test is highly salient (i.e., attracts attention), its associative strength with the US will be given greater weight than the equivalent associative strength between a less salient cue representation and the US. This greater weight will result in a greater manifestation of that cue's specific associative strength. Moreover, attention is also assumed to influence acquisition, as is assumed by many earlier attentional models. Therefore, we chose a name that captures these two characteristics: *attention as an acquisition and performance variable* (AAPV). Although AAPV concerns itself with information processing during *both* acquisition and test trials, we will see that many differences between groups (e.g., an overshadowing group and its conventional control group) are due largely to differences in the saliences of cue representations at the time of testing. This makes AAPV primarily (but not exclusively) a performance-focused model. As such, AAPV shares this characteristic with comparator models (e.g., Gallistel & Gibbon, 2000; Stout & Miller, 2007) and distinguishes it from total-error reduction models (e.g., Rescorla & Wagner, 1972).

Indeed, several prior models have employed attention in their accounts of Pavlovian conditioning (e.g., Esber & Haselgrove, 2011; Le Pelley, 2004; Mackintosh, 1975; Pearce & Hall, 1980; Pearce & Mackintosh, 2010), but

the trial-by-trial variation in cue-specific salience they suggested is determined by the amount of associative strength already acquired or yet to be learned, and regulates the amount of associative strength learned on the next trial. For example, Mackintosh's model assumes that the saliences of cues vary on each trial as a function of their potential to predict the US and the US that actually occurs. If a cue is a good predictor of the US, its salience will increase. In contrast, the Pearce and Hall model assumes that a cue that is an accurate predictor of its outcome loses associability (notably, Pearce and Hall differentiate between two types of attention: *associability*, which is a function of reinforcement history with the cue in question, and *saliency*, which is fixed for any given cue). Efforts have been made recently to reconcile the apparently opposite predictions concerning evolution of attention made by these two models (see Esber & Haselgrove, 2011; Le Pelley, 2004; Pearce & Mackintosh, 2010). Critically, in all of these attentional models, salience (associability for Pearce & Hall, 1980) is seen as regulating the amount of associative strength learned on each training trial, and reciprocally, associative strength is assumed to determine trial-by-trial variation of salience; these models explain conditioned responding as being solely dependent upon the cue–US associative strength at test. (A few researchers have suggested that attention at the time of testing may also influence conditioned responding, but they did not make this part of their formal models [e.g., Mackintosh, 1975].) As we will see, AAPV too posits that salience affects the rate of acquisition of associative strength on a trial-by-trial basis. But it focuses more on how cue salience at test influences test performance by modulating the expression of the cue–US association.

### Learning simple associations between a cue and its outcome

We define  $V_{X \rightarrow o}$  as the associative strength between a cue X and its outcome (o), where o is often a US but could be a neutral stimulus. A high value of  $V_{X \rightarrow o}$  indicates that o is no longer a surprising event when it follows X. As will be shown later, positive values of  $V_{X \rightarrow o}$  are a necessary condition for X to have the potential to elicit a response appropriate for o (e.g., the CR). The model can be formalized with the following equation for the acquisition of associative strength when cue X and outcome o are paired (adapted from Bush & Mosteller, 1955; see also Stout & Miller, 2007):

$$\Delta V_{X \rightarrow o}^n = S_X^{n-1} * (1 - V_{X \rightarrow o}^{n-1}), \quad (1)$$

where  $\Delta V_{X \rightarrow o}$  denotes the change of associative strength ( $V$ ) between X and o, as a result of X being paired with the o on trial  $n$  ( $0 \leq V_{X \rightarrow o} \leq 1$ ).  $S_X$  corresponds to cue X's salience ( $0 < S_X < 1$ ). Saliency corresponds to the cue's

physical salience when it is presented for the first time, but it decreases or increases depending on the outcome's salience, as we will explain later. Because, in the equation for modulation of salience (Eq. 2),  $S_X$ 's increases and decreases are a function of the outcome's salience, the outcome's salience is assumed to indirectly influence the evolution of  $V_X$ . Importantly, as Eq. 1 indicates, the asymptote toward which  $V$  increases is 1. A major characteristic of AAPV is that no decrease in the associative strength is predicted by the model. Most models of Pavlovian conditioning previously mentioned allow  $V$  to decrease during extinction. However, empirical evidence challenges the hypothesis of a decrease in the associative strength during extinction trials (see Bouton, 2004, for a review). For example, spontaneous recovery (i.e., postextinction recovery of the CR when a long delay is imposed between extinction and test; e.g., Pavlov, 1927) is difficult to interpret if  $V$  tends toward zero during extinction. Other phenomena, such as the renewal effect (i.e., recovery of the CR when the extinguished cue is tested in a context different from the context of extinction; e.g., Bouton & Bolles, 1979) and the reinstatement effect (i.e., recovery of the CR to  $X$  when the US is presented alone between extinction and test; e.g., Rescorla & Heth, 1975), also argue in favor of the view that  $V$  does not decrease during extinction.

One of the model's most important characteristics is that a cue's salience increases or decreases according to its outcome's salience. Hence, a cue that is repeatedly followed by a biologically significant outcome (e.g., a US) will become more and more salient until its salience eventually reaches the outcome's salience ( $S_o$ ). Just as increases in  $V_{X \rightarrow o}$  allocate to  $X$  the potential to elicit the response generated by  $o$ , increases in  $S_X$  will allow  $X$  to have the same salience as the outcome. The intuition here is that if a cue produces the outcome's response, it should acquire its salience. In like manner, if  $X$  is not followed by any specific outcome, its salience will decrease to eventually reach the low salience of the context, which effectively is the outcome of such a cue. Thus, the cue is now effectively treated as a part of the context because it has become familiar and has the context's salience. Hence, both modulation of salience and acquisition of associative strength depend on stimulus contiguity.

Equation 2 represents the change in  $S_X$  on trial  $n$ . It was designed so that the salience of a cue can evolve in one direction or the other on the basis of what follows activation of the cue's representation:

$$\Delta S_X^n = (S_o^n - S_X^{n-1}) * S_o^n, \quad (2)$$

where  $S_o^n$  represents the outcome's salience on trial  $n$ .  $S_o^n$  is assigned the context's salience (i.e., .1 in the simulations presented here) when  $X$  is followed by no nominal stimulus. Because variations of  $S_X^n$  are presumed to be modulated by the

outcome's salience ( $S_o^n$ ) on trial  $n$ , decreases in the cue's salience will be slower than increases.

Equation 3 provides AAPV's account of how the two variables  $V$  and  $S$  interact at test to cause  $X$  to elicit a response appropriate for the outcome. Presumably, as is commonly acknowledged, the response appropriate for the outcome is mediated by an anticipatory representation of the outcome:

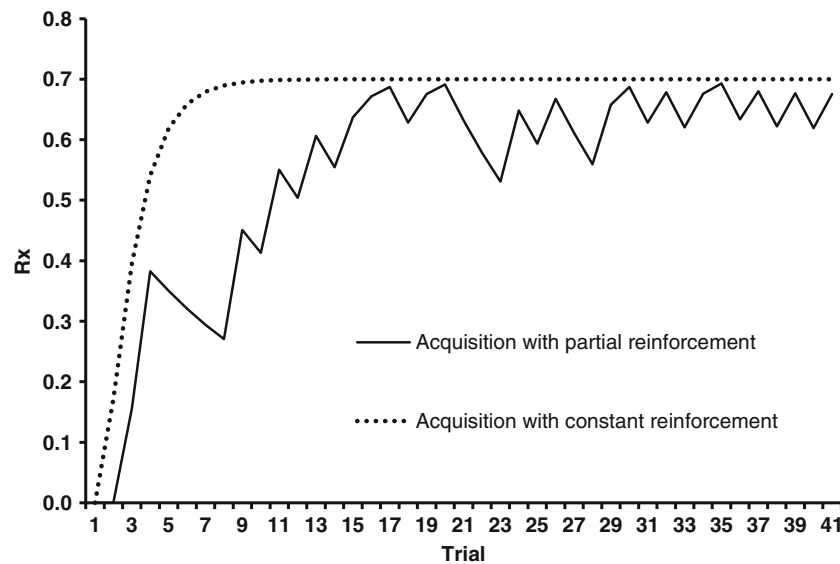
$$R_{X/o} = k * S_X * V_{X \rightarrow o} \quad (3)$$

where  $S_X$  represents  $X$ 's current salience and  $V_{X \rightarrow o}$  the current associative strength between  $X$  and the outcome.  $k$  is a scaling factor that can take any positive value; it is designed to balance dimensionality in the equations and to scale AAPV's predictions to the actual scores of each data set.  $k$  is set different from 1 in the present article only when AAPV's predictions are to be compared with a specific data set depending on the relevant behavioral metric of the data, because the experiments simulated here differed in many procedural details (e.g., nature of the CSs, CS durations, CS intensities, nature of the USs, US duration, interstimulus interval, intertrial interval, session duration, and number of sessions).

For the simulations presented in this article, we assume that a familiar context that no longer attracts appreciable attention has a salience of .1. This low value has been chosen to be as distinct as possible from the salience of .7, which we have allocated to the US, and .1, being near 0, emphasizes the assumption that this context no longer draws much attention. In contrast, a biologically significant event like a US should have a high salience (.7). An ordinary novel cue introduced in the context is presumed to initially have a moderate salience of .4.

During acquisition (i.e.,  $X$ -US pairings), not only does  $V_{X \rightarrow US}$  increase (Eq. 1; the outcome is the US), but so too does  $S_X$  (Eq. 2). After a sufficient number of trials,  $V_{X \rightarrow US}$  tends toward 1. At that stage,  $X$  has acquired the potential to elicit the CR, which it will never lose (see Eq. 1). Whether this potential will be expressed will depend upon  $X$ 's salience at test. During asymptotic acquisition,  $S_X$  reaches the US's salience (i.e., .7); consequently,  $R_{X/US}$  (also designated  $R_X$ ) also reaches the value of .7 (i.e., the product of the  $X$ -US association and the salience of the CS). The evolution of  $R_X$ 's values during acquisition trial-by-trial is illustrated in Fig. 1.

Second-order conditioning (Pavlov, 1927) in which  $X$  is not directly paired with the US but with a cue that has previously been paired with the US (i.e.,  $A$ -US followed by  $X$ - $A$ ) is also readily explained by AAPV. Responding to  $X$  occurs because  $X$ 's association with  $A$  (i.e.,  $V_{X \rightarrow A}$ ) is high. Because  $A$  produces the CR and  $V_{X \rightarrow A}$  is high,  $X$  also acquires the potential to elicit the CR, due to activation at test of the anticipatory representation of a cue (i.e.,  $A$ ), which evokes the anticipatory representation of the US. Moreover,  $A$ 's salience is high because it was followed by the US and  $X$  and then



**Fig. 1** Evolution of  $R_X$  as a function of the number of acquisition trials of (i.e., X–US) with constant reinforcement and with partial reinforcement (20 trials reinforced and 20 nonreinforced, randomly interspersed). For this data set,  $k$  was set at 1

acquires A's salience because it is repeatedly followed by A. X's high salience allows the CR to be expressed. If the two phases are reversed, as is the case in a sensory preconditioning preparation (i.e., X–A followed by A–US), conditioning of X is also observed (Brogden, 1939). AAPV assumes that, because X has acquired the capacity of producing A's responses (among which is the CR after phase 2), X also is conditioned by way of the simple associative chain X–A–US during sensory preconditioning. However, AAPV anticipates that sensory preconditioning would yield weaker responding to X than second-order conditioning because X–A pairings are presented before  $S_A$ 's increase in the former condition and after  $S_A$ 's increase in the latter condition. Therefore, only during second-order conditioning is X paired with a highly salient cue that increases  $S_X$ .

One major difficulty in modeling acquisition concerns procedures in which partial reinforcement is administered. A model of acquisition should account for the fact that, even if the cue is only sometimes followed by the US, conditioned responding to the cue (albeit often weaker) develops. Simulations of partial reinforcement of X with AAPV show trial-by-trial variations in  $R_X$  due to the variations of X's salience. However, these variations are kept rather small by the fact that variations of salience are slow following nonreinforced trials, as compared with the rapid increases resulting from reinforcements of the cue (Fig. 1) until an asymptote for  $R_X$  is reached. Indeed, Eq. 2 shows that the variations of saliences are dependent upon the outcome's salience, which is small during nonreinforced trials (i.e.,  $S_o = S_{\text{context}}$ ) and large during reinforced trials (i.e.,  $S_o = S_{\text{US}}$ ). Hence, AAPV predicts a high level of responding as a result of intermittent reinforcement procedures, due to salience not being appreciably decreased within the session (see Fig. 1). Moreover,

resistance to extinction is known to be stronger (i.e., extinction takes more nonreinforced trials) when acquisition has been performed with partial reinforcement, as compared with when constant reinforcement has been used (e.g., Haselgrove, Aydin & Pearce, 2004; Pearce, Redhead & Aydin, 1997). This partial reinforcement extinction effect has challenged most of the models of Pavlovian conditioning and challenges AAPV too. AAPV anticipates a similar rate of decrease of responding during extinction whether acquisition occurred with partial reinforcement or constant reinforcement because X's salience is high after acquisition whatever the type of reinforcement.

However, if a simple constant acquisition phase is followed by extinction (i.e., X–US followed by X–no-US trials),  $V_{X \rightarrow \text{US}}$  will not decrease, because learned associative strengths do not decrease according to AAPV. But across extinction trials,  $S_X$  will decrease according to Eq. 2 to finally reach the context's salience after many extinction trials. Consequently,  $R_X$  progressively approaches a low asymptotic value after many extinction trials (i.e.,  $R_X = .1$ ; see Fig. 2). Hence, AAPV does not assume that extinction weakens associative strength, as the Rescorla–Wagner model does. Moreover, AAPV's explanation of extinction is not based on forming an association between the cue and a no-outcome representation, as some other models propose (e.g., Pearce & Hall, 1980); rather, AAPV assumes that extinction results from a decrease of the cue's salience. This view has already received support from experimental data (e.g., Robbins, 1990); however, other observations, such as counterconditioning (i.e., X–US1 followed by X–US2), are difficult to reconcile with the view that the decrease in a CR during extinction is due to a decrease of attention to the CS. AAPV assumes that the CS's salience should not decrease during phase 2 of

counterconditioning, because it is followed by a salient cue (i.e., US2). However, the CR appropriate for US1 decreases. Presumably, this is due to response competition, which is beyond the domain of AAPV.

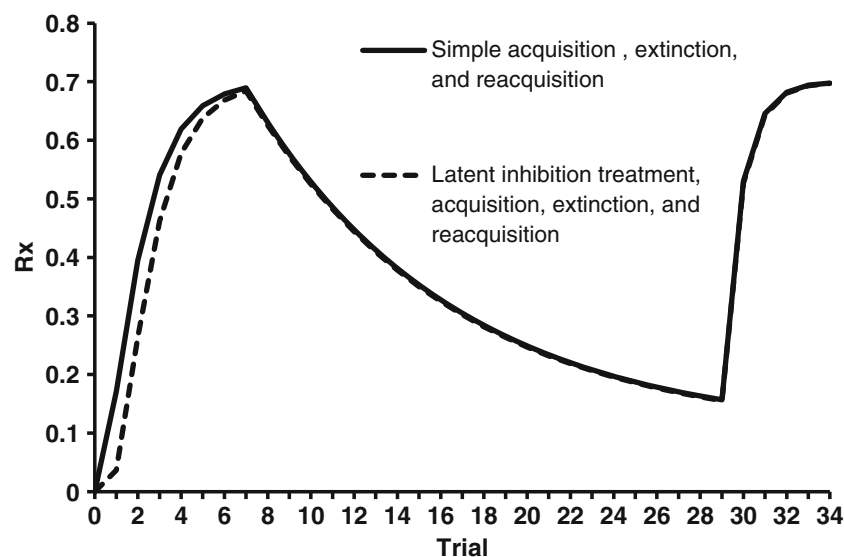
Rapid reacquisition of behavioral control after extinction (e.g., Napier, Macrae & Kehoe, 1992; but see Bouton & Swatzentruber, 1989) is anticipated by AAPV (see Fig. 2) because  $V_{X \rightarrow US}$  is already high when reacquisition begins, as opposed to the null associative strength that exists prior to the first acquisition phase. The model also predicts that reacquisition will be slower if extinction is prolonged, because  $S_X$  progressively decreases during extinction to eventually reach the asymptotic value of .1 if extinction is sufficiently prolonged.

### Direct and retrieved representations

Hitherto, we have considered what happens when a cue is trained alone, but a model of Pavlovian conditioning should also be able to make predictions about what happens when compound cues are paired with the US. We suggest, as others did before (e.g., Wagner, 1981), that when a stimulus X is presented in compound with another stimulus A, it acquires the potential to later activate a representation of A, which we call the *retrieved* representation of A (Rr[A]), when X is presented alone because of the resultant within-compound link between X and A. For the sake of simplicity, the current version of the model makes no prediction concerning the evolution of the within-compound link between X and A. Here, the concept *retrieved representation* (Rr) is specifically used to describe the representation that is activated by a

companion cue such as X when the stimulus in question (A) is not present in the environment. Rr(A) is the retrieved memory of A activated by a cue currently present in the environment with which it has previously been presented in compound. Thus, *companion* refers to stimuli that are simultaneous with a cue, whereas *associate* refers to stimuli that follow a cue; this distinction explains why simultaneous conditioning (i.e., simultaneous CS–US pairings) result in little or no behavioral control by the CS. Therefore, Rr(A) activated by a test of X alone after overshadowing treatment (i.e., AX–US pairings) is distinct from the direct representation of A (Dr[A]), which is the representation directly activated by the presence of A in the environment during the AX presentations. Rr(A) is not activated by X if X and A are presented in compound (i.e., activation of Dr[A] inhibits activation of Rr[A]). We propose that Drs and Rrs bring distinctly different contributions to the generation of any conditioned response in Pavlovian conditioning.

In Miller and Matzel's (1988) comparator hypothesis, when X is tested alone, the representation of A is assumed to be retrieved if X and A have previously been presented together. It is the learned association between comparator stimulus A and the US acquired while A was actually present, in conjunction with the X–A and X–US associations, which modulates X's potential to elicit a CR. In contrast, the current model allocates a role to the relationship that Rr(A), not Dr(A), entertains with the US when X is tested alone. This aspect establishes one of several sharp differences between the comparator hypothesis and AAPV. Here, no role is allocated to the associative strength between Dr(A) and the US when X is tested in the absence of A. What counts is not the degree to which the directly activated representation of the competitor A



**Fig. 2** Evolution of  $R_X$  as a function of number of acquisition trials of (i.e., X–US) and for a latent inhibition design (i.e., phase 2, consisting of X–US during trials 1–7 following X-alone presentations). Trials 8–29

correspond to extinction. For both conditions, extinction is followed by five trials of reacquisition (trials 30 to 34). For this data set,  $k$  was set at 1



(i.e.,  $Dr[A]$ ) is associated with the US but the degree to which the retrieved representation of A ( $Rr[A]$ ) is associated with the US. The role of associations between the trace of a stimulus and the US has already been considered (e.g., Lin & Honey, 2011); however, the trace these authors considered concerns the memory of stimuli that had been presented 40 s before the US onset. In the present model, the trace ( $Rr[A]$ ) is not the remnant of  $Dr(A)$ 's previous activation but the activation that results from  $Dr(X)$ 's activation without  $Dr(A)$ . The core hypothesis is that only the representations activated by X on a test of X (i.e.,  $Dr[X]$  and  $Rr[A]$ ) can express their associative strengths with the US. Given that  $Dr(A)$  is not active when X is tested alone (and, moreover, it is never activated by X), its learned association with the US does not influence the CR to X, contrary to the comparator hypothesis. Obviously, when cue X is followed by the US, the putative associative strength that builds up between X and the US is not between these stimuli but between their representations. Just as Pavlovian conditioning is assumed to result from the increasing associative strength between the directly activated representation of X (i.e.,  $Dr[X]$ ) and the directly activated representation of the US (i.e.,  $Dr[US]$ ) when they are paired, AAPV anticipates that associations will also be formed between activated retrieved representations and  $Dr(US)$  when activation of these Rrs is followed by the US.

To illustrate the mechanics of AAPV when compound cues are presented during training, we will use overshadowing (i.e., AX–US) as an iconic example. AAPV does not imply any competition for a limited resource of associative strength; that is, it does not subscribe to the total-error reduction rule of the Rescorla–Wagner (1972) model. Any activated cue representation (direct or retrieved) followed by the US, even in compound with other cues, has the opportunity to acquire a strong association with the US. Also, AAPV assumes that, when X is presented alone at test following AX–US pairings, the resultant CR to X will depend on the mean of the weighted associative strength between  $Dr(X)$  and the US and the weighted associative strength between  $Rr(A)$  and the US. Hence, in contrast to the comparator hypothesis (e.g., Miller & Matzel, 1988), we do not suggest that an opposition between representations' associative strengths to the US occurs at test. This difference of views arises from AAPV's assumption that the associative strength with the US of the representation of an absent former companion cue positively contributes to responding to X at test, whereas the comparator hypothesis invokes only the modulatory role of the association between the representation of the companion cue and the US. As a consequence, AAPV anticipates the strongest CRs when both  $Dr(X)$  and  $Rr(A)$  are strongly associated with the US. If  $Rr(A)$  has not been paired with the US, only a part of the cue representations that are activated on a test of X are associated with the US (i.e.,  $Dr[X]$  but not  $Rr[A]$ ). The organism either has not yet had a chance to learn whether the US appears in the

absence of A (e.g., as in overshadowing treatment with one X-alone test trial) or has learned that the omission of A on repeated X-alone test trials following overshadowing treatment signals nonreinforcement (i.e., activation of  $Rr[A]$  has been repeatedly followed by the absence of the US over multiple X-alone test trials). Both of these situations result in  $Rr(A)$  not being associated with the US.

Equation 1 can, therefore, be generalized to any representation that has been paired with the US (Eq. 4). We define  $V_{i \rightarrow o}$  as the associative strength between any representation  $i$  (either  $Dr$  or  $Rr$ ) and the outcome ( $o$ ), where  $o$  is often but need not be a US. A high value of  $V_{i \rightarrow o}$  indicates that  $o$  is no longer a surprising event following  $i$ . Positive values of  $V_{i \rightarrow o}$  are a necessary condition for  $i$  to have the potential to elicit a response appropriate for  $o$  (e.g., the CR).

$$\Delta V_{i \rightarrow o}^n = S_i^{n-1} * (1 - V_{i \rightarrow o}^{n-1}), \quad (4)$$

where  $\Delta V_{i \rightarrow o}$  denotes the change of associative strength ( $V$ ) between  $i$  and  $o$ , as a result of  $i$  being paired with the  $o$  on trial  $n$  ( $0 \leq V_{i \rightarrow o} \leq 1$ ).  $S_i$  corresponds to the salience of the representation of cue  $i$  ( $0 < S_i < 1$ ).

The salience of the retrieved representation of A ( $S_{Rr(A)}$ ) activated by presenting A's companion cue X alone is equal to  $S_{Dr(A)}$  (also designated by  $S_A$ ) the first time that  $Rr(A)$  is activated but evolves independently from  $S_A$  when  $Rr(A)$  is subsequently activated. The evolution of saliences over trials will be detailed later.

### Global associative strength

As has already been stated, if more than one representation is activated by X when X is tested (e.g.,  $Dr[X]$  and  $Rr[A]$ ), each associative strength between an active representation and the outcome (i.e.,  $V_{X \rightarrow O}$  and  $V_{Rr(A) \rightarrow O}$ ) will contribute to the generation of the response appropriate for the outcome. It is commonly accepted that the CR is activated by an anticipatory representation of the US that is retrieved by the CS as a result of repeated CS–US pairings. AAPV assumes that this anticipatory representation of the outcome is activated by  $Dr(X)$  and  $Rr(A)$  as a result of  $Dr(X)$ 's and  $Rr(A)$ 's association with the outcome. Hence, Drs and Rrs are viewed as representations distinct from anticipatory representations, in that Drs and Rrs have the potential to activate anticipatory representations of the US, which, in turn, will activate the CR. It is their level of association with the outcome that determines the degree to which the anticipatory representation will be activated. Finally, anticipatory representations inform the organism that a cue is about to appear in the environment. According to AAPV, the anticipated CR is determined by the mean of the activated representations' associative strengths at test, rather than only by  $V_{X \rightarrow US}$ , as described in Eq. 3 when a single cue is trained.

Since more than one cue representation can be activated by X at test, we now need to speak to the relative importance of each of the associations activated by presentation of X above and beyond the relative strengths of these associations. Let us imagine what would happen if, for instance,  $V_{X \rightarrow US}$  was strong, whereas  $V_{Rr(A) \rightarrow US}$  was weak. This happens, for example, on a test of overshadowing (i.e., AX–US pairings) in which  $V_{X \rightarrow US}$  increases over the AX–US pairings but in which  $V_{Rr(A) \rightarrow US}$  remains null because Rr(A) is never activated prior to test (i.e., the organism has not yet learned whether or not the US occurs following activation of Rr[A]). Indeed, when the two representations (i.e., Dr[X] and Rr[A]) are activated by X at test, the information provided by each representation concerning the likelihood of the US is different. Should we assume that these two associative strengths are simply averaged (i.e., given equivalent weight)? We suggest this is not the case. Instead, the more salient a representation, the more its associative strength will influence responding. For example, if during a test of overshadowing, Dr(X) and Rr(A) are both activated, more importance (weight) will be allocated to the associative strength between Dr(X) and the US because Dr(X) is more salient than Rr(A). The relative attentional weight of each associative strength at test can be estimated by allocating a coefficient to each activated  $V_{i \rightarrow o}$  depending on the respective saliences of the  $i$  representations. Therefore,  $R_X$  is determined by the salience-weighted average of the different associative strengths to the US that are activated at test.

The weight given to each representation's associative strength with the outcome is proportional to the representation's salience. By  $i$ 's salience ( $S_i$ ), we refer to the current salience a representation has at test, which is a function of its salience when it was first activated, as well as subsequent increases and decreases dependent upon whether the representation has been repeatedly paired with the US (or any salient outcome) or followed by no stimulus (or a low salient event). After AX–US pairings in phase 1, if A's companion cue X is presented alone and followed by the US (e.g., AX–US followed by X–US), Rr(A)'s salience will progressively increase as well as will Dr(X)'s salience because both representations are repeatedly paired with the US during phase 2. These increases in salience incorporate the idea that Dr(X) and Rr(A) now command more of the organism's attention because they preceded a significant event. Their saliences progressively approach the US's salience, which is the asymptote for salience of a reinforced representation. AAPV assumes that the salience of the US does not decrease significantly because US is biologically significant; in other words, a stimulus that is unconditionally significant for the organism does not readily lose its salience. This is essential because, otherwise, a trial-by-trial decrease in the US's salience would be expected over the X–US pairings because the US is followed by no salient event. However, assuming a relatively gradual decrease in a

US's salience would allow AAPV to predict both habituation and the decrease in the CR observed when the number of conditioning trials is great (i.e., the overtraining effect; Pavlov, 1927; Urcelay, Witnauer & Miller, 2012). Analogously,  $S_{Rr(A)}$  will decrease when Rr(A) is repeatedly activated without being followed by the US. Rr(A)'s salience (i.e.,  $S_{Rr[A]}$ ) is determined by Dr(A)'s current salience only when Rr(A) is activated for the first time, because it is assumed that more attention is paid to a missing salient cue than to a missing nonsalient cue. Rr(A)'s salience increases and decreases independently of Dr(A)'s salience with the repeated pairings of Rr(A) with the US or with no stimulus. For example, after repeated presentations of the compound AX paired with the US, when X is tested, the current value of  $S_{Dr(A)}$  is allocated to Rr(A) because Rr(A) is activated for the first time at test (since Rr[A] is activated by X only when both A is absent and X has previously been simultaneously paired with A). If X–US pairings are presented after the AX–US pairings, Rr(A)'s salience will correspond to Dr(A)'s current salience only the first time it is activated. Rr(A)'s salience will then increase (due to Rr[A]'s activation being paired with the US) during the X–US pairings, whereas Dr(A)'s salience will remain the same. Admittedly, it is unusual to allocate salience to a retrieved representation, because psychological salience is often equated with stimulus intensity and is conventionally thought to apply only to stimuli that are physically present. If physical objects are identified as being highly salient, it is presumably because their impact on the brain is greater than that of other objects in the environment. In like manner, we assume that the absence of cue A on an X-alone trial (activating Rr[A]) can have a greater impact on the brain (represented by  $S_{Rr[A]}$ ) either because Rr(A) is activated for the first time when A itself was salient or because Rr(A) is known to be followed by the US independently of A's salience.

### Mathematical implementation of global associative strength

As Eq. 5a indicates and as we have already stated, the weight allocated to  $V_{i \rightarrow US}$  is a function of  $S_i$ .

$$R_{X/o} = k * \frac{\sum (S_i * V_{i \rightarrow o})}{\sum S_i}, \quad (5a)$$

where  $k$  is a scaling factor that can take any positive value; it is designed to balance dimensionality in the equations and to scale AAPV's predictions to the metric of each data set.  $S_i$  represents the saliences of the different representations activated at test.

The amplitude of the CR elicited by X at test is assumed to be proportional to the average salience-weighted strength of the associations to the US active at test. For example, if X is

tested after AX–US pairings (i.e., overshadowing treatment), the relevant associative strengths will be  $V_{Dr(X) \rightarrow US}$  and  $V_{Rr(A) \rightarrow US}$ , to which  $Dr(X)$ 's salience and  $Rr(A)$ 's salience will be allocated as weights, respectively, normalized by the sum of saliences of all cues present (i.e., global salience) in the denominator. The view that the CR is determined by the weighted average of the associations to the US activated at test is fundamental to AAPV. Equation 5a is similar to Pearce's (1994) model's Eq. 3, which assumes that the expression of  $V_T$  (i.e., the associative strength of the compound AX) is determined by the weighted average of  $Dr(A)$ 's and  $Dr(X)$ 's associations with the US. AAPV's Eq. 5a concerns associations with the US of  $Dr(X)$  and the different Rrs activated by X at test. However, in contrast with Pearce's proposal that the presence of another cue (A) with X on a given trial would reduce  $\Delta V_X$  proportionally to A's salience (Pearce's Eq. 2), AAPV's Eq. 4 does not assume that a representation's increase in associative strength is modulated by other representations being activated at the same time. Equation 5a presents the CR to X as being solely dependent on the weighted average of the associative strengths to the US. However, AAPV views this weighted average only as the potential that  $Dr(X)$  together with the activated Rrs have to produce the response appropriate for the outcome. This potential is modulated by the salience that X (i.e., the tested cue) has at the time of test. Analogously, a US has the maximal potential of producing its own response, but the amplitude of the unconditioned response is proportional to the US salience. Equation 5b accounts for this further role of salience of cues that are present by multiplying Eq. 5a by the tested  $Dr$ 's salience:

$$R_{X/o} = k * S_X * \frac{\sum (S_i * V_{i \rightarrow o})}{\sum S_i} \tag{5b}$$

When X is trained alone (i.e., never presented in compound with an associate cue), only  $Dr(X)$  will be activated at test, and Eq. 5b will reduce to  $k * S_X * \frac{S_X * V_{X \rightarrow o}}{S_X}$ , which corresponds to Eq. 3.

**Evolution of salience when multiple representations are active**

As has previously been stated, a stimulus regularly followed by a highly salient outcome will increase in salience toward some asymptote, as well as increase in associative strength with the outcome. In like manner, a stimulus can lose its potential to attract attention if it is regularly presented followed by no stimulus (i.e., extinction). Equation 2 has already introduced AAPV's position concerning how salience evolves when only one cue is trained. Equation 6 generalizes Eq. 2 to any representation, whether  $Dr$  or  $Rr$ .

Equation 6 represents the modulation of  $S_i$  on trial  $n$ . It was designed so that the salience of a cue's representation can evolve in one direction or the other based on what follows activation of the cue's representation.

$$\Delta S_i^n = (S_o^n - S_i^{n-1}) * |S_o^n - S_a^{n-1}| \tag{6}$$

where  $i$  represents any  $Rr$  or  $Dr$  activated on trial  $n$ .  $S_o^n$  represents the outcome's salience on trial  $n$ ;  $S_o^n$  is assigned the context's salience (i.e., .1 in the simulations here presented) when  $i$  is followed by no nominal stimulus.  $S_a^{n-1}$  corresponds to the salience of  $i$ 's companion cue (a) when  $i$  is a  $Dr$  and it is activated in compound with  $Dr(a)$ . Hence, Eq. 6 introduces a differential speed of variation for a  $Dr$ 's salience when it is presented in compound with another  $Dr$ . If the  $Dr$  has no companion cue (i.e., if it is an elemental  $Dr$ ),  $S_a^{n-1} = 0$ . Finally,  $S_i^{n-1}$  designates the salience that the  $Rr$  or the  $Dr$  had immediately before trial  $n$ .

Therefore,  $S_{Dr(i)}$  will eventually conform to the outcome's salience if no other  $Dr$  is presented in compound or if it is activated in compound with another  $Dr$  at a different rate. The companion cue's salience will slow down the increase of  $Dr(i)$ 's salience during the AX–US trials but will accelerate its decrease when  $|S_o - S_a| > S_o$  which is the case during nonreinforced presentations of AX. This view is essential in understanding the current model in that it suggests that any representation tends to become more like the representation of what follows it in terms of both behavioral control and attention. Note that a representation's salience increases more quickly when the discrepancy between  $S_o^n$  and  $S_i^{n-1}$  is large (e.g., if the outcome is salient when cue  $i$  is not).

**Cue competition**

If X is tested after overshadowing treatment (i.e., repeated AX–US pairings),  $Dr(X)$  is associated with the US, but  $Rr(A)$  is not because it is activated by X in the absence of A for the first time at test. Thus, the CR during a test of X will be less than in a control condition in which X alone has been paired with the US. As we already mentioned, AAPV accounts for overshadowing because only a part of the representations present at test are associated with the US—namely,  $Dr(X)$  but not  $Rr(A)$ , because every previous presentation of X was accompanied by  $Dr(A)$  and not  $Rr(A)$ . As a consequence,  $V_{X \rightarrow US}$  approaches 1 (because  $Dr[X]$  was repeatedly paired with the US) and  $V_{Rr(A) \rightarrow US}$  remains at 0. Both  $S_X$  and  $S_A$  are high (i.e., .7) after many AX–US pairings because A's and X's saliences have increased at the same rate (Eq. 6) starting from the same initial value of .4. If one cue had a lower salience than the other, only the bigger of the two would have reached .7 while the other would have increased at a slower rate, never reaching .7 (Eq. 6). If  $S_A = .7$ ,  $S_{Rr(A)} = S_A = .7$ ,



because  $R_r(A)$  is activated of the first time at test. Consequently,  $R_X = S_X * [(S_X * 1) + (S_{R_r(A)} * 0)] / (S_X + S_{R_r(A)}) = .7 * [(.7 * 1) + [.7 * 0]] / (.7 + .7) = .35$ . This moderate value of  $R_X$  predicts that the CR will be attenuated, as compared with the condition in which X alone was paired with the US ( $R_X = .7$ ).

A similar account applies to overexpectation, a situation in which both cues A and X have been paired individually with the US prior to being paired in compound with the US (i.e., A–US interspersed with X–US followed by AX–US, which results in weaker responding to X than with a control group lacking either the A–US trials or the AX–US trials; see, e.g., Rescorla, 1970). However, X's and A's saliences increase more rapidly during phase 1 of an overexpectation procedure, as compared with an overshadowing procedure, because the rate of increase is not slowed down by any companion cue (Eq. 6). Thus, after phase 2 of overexpectation treatment (AX–US), X when tested alone has lost part of its potential to elicit the CR that it had acquired during phase 1, because it now activates a representation (i.e.,  $R_r[A]$ ) that is not associated with the US.

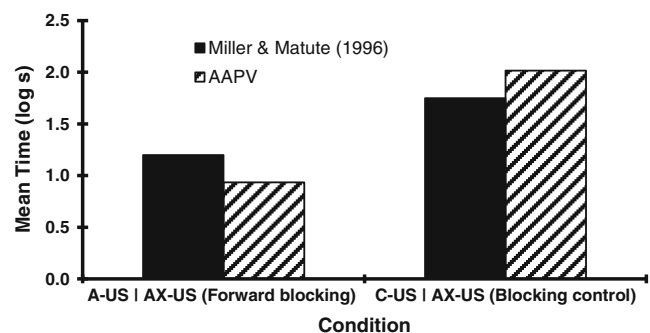
AAPV also makes valid predictions with respect to cue competition as seen in the relative stimulus validity effect. Wagner, Logan, Haberlandt, and Price (1968; see also Cole, Barnet & Miller, 1995) showed that cue X has less control over responding when it is trained in compound with a more valid predictor of the US (i.e., AX–US trials interspersed with BX–no-US trials) than with an equally valid predictor of the US (i.e., AX ± trials interspersed with BX ± trials) (where ± indicates 50 % partial reinforcement). According to AAPV, the more valid predictor's salience ( $S_A$ ) will reach the US's salience, thereby keeping X's salience ( $S_X$ ) from further increasing on subsequent AX–US trials, and  $S_X$  will decrease on BX–trials until  $S_B$  reaches the context's salience of .1 (see Eq. 6). Moreover, in the less valid condition, the strong  $R_r(A)$ –US association formed on BX–US trials will increase  $R_X$ .

Similarly, Hall, Mackintosh, Goodall, and Dal Martelo (1977; see also Arcediano, Escobar & Miller, 2004) conditioned an initially low-salience cue X alone (i.e., X–US). Cue X was then presented in reinforced compound with a novel initially more salient cue A (i.e., AX–US). The authors observed that the compound cue trials resulted in attenuation of behavioral control by X. More recently, these observations were supported by studies with humans (Denton & Kruschke, 2006). These authors argued that such results could be adequately interpreted only by models using variable attention. The observed decrease in responding to X is predicted by AAPV because the initial X–US pairings presumably increased both  $S_X$  and  $V_{D_r(X) \rightarrow US}$ . However, as a result of X being presented in compound with A, the test of X activated  $R_r(A)$  for which associative strength with the US was null. Consequently,  $R_X$  is no longer  $(S_X * V_{X \rightarrow US}) = .7 * 1 = .7$ , which corresponds to what happens as a result of the initial X–US pairings (Eq. 3), but is now  $S_X * [(S_X * 1) + (S_{R_r(A)} * 0)] / (S_X + S_{R_r(A)})$

$= S_{R_r(A)} = .7 * [(.7 + 0) / (.7 + .5)] = .41$ . Here,  $S_{R_r(A)}$  is set at .5 (and not .4) because A is described as a relatively salient stimulus. This interpretation of Hall et al.'s (1977) results is based on the principle that, because X is presented in compound with A, estimations of the CR amplitude to X are down modulated due to the intervention of another representation activated at test (i.e.,  $R_r[A]$ ) that has never been paired with the US.

Forward blocking is another situation in which the presence of a second cue during training of cue X down modulates the CR to X at test (e.g., Kamin, 1968; Miller & Matute, 1996; Shanks, 1985). In forward blocking, conditioning of A (i.e., A–US) is followed by a phase in which the compound AX is paired with the US (i.e., AX–US), with the result that responding to X is low, relative to a control group that receives only the AX–US pairings. Blocking is also explained by Eq. 5b (see Fig. 3). According to AAPV, the A–US pairings of phase 1 have a deflating effect on the manifestation of X's associative strength with the US that is mediated by a relatively low value for  $S_{D_r(X)}$  and a high value for  $S_{R_r(A)}$ . When X is tested,  $D_r(X)$  and  $R_r(A)$  are activated, and, because  $R_r(A)$  is activated for the first time at test (as was the case in overshadowing),  $V_{R_r(A) \rightarrow US}$  is equal to 0. At the time of test,  $S_{R_r(A)}$  is high because  $R_r(A)$ 's salience is assigned A's salience, which is high due to phase 1 treatment. That is, A's increased salience is transmitted to  $R_r(A)$ , when it is activated for the first time at test. However, according to Eq. 6, X's salience remains moderate (i.e., .4) during phase 2 because A's high salience prevents any increase ( $S_a = S_A$  and  $[S_o - S_A] = 0$  after phase 1). In this condition,  $R_X$ 's numerator is again reduced to  $S_X * S_X * 1$  because  $V_{R_r(A) \rightarrow US}$  is null, and  $V_{D_r(X) \rightarrow US} = 1$ . But if, this time, the denominator is smaller than it was in overshadowing because X's salience remains moderate, the numerator is much smaller also. Here,  $R_X = S_X * [(S_X * 1) + (S_{R_r(A)} * 0)] / (S_X + S_{R_r(A)}) = .4 * [(.4 * 1) + (.7 * 0)] / [.4 + .7] = .15$  after repeated AX–US pairings.

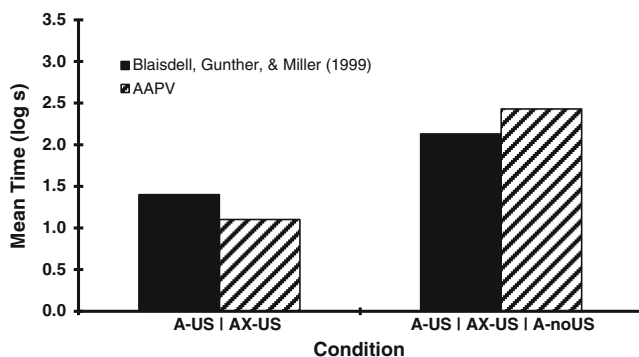
Posttraining extinction of X's companion cue (i.e., A–US followed by AX–US followed by A alone) increases  $R_X$ 's



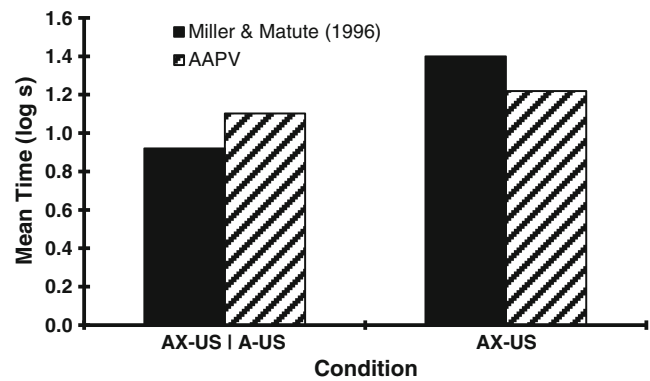
**Fig. 3** Miller and Matute's (1996, Experiment 3) results for forward blocking and its control, as compared with AAPV's predictions for  $R_X$ .  $k$  was determined by a least mean square fit of group means. For this data set,  $k = 7.41$

value (see Fig. 4), consistent with reports of the CR to X recovering from the blocking effect (i.e., retrospective reevaluation; see, e.g., Arcediano, Escobar & Matute, 2001; Blaisdell, Gunther & Miller, 1999; but see Dopson, Pearce & Haselgrove, 2009). Indeed, posttraining extinction of A (one manipulation that often induces retrospective reevaluation) will decrease A's salience after many trials. Therefore, when  $R_r(A)$  is activated for the first time at test of X,  $S_{R_r(A)}$  is low (i.e., .1) because A's current salience at test, which has decreased during phase 3 extinction treatment, is allocated to  $R_r(A)$  when it is activated for the first time. Therefore,  $R_X$ 's denominator is lower than what it would be in the simple blocking condition, and  $R_X$  is increased (i.e.,  $R_X = S_X * [(S_X * 1) + (S_{R_r(A)} * 0)] / (S_X + S_{R_r(A)}) = .4 * .4 + 0 / (.4 + .1) = .32$  after many A–no-US trials).

The view that forward blocking occurs because of a reduction in the processing of the blocked cue during the AX–US pairings has received support from recent studies (Kruschke & Blair, 2000; Le Pelley, Beesley & Suret, 2007). According to the Rescorla and Wagner (1972) model, it occurs because the association between A and the US has already been learned and, consequently, the US during the AX–US pairings is less surprising and, therefore, less processed. As we saw, AAPV does not primarily view blocking as a change in processing of either the target cue or the outcome at the time of acquisition but as the result of the effects of a salient retrieved representation of the blocking cue ( $R_r[A]$ ) at test, which has a null associative strength with the US. Admittedly, X's salience will be prevented from increasing during compound training as a result of A's high salience, but no decrease in the blocked cue's salience is predicted, and the blocked cue's associative strength will increase during the AX–US trials. Therefore, AAPV does not centrally explain blocking by a loss of salience or a decrease in accrued associative strength of the blocked cue but by the intervention at test of a highly salient representation of the blocking cue A (i.e.,  $R_r[A]$ ), the associative strength of which is null. In like manner, overshadowing



**Fig. 4** Blaisdell et al (1999, Experiment 3) results for responding to X following extinction of A after forward blocking compared to AAPV's predictions for  $R_X$ .  $k$  was determined by a least mean square fit of group means. For this data set,  $k = 8.74$



**Fig. 5** Miller and Matute's (1996, Experiment 2) results for backward blocking and a control group lacking the A–US trials, as compared with AAPV's predictions for  $R_X$ .  $k$  was determined by a least mean square fit of group means For this data set,  $k = 4.48$

is viewed primarily not as a lack of acquisition of the associative strength, but as resulting from the activation at test of a novel representation ( $R_r[A]$ ), the associative strength of which is null.

A reduction of blocking is observed when a second US is added during the compound training phase (i.e., A–US1 followed by AX–US1–US2; see, e.g., Dickinson, Hall & Mackintosh, 1976; Dickinson & Mackintosh, 1979). This observation is anticipated by AAPV because the addition of a second US increases the compound outcome's salience. If  $S_A$  is increased during phase 1, it is increased only to the level of US1's salience. Therefore, when phase 2 begins,  $S_A$  is weaker than the new compound outcome's salience and Eq. 6's  $|S_o - S_a|$  (which corresponds here to  $|S_{US1 + US2} - S_A|$ ) is positive, allowing  $S_X$  to increase. Interestingly, these authors also showed that omitting the second US during the compound training phase (i.e., A–US1–US2 followed by AX–US1) also results in reduced blocking despite the fact that the surprising outcome seemingly corresponds to a loss of salience. AAPV also accounts for that result by assuming that A's salience in this case will decrease during the compound training phase to conform to the reduced outcome's salience. Here again, the absolute value of the discrepancy between  $S_o$  and  $S_A$  at the beginning of phase 2 makes  $S_X$ 's increase possible. If  $S_X$  increases, blocking is reduced, as previously explained.

Kruschke and Blair (2000) showed that one consequence of blocking is subsequent retardation in the learning process with the blocked cue, and they explained this in terms of learned inattention. AAPV accounts for this observation by assuming that in Eq. 5b, the denominator's high value of  $S_{R_r(A)}$  (i.e.,  $S_{R_r(A)} = .7$ ) will remain a down modulating factor for  $R_X$  until  $V_{R_r(A)}$  has reached 1—that is, until many X–US pairings (following the AX–US pairings) have been completed.

Backward blocking (AX–US followed by A–US; see, e.g., Miller & Matute, 1996; Shanks, 1985) is also predicted by AAPV, but it is expected to be weaker than what is observed

with forward blocking (see Fig. 5). The reason for this is that the saliences of both A and X will increase during phase 1 of backward blocking so that eventually both of them approach .7 (assuming that X and A were of equivalent initial saliences). Therefore, backward blocking is predicted to occur only when few AX–US pairings have been made, as was the case in Miller and Matute’s study. Admittedly, those authors attributed their success in obtaining backward blocking to the sensory preconditioning procedure they used. Interestingly, Miller, Hallam, and Grahame (1990) failed to obtain backward blocking in Pavlovian conditioning with nonhumans after relatively many AX–US pairings and when no sensory preconditioning procedure was used to make A excitatory. These failures indicate, as AAPV anticipates, that backward blocking is less readily observed than forward blocking. But the relative contributions of sensory preconditioning and number of compound trials are still unclear.

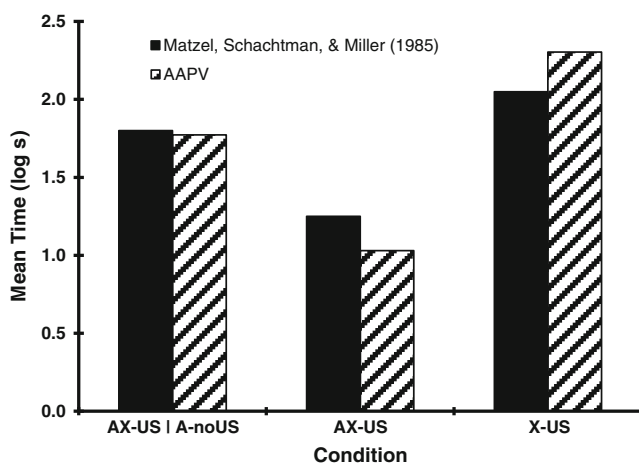
More generally, retrospective revaluation refers to a change in behavioral control by a target cue as a result of associative inflation or deflation of its companion cue with which it was previously paired. Retrospective revaluation is a serious challenge for many earlier models of learning because these models do not anticipate any change in the associative strength between a cue and the US on a trial on which the cue is absent, and no other variable (such as salience) is expected to influence conditioned responding. Backward blocking (i.e., AX–US followed by A–US) is one of the simplest instances of retrospective revaluation and, as we saw, is explained by AAPV. In like manner, release from overshadowing, in which overshadowing of X by A is attenuated as a result of posttraining extinction of A (i.e., AX–US followed by A–no-US; see, e.g., Kaufman & Bolles, 1981; Matzel et al., 1985), is explained by AAPV’s assumption that  $S_A$  decreases towards .1 (i.e., the context’s salience) during the

second phase (see Fig. 6). When X is tested,  $Rr(A)$  is activated for the first time.  $V_{Rr(A)}$  is null, and  $S_{Rr(A)}$ , receiving A’s salience, is also low due to the extinction of A. X’s salience will be high after many AX–US pairings because it progressively increases during phase 1. Therefore,  $R_X = S_X * [(S_X * V_X) + (S_{Rr(A)} * V_{Rr(A)})] / (S_X + S_{Rr(A)}) = .7 * [(.7 * 1) + (.1 * 0)] / (.7 + .1) = .6$  if phases 1 and 2 are prolonged. This instance of retrospective revaluation (i.e., recovery from overshadowing) illustrates that the model is able to anticipate modification of the manifestation of a cue’s associative strength with the US even when the cue is no longer presented, by modulating the salience of the representation of the target cue’s companion cue.

### Conditions involving more than one nontarget cue

AAPV makes clear predictions concerning what happens when cue X is presented in compound with different nontarget cues. For example, if AX–US pairings are intermixed with BX–US pairings, AAPV anticipates a high CR to X. Because test of X activates  $Rr(A)$  as well as  $Rr(B)$ ,  $R_X = S_X * [(S_X * V_X) + (S_{Rr(A)} * V_{Rr(A)}) + (S_{Rr(B)} * V_{Rr(B)})] / (S_X + S_{Rr(A)} + S_{Rr(B)}) = .7 * [(.7 * 1) + (.7 * 1) + (.7 * 1)] / (.7 + .7 + .7) = .7$ . The saliences of  $Rr(A)$  and  $Rr(B)$  are high because  $Rr(A)$  is activated during the BX–US pairings and  $Rr(B)$ , during the AX–US pairings. Both of these representations are repeatedly followed by the US. Hence, the principle of Eq. 5b is easily applied to conditions in which multiple Rrs are activated.

Complex forms of retrospective revaluation involving more than one nontarget cue have been described in the associative learning literature. For example, Denniston, Savastano, Blaisdell and Miller (2003) reported that after AB–US had been experienced in phase 1 and AX–US in phase 2, adding in a third-phase B–no-US trials decreased responding to X, as compared with the condition in which no B–no-US trials were administered in phase 3. This result is generally described as second-order retrospective revaluation in which associative deflation of B increases the manifestation of A’s associative strength, which, in turn, results in down modulating the manifestation of X’s associative strength. De Houwer and Beckers (2002) obtained comparable results using a causal learning task with humans. The latter authors also showed an increase in the response to X when B was reinforced in phase 3 rather than extinguished (i.e., AB–US followed by AX–US and followed by B–US pairings), as compared with what happens if phase 3 does not occur. In general, AAPV is able to explain second-order retrospective revaluation. AAPV accounts for the increase in  $R_X$  when B is reinforced during phase 3 in De Houwer and Becker’s experiment as follows. The B–US trials cause  $V_{Rr(A)}$  to increase to 1 and also increase  $S_{Rr(A)}$  as a result of the  $Rr(A)$ –US pairings. As for  $S_{Rr(B)}$ , it was increased during phase 2.  $Dr(X)$ , being



**Fig. 6** Matzel et al (1985, Experiment 1) results for release from overshadowing, as compared with AAPV’s predictions for  $R_X$ .  $k$  was determined by a least mean square fit of group means. For this data set,  $k = 3.32$

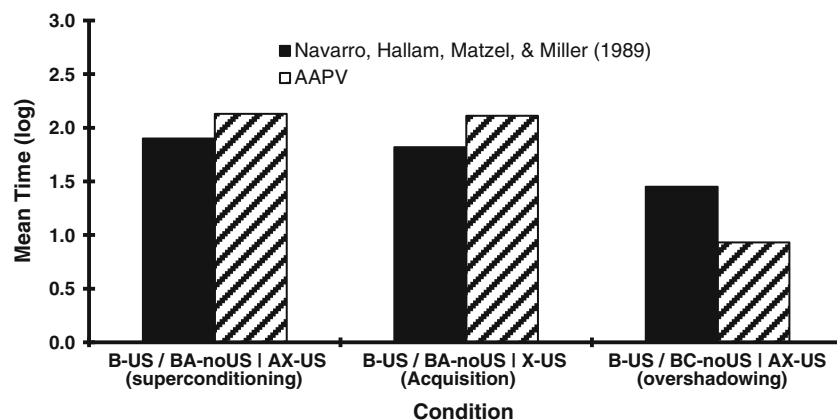
repeatedly compounded with Rr(B) (which was activated by A) during phase 2, acquires the potential to activate Rr(B) at test. Applying Eq. 5b,  $R_X = S_X * [(S_X * 1) + (S_{Rr(A)} * 1) + (S_{Rr(B)} * 1)] / (S_X + S_{Rr(A)} + S_{Rr(B)}) = .4 * (.4 + .7 + .7) / (.4 + .7 + .7) = .4$ .  $S_X$  does not increase during phase 2 because, if phase 1 is prolonged,  $S_A$  becomes maximal, thereby preventing  $S_X$  from increasing (see Eq. 6). As each  $V_i$  tends toward 1,  $R_X$  approaches 1, which is larger than the .23 predicted if phase 3 were omitted (the numerator in the latter condition is smaller as a result of  $V_{Rr(A)}$  being null). However, while SOCR makes relatively accurate predictions for both phase 3 conditions (i.e., B–US trials and Bno–US trials), AAPV fails to account for the decreased responding to X caused by phase 3's extinction of B in the specific condition used by Denniston et al. (2003) because B–no–US trials decrease Rr(A)'s salience. Consequently, Eq. 5b's denominator is decreased, and not increased, as compared with the no phase 3 condition.

Another phenomenon in which multiple nontarget cues are involved is superconditioning, in which reinforcement of the compound AX (i.e., AX–US) in phase 2 has been preceded by inhibitory training with A (i.e., interspersed B–US and AB–no–US trials) in phase 1. The finding is stronger responding to X than if A had not undergone inhibitory training. When X is tested,  $V_X$ ,  $V_{Rr(A)}$ , and  $V_{Rr(B)}$  are high because activation of these three representations has been repeatedly paired with the US.  $S_{Rr(A)}$  and  $S_{Rr(B)}$  increase from .4 toward .7 during phase 1 and phase 2, respectively. Consequently,  $R_X$  is expected to be high, as was observed by Navarro, Hallam, Matzel, and Miller (1989; see Fig. 7). However, these authors did not observe a greater response in the superconditioning condition, as compared with the condition with elemental training with X (i.e., X–US). Adding a third phase in which A is extinguished should decrease Rr(B)'s salience and, consequently, decrease  $R_X$ , consistent with what Urushihara, Wheeler, Pineño and Miller (2005) observed. Therefore, AAPV can also explain a number of observations obtained with complex designs involving more than one Rr activated by cue X.

## Latent inhibition

Latent inhibition refers to weak responding to X during and following X–US pairings in phase 2 as a result of X–no–US trials in phase 1. AAPV explains latent inhibition through the reduction of X's salience from .4 to .1 during phase 1 as predicted by Eq. 3. When X is paired with the US in phase 2,  $S_X$  will slowly grow from .1 to .7. Because  $S_X$  is weak at the beginning of phase 2,  $V_X$  will increase only slowly, as compared with unfettered acquisition in the absence of phase 1. Hence, latent inhibition is viewed as resulting from the loss of salience that X undergoes when it is preexposed. Because  $V_X$  is a function of the cue's salience,  $V_X$ 's increase will be slowed down when the cue's salience is low. In sum, inhibition results both from low  $S_X$  and from low  $V_X$ . We later discuss AAPV's account of why latent inhibition is context specific.

Schmajuk, Lam and Gray (1996) proposed in their SLG model that latent inhibition is manifested because preexposure of X decreases X's novelty (i.e., total novelty [Novelty] is limited to X's novelty if X alone is presented and no other cue is expected), and this decrease in novelty results in a slow rate of acquisition of  $V$ . Schmajuk et al. refer to the orienting response's amplitude as an index of the amount of processing afforded to a cue as a result of a mismatch between what was expected and external reality (Sokolov, 1963). According to SLG, because the amount of processing a cue receives is proportional to the cue's novelty (i.e., the degree to which a cue was unexpected), a cue introduced for the first time is assumed to receive more processing (i.e., it is more salient) than a preexposed cue (i.e., a familiar cue) that has been followed by no significant event, which is similar to AAPV's account. Thus, both the SLG model and AAPV explain latent inhibition by a decreased attention to the cue as a result of preexposure, a position similar to that of several other models as well. Moreover, the SLG model anticipates Holmes and Harris's (2009) observation that latent inhibition to X disappears with prolonged compound conditioning of AX. Simulations



**Fig. 7** Navarro et al (1989, Experiment 4) results for superconditioning, as compared with AAPV's predictions for  $R_X$ .  $k$  was determined by a least mean square fit of group means. For this data set,  $k = 3.11$



from the SLG model show less conditioned responding to the preexposed cue X than to the novel cue A after four AX–US pairings, but this difference is decreased after 20 AX–US pairings (Schmajuk & Kutlu, 2011). AAPV likewise anticipates a discrepancy between responding to the preexposed cue, as compared with the novel cue, that decreases over the AX–US pairings. AAPV explains the initial discrepancy between A's and X's response amplitudes by assuming that X has lost some salience during preexposure and, consequently,  $V_X$  increases more slowly than  $V_A$ . X's and A's saliences should both increase during the AX–US pairings. As a consequence,  $V_X$  will keep increasing even after  $V_A$  reaches asymptote, thereby reducing the discrepancy between responding to X and to A. However, consistent with Holmes and Harris's observation, the SLG model anticipates a greater decrease in the discrepancy between the responses to the two cues over the AX–US pairings than does AAPV.

Mercier and Baker (1985) observed a latent inhibition effect to X even if X preexposure was within compound AX (i.e., AX–no–outcome followed by X–US trials). More recently, Rodriguez and Hall (2008) observed even greater retardation of behavioral control by a cue (X, an odor) when rats had been preexposed to the cue within a compound including that cue (i.e., AX–no–outcome), as compared with a condition in which subjects were preexposed to the cue alone. According to AAPV, in the condition in which the compound is preexposed, the denominator of Eq. 5b is larger because  $S_X$  and  $S_{R(A)}$  are summed. In that condition, both  $V_{X \rightarrow US}$  and  $V_{R(A) \rightarrow US}$  are weak after one or two cue–US pairings of phase 2; therefore, the numerator is small and the denominator is large. Hall and Rodriguez (2011) proposed an account of latent inhibition based on the Pearce and Hall (1980) model. The results of their experiment confirm a greater retardation effect when subjects are subjected to preexposure of the cue within a compound, relative to the cue alone. They further demonstrated that this effect could be attenuated by introducing into the design a blocking-like phase—that is, a phase in which X's companion cue A is preexposed alone before being preexposed in compound with X (i.e., A–no–stimulus followed by AX–no–stimulus, followed by X–US pairings). According to AAPV, after six presentations of A alone in phase 1, A's salience has decreased from .4 to .26. During the six preexposures of the compound AX in phase 2, A's salience will continue its decrease and eventually slow down the decrease of X's salience because of the small value of  $|S_o - S_a|$  (i.e.,  $|S_{context} - S_a|$ ; see Eq. 6). Consequently, X will not lose as much of its salience (i.e.,  $S_X = .27$ ) as in the condition in which no blocking phase was introduced (i.e.,  $S_X = .19$ ). X's salience will be larger in the former condition; consequently, the retardation effect will be less pronounced. That is,  $V_X$  will increase more rapidly because acquisition of  $V_X$  is a function of  $S_X$ . However, Reed, Anderson and Foster (1999) observed that this blocking condition provoked enhanced latent inhibition, as compared with

a condition in which phase 2's AX–no–US trials are replaced by X–no–US trials (i.e., A–no–US followed by X–no–US, followed by X–US trials). AAPV predicts this result provided A is not extensively nonreinforced during phase 1. The presence of A in the compound AX accelerates the decrease of X's salience if  $S_A > .2$ . When  $.1 < S_A < .2$ ,  $|S_o - S_a| < S_o$ , so that the factor modulating the variations of  $S_X$  within the compound AX (i.e.,  $|S_o - S_a|$ ; see Eq. 6) is smaller than the factor that modulates  $S_X$  when it is presented alone (i.e.,  $S_o$ ; see Eq. 2).

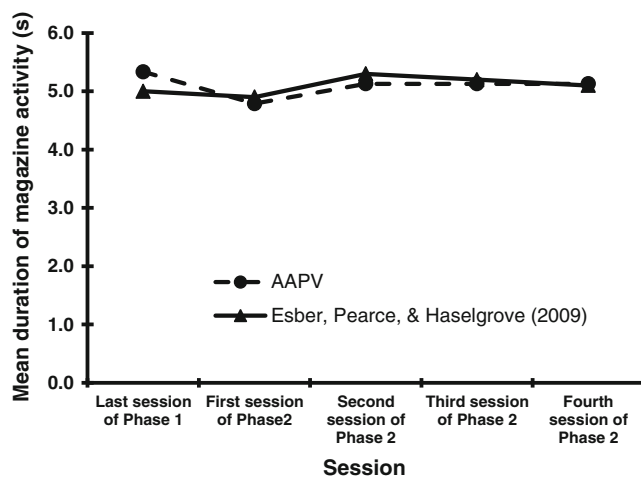
As was previously mentioned, counterconditioning refers to X–US1 trials being followed by X–US2 trials, where US1 and US2 are of opposing valences (Pavlov, 1927). The usual observation is that the X–US2 trials reduce responding to X that reflects the X–US1 pairings. AAPV predicts that  $R_{X \rightarrow US2}$  (i.e., a conditioned response appropriate for US2) will increase more rapidly than it would in a latent inhibition situation (i.e., X–no–US followed by X–US). As before, following X–no–US trials, learning of  $V_{X \rightarrow US}$  is retarded by X's low salience. However, during counterconditioning, learning of  $V_{X \rightarrow US2}$  takes place when  $S_X$  is high due to X–US1 pairings during phase 1. Because  $S_X$  modulates  $V_X$ 's rate of increase (Eq. 1),  $V_{X \rightarrow US2}$  will rapidly increase. This prediction is consistent with the observation of facilitation rather than retardation of conditioning when a CS has previously been paired with another outcome (i.e., counterconditioning), as compared with conditioning when the CS has previously been preexposed alone (i.e., latent inhibition; see, e.g., Dickinson & Pearce, 1977; Killcross, Dickinson & Robbins, 1995). However, seemingly opposite results were reported by Hall and Pearce (1979). They showed retardation rather than facilitation of conditioning when the CS was reinforced with a relatively weak shock in a first phase before being paired with a stronger shock, as compared with a group in which a new CS was paired with the stronger shock. AAPV explains their results by assuming that, if the shock was weak during phase 1, its salience weakened during that phase because it was followed by no event. Admittedly, we earlier stated that a US does not lose its salience, but this is assumed only for biologically significant outcomes. If Hall and Pearce's weak shock of phase 1 was not strong enough to be a biologically significant outcome (although the authors did see some responding following low shock in phase 1), its salience is bound to decrease during phase 1 because a cue not followed by any other cue loses its salience. While the outcome's salience decreases, the  $S_{CS}$  decreases too according to AAPV. Therefore, the Hall–Pearce effect is explained the same way as the latent inhibition effect, by a weak CS's salience.

## Recent data

As we saw, many of the different benchmark learning phenomena are readily explained by AAPV. Moreover, recent

years have seen reports of a growing number of results that, although not of benchmark status (in part, because they have not yet been widely replicated), are, in good part, adequately predicted by AAPV. For example, the comparator hypothesis (Stout & Miller, 2007) has recently been challenged by Esber, Pearce and Haselgrove (2009) because one of its assumptions is that, once X has been presented in compound, the expression of its associative strength is regulated by the associative strength between its companion cue and the US. As Esber et al. point out, according to the comparator hypothesis, introducing AX–US pairings interspersed among X–US pairings down modulates the CR to X because of the A–US pairings within the AX–US pairings. However, contrary to the comparator hypothesis' predictions, Esber et al. found that introducing AX–US pairings interspersed among X–US pairings after a first phase of X–US pairings did not alter X's potential to elicit the CR. AAPV anticipates only a slight reduction of the CR to X as a result of introducing the AX–US pairings, but importantly, it predicts a rapid recovery of X's full CR with continuing interspersed X–US and AX–US trials (see Fig. 8). This occurs due to the intermixed presentations of the X–US and the AX–US pairings causing  $Rr(A)$  to be paired with the US on each of the X–US pairings. The  $Rr(A)$  pairings with the US will increase both  $S_{Rr(A)}$  and  $V_{Rr(A) \rightarrow US}$ . Consequently,  $R_X = S_X * [(S_X * 1) + (S_{Rr(A)} * 1)] / (S_X + S_{Rr(A)}) = .7 * (.7 + .7) / (.7 + .7) = .7$ , which produces the same result as when only X–US pairings are presented (i.e.,  $[S_X * V_{X \rightarrow US}] = .7 * 1 = .7$ ). As Fig. 8 shows, AAPV makes predictions close to what Esber et al. observed.

Some recent studies have provided support for either the Mackintosh (1975) model or the Pearce and Hall (1980)



**Fig. 8** Esber et al (2009) results and AAPV's predictions concerning responding to X when AX–US pairings interspersed with X–US pairings followed a first phase in which X–US pairings were presented.  $k$  was determined by a least mean square fit of group means. For this data set,  $k = 7.62$ . The last session of phase 1 corresponds to the last session in which X was paired alone with the US. The four phase 2 sessions correspond to tests of X when X-alone pairings with the US are intermixed with AX–US pairings

**Table 1** Design of the experiment analyzed by Le Pelley et al. (2011)

Phase 1	Phase 2	Test
AV – o1	AX – o3	AC
AW – o1	BY – o4	BD
BV – o2	CV – o3	VX
BW – o2	DW – o4	WY
CX – o2		
CY – o2		
DX – o1		
DY – o1		

*Note.* Letters A to D and V to Y refer to different cues; o1 to o4 refer to outcomes. At test, ratings of the indicated cue compounds were obtained with respect to outcomes o3 and o4. All participants experienced all types of trials listed during a given stage of training. Results show that AC and BD are more predictive of o3 and o4, respectively, than are VX and WY, presumably because A, B, C, and D, in contrast with V, X, W, and Y, consistently predicted a unique outcome during phase 1.

model. For example, Le Pelley, Beesley and Griffiths (2011) concluded that Le Pelley and McLaren's (2003) observation that predictiveness influences attention in human causal learning could be explained by the Mackintosh model, but not by the Pearce and Hall model (see Table 1). In this study, compounds AC and BD were more predictive of o3 and o4, respectively, than were VX and WY because A, B, C, and D, in contrast with V, X, W, and Y, consistently predicted a unique outcome during phase 1. The former set of cues, being better predictors of their outcomes, received more attention than the latter according to Mackintosh's model. AAPV assumes that, because increases in the salience of a cue depends on its associative strength with its outcome (see Eq. 2), cues V, W, X, and Y had less chance than cues A, B, C, and D during phase 1 to increase their salience because their associative strengths with the two outcomes with which they were paired were presumably not asymptotic. Due to associative strength ( $V$ ) being smaller for these cues, the increase of salience for these cues presumably was slowed down.

Some of the results observed when subjects have to learn to discriminate between two cues on the basis of their outcomes could possibly be explained without reference to changes in attention. Hall (1991; see also Hall & Rodriguez, 2010) proposed an interpretation in terms of acquired distinctiveness as an alternative to explanations based on learned attention. The basic principle of Hall's theory is that, if cue A and cue X are separately trained with different outcomes (o1 and o2, respectively) during phase 1, A and X increase their distinctiveness, as compared with cues that are trained with the same outcomes. Hall's proposal is that A and X can be differentiated not only on the basis of the Dr they activate, but also on the type of anticipatory outcome representation that they activate. Hence, if A and X are trained with two different outcomes (o3 and o4) during phase 2 of discrimination test, A and X will

more readily be discriminated (i.e., there will be a positive transfer of what was learned during phase 1) because of the distinct outcome representations they activate, as compared with a condition in which phase 1 was omitted or in which the two cues had the same outcome. Hall's theory can explain the difference between interdimensionality shifts (IDSs) and extradimensionality shifts (EDS; i.e., the IDS–EDS effect, which is the superiority in learning after an IDS, as compared with an EDS). Although acquired distinctiveness is a valid alternative to theories of attention to explain the IDS–EDS effect, at least one set of results challenges this type of interpretation. Dopson, Esber, and Pearce (2010; see also Pearce & Mackintosh, 2010) trained compound cues during a first phase in which cue A was consistently followed by the US regardless of its companion cue (e.g., Av–US; Ax–US), while C was never followed by the US regardless of its companion cue (e.g., Cv–no-US, Cx–no-US). Both A and C were relevant cues because they were sufficient to respond correctly, whereas v and x were not. In like manner, B and C were the relevant cues for discrimination between the two conditions Bw–US and Cw–no-US. During test, pigeons had to learn to discriminate between Av–no-US and a new reinforced compound (Aw–US) and between this new compound (i.e., Aw–US) and a Bw–no-US condition. More rapid discrimination was observed between Bw and Aw than between Aw and Av, whereas Bw and Av could not be differentiated on the basis on their activation of an anticipated outcome, making difficult any interpretation in terms of acquired distinctiveness. AAPV explains these results in terms of increased attention (i.e., increased salience) granted to A and B, as compared with v and w, because the latter cues were only sometimes followed by the US, while the former were well associated with the US. Therefore, AAPV's interpretation is equivalent to Dopson et al.'s account according to which more attention is paid to relevant cues than to irrelevant cues in solving a discrimination task.

### Challenges to AAPV

A major phenomenon that a model of Pavlovian conditioning should be able to explain is conditioned inhibition. Typical procedures to produce conditioned inhibition involve interspersing trials on which a cue is reinforced when it is presented alone (i.e., X–US) and not reinforced when in compound with another cue (i.e., AX–no-US). A test of X (or a transfer excitator Y) alone results in a strong CR, whereas a test of AX (or AY) results in a weak CR. Because above we considered only the test of a single cue X and not of a compound AX, AAPV in its present form cannot explain conditioned inhibition. To do so, AAPV would have to assume that the degree of inhibition exercised by A on the CR to X during the AX–no-US presentations (or during an AY summation test, where Y is

a transfer conditioned excitator) is proportional to the degree to which A's salience accelerates the decrease of X's salience during inhibition training according to Eq. 6 (i.e.,  $S_A$  accelerates the decrease of  $S_X$  when  $S_A > 0.2$  because Eq. 6's  $|S_O - S_A|$  is larger than Eq. 2's  $S_O$ ). This assumption would also permit an account of renewal, provided configuring the CS and extinction context is assumed to occur after many extinction trials. But this assumption introduces appreciable complexity to the model, so we refrained from integrating it into the model presented here. Instead, it is a possibility that might be worth considering in the future.

Other conditions constituting real limitations in the predictive potential of the model need be considered. In common with many other models of Pavlovian conditioning, spontaneous recovery from extinction (Pavlov, 1927) cannot be explained by AAPV unless one assumes that increasing novelty or the degree of surprise generated by the reintroduction of the cue somehow increases the cue's salience. This seems to be a reasonable assumption, but it is difficult to capture in a principled manner within the tenets of the current model. AAPV is also limited in its prediction of what happens as far as the type of CR is concerned during counterconditioning (i.e., X–US1 followed by X–US2). As previously described, rapid acquisition of the second associative strength is predicted by the current model, but it is not clear which CR will be manifested after phase 2 because both  $R_{X/US1}$  and  $R_{X/US2}$  are high. Presumably, the second CR replaces the first one through some sort of response competition, but here again it is difficult to describe the sequential occurrence of this in a principled manner within the current framework of this model.

Another problem with AAPV is that the potential of X to activate  $R_r(A)$  at test depends on the within-compound link between X and A. However, nothing in the present equations represents the link between two compounded cues. Presumably, the link between a cue and its associate cue is rapidly learned given simultaneous presentation of the two cues, and for the sake of simplicity we have assumed that the within-compound link is maximal after only one presentation of the compound. The omission of a parameter modulating the rate of increase of the within-compound link is a flaw that impacts a few of AAPV's predictions, but has the advantage of keeping the equations tractable. The simplest way to consider such a parameter would be to assume that this link progressively increases during compound presentations of X and A in a manner paralleling acquisition of associations between a CS and a US. No mutual modulation of saliences would be expected, because AAPV assumes that saliences increase or decrease only according to the outcome's salience and with simultaneous presentation of two cues there is no clear outcome. Among the phenomena that AAPV could account for if such a parameter was added is the unblocking effect observed by Mackintosh and Turner (1971). These authors found that, if a stimulus A is paired with a weak shock (us) during phase 1

before being paired with a strong shock (US) in compound with the cue X (i.e., A–us followed by AX–US), blocking of the CR to X will be weaker, as compared with a condition in which AX has been paired several times with the weak outcome (us) prior to being paired with the strong US (i.e., A–us before AX–us before AX–US). Presumably, the within-compound link between A and X is initiated during the AX–us pairings of phase 2 of the latter condition. Consequently, A can exercise a greater blocking effect than in the former condition, in which the within-compound link is initiated only during the AX–US pairings of last phase.

Yet another challenge comes from the observation that a cue that has been reinforced on a partial reinforcement schedule attracts more attention than a cue reinforced on a constant reinforcement schedule (Kaye & Pearce, 1984). This effect is not captured by AAPV's equations, because they predict (like Mackintosh, 1975) that attention growth is controlled by the outcome's salience, and not by the unexpected nature of the outcome on a given trial.

Despite the limitations here presented, AAPV makes accurate predictions for many of the most well-known phenomena of Pavlovian conditioning and competes with other models by offering another perspective on the interpretation of these phenomena.

## General discussion

AAPV can be summarized by stating the three central tenets on which it is based: First, AAPV proposes not only that the associative strength between a cue and its outcome increases with repeated pairings of the cue and the outcome, but also that its salience will increase or decrease to finally reach the outcome's salience. A stimulus followed by a US ends up eliciting a response appropriate for the US not only because of the strong X–US association, but also because it has become as salient as the US.

Second, when X is tested alone, not only will X's direct representation (i.e.,  $Dr[X]$ ) be activated, but also activated will be the retrieved representations of X's former companion cues (i.e., first-order companions), as well as retrieved representations that have repeatedly been activated in the presence of X by one of its companion cues (i.e., second-order companions). The Rrs activated at test correspond to cues that are not present in the environment at the time of test but that have formerly been directly or indirectly presented in compound with cues that are present. This assumption differs from the similar one of Wagner's (1981) SOP model in that in AAPV the representation of presented cues does not decay into representations indistinguishable from retrieval activated representations. AAPV assumes that, because associative strengths are not between stimuli but between representations of stimuli,

retrieved representations, as well as direct representations, should express their associative strengths with the US. In that sense (among others), AAPV distinguishes itself from Stout and Miller's (2007) comparator hypothesis, which anticipates that the expression of the associated strength of a test cue X is controlled by  $Dr(A)$ 's (but not  $Rr[A]$ 's) associative strength with the US's representation. AAPV, contrary to the comparator model, is not concerned with the associative strength of the direct representation of A when X is tested alone. Third, attention at the time of performance influences responding. The amplitude of the CR to X is the result of the combination of these representations' associative strengths with the US, each accordingly weighted by current attention (i.e., salience) to that representation.

In summary, AAPV is a model of learning in which changes in attention are central. This aspect is certainly not unique to AAPV, since numerous prior models introduced attention as a variable in the learning process. However, AAPV's uniqueness as a model of attention is that it assumes that attention varies independently according to what follows activation of retrieved representations and what follows activation of direct representations. If what follows activation of a representation (directly activated or retrieval activated) is of high salience, the representation itself becomes more salient. If what follows is of low salience, the representation's salience will decrease to conform to the outcome's low salience. Importantly, AAPV assumes that attention is critical not only during learning, but also during performance. Considering attention as a performance, as well as a learning, variable has been defended by Kruschke in his ADIT and EXIT models (e.g., Kruschke, 2001) and by Schmajuk et al. (1996) in their SLG model. In the latter model, those authors assume that postconditioning manipulations such as extinction of the training context will increase total novelty of the target cue, thereby attenuating any potential latent inhibition effect by incrementing the magnitude of the CR to X. As such, they suggest that attention influences not only learning, but also performance. However, to our knowledge, the first to suggest implementation of attention as a performance variable in a model of learning was Mackintosh (1975). By including salience as a performance variable, even if  $V_X$  has increased, its expression will be inhibited if  $S_X$  is weak. Recently, Le Pelley, Suret and Beesley (2009) developed Mackintosh's model, incorporating  $S_X$  into the equation for responding to X at test (i.e.,  $R_X = k * S_X * V_X$ ). Equation 2b in the present article resembles those authors' equation for  $R_X$ , but AAPV suggests inclusion of all the representations activated at test, not just cues that are presented. Not only is  $Dr(X)$ 's salience thought to modulate performance at test, but additionally the saliences of other activated representations are assumed to have the same effect. This extension to all of the activated representations, as demonstrated, considerably enlarges the scope of empirical data that AAPV can explain.



As we saw, these fundamentals led to successful simulations of the different benchmark learning phenomena considered in this article. AAPV surely is in need of refinement, but its current presentation can explain numerous published observations. It is based on simple equations incorporating the principle that all activated representations contribute to the modulation of the CR to X at test according to their saliences, which are determined by the outcome that followed them during training. Moreover, it emphasizes the role of attention at the time of testing, as well as training. It is these principles, as opposed to the details of the model proposed here, that we anticipate will be vindicated in the future.

**Author notes** Support for the preparation of this article was provided by NIMH grant 33881. We thank Henry Cham, Mario Laborda, Gonzalo Miguez, Mikael Molet, and Cody Polack for their comments on an earlier version of the manuscript. Electronic supplementary material (simulator) is available at <http://psitec.recherche.univ-lille3.fr/membres/craddock/aapvsimulator.html>.

## References

- Arcediano, F., Escobar, M., & Matute, H. (2001). Reversal from blocking in humans as a result of posttraining extinction of the blocking stimulus. *Animal Learning & Behavior*, *29*, 354–366.
- Arcediano, F., Escobar, M., & Miller, R. R. (2004). Is stimulus competition an acquisition deficit or a performance deficit? *Psychonomic Bulletin & Review*, *11*, 1105–1110.
- Blaisdell, A. P., Gunther, L. M., & Miller, R. R. (1999). Recovery from blocking through deflation of the blocking stimulus. *Animal Learning & Behavior*, *27*, 63–76.
- Bouton, M. E. (2004). Context and behavioural processes in extinction. *Learning & Memory*, *11*, 485–494.
- Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation*, *10*, 445–466.
- Bouton, M. E., & Swatzentruber, D. (1989). Slow reacquisition following extinction: Context, encoding, and retrieval mechanisms. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*, 43–53.
- Brogden, W. J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology*, *25*, 323–332.
- Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: Wiley.
- Cole, R. P., Barnet, R. C., & Miller, R. R. (1995). Effect of relative stimulus validity : Learning or performance deficit? *Journal of Experimental Psychology: Animal Behavior Processes*, *21*, 293–303.
- De Houwer, J., & Beckers, T. (2002). Second-order backward blocking and overshadowing in human causal learning. *Experimental Psychology*, *49*, 27–33.
- Denniston, J. C., Savastano, H. I., Blaisdell, A. P., & Miller, R. R. (2003). Cue competition as a retrieval deficit. *Learning and Motivation*, *34*, 1–31.
- Denton, S. E., & Kruschke, J. K. (2006). Attention and salience in associative blocking. *Learning and Behavior*, *34*, 285–304.
- Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective reevaluation of causality judgments. *Quarterly Journal of Experimental Psychology*, *49B*, 60–80.
- Dickinson, A., Hall, G., & Mackintosh, N. J. (1976). Surprise and the attenuation of blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*, 313–322.
- Dickinson, A., & Mackintosh, N. J. (1979). Reinforcer specificity in the enhancement of conditioning by posttrial surprise. *Journal of Experimental Psychology: Animal Behavior Processes*, *5*, 162–177.
- Dickinson, A., & Pearce, J. M. (1977). Inhibitory interactions between appetitive and aversive stimuli. *Psychological Bulletin*, *84*, 690–711.
- Dopson, J. C., Esber, G. O. R., & Pearce, J. M. (2010). Differences in the associability of relevant and irrelevant stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*, 258–267.
- Dopson, J. C., Pearce, J. M., & Haselgrove, M. (2009). Failure of retrospective reevaluation to influence blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 473–484.
- Esber, G., & 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.
- Esber, G., Pearce, J. M., & Haselgrove, M. (2009). Enhancement of responding to A after A+ / AX + training: Challenges for a comparator theory of learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 485–497.
- Gallistel, C. R., & Gibbon, J. (2000). Time, rate and conditioning. *Psychological Review*, *107*, 289–344.
- Hall, G. (1991). *Perceptual and associative learning*. Oxford: Clarendon Press.
- Hall, G., Mackintosh, N. J., Goodall, G., & Dal Martelo, M. (1977). Loss of control by a less valid or by a less salient stimulus compounded with a better predictor of reinforcement. *Learning and Motivation*, *8*, 145–158.
- Hall, G., & Pearce, J. (1979). Latent inhibition of a CS during CS-US pairings. *Journal of Experimental Psychology: Animal Behavior Processes*, *5*, 31–42.
- Hall, G., & Rodriguez, G. (2010). Attentional learning. In C. Mitchell & M. E. LePelley (Eds.), *Attention and learning* (pp. 41–70). Oxford: Oxford University Press.
- Hall, G., & Rodriguez, G. (2011). Blocking of potentiation of latent inhibition. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*, 127–131.
- Haselgrove, M., Aydin, A., & Pearce, J. (2004). A partial reinforcement effect despite equal rates of reinforcement during Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*, 240–250.
- Holland, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing and blocking cues. *Quarterly Journal of Experimental Psychology*, *52B*, 307–333.
- Holmes, N. M., & Harris, J. A. (2009). The loss of latent inhibition across compound conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 328–339.
- Kamin, L. J. (1968). “Attention-like” processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium on the prediction of behavior: Aversive stimulation* (pp. 9–31). Miami: University of Miami Press.
- Kaufman, M. A., & Bolles, R. C. (1981). A nonassociative aspect of overshadowing. *Bulletin of the Psychonomic Society*, *18*, 318–320.
- Kaye, H., & Pearce, J. M. (1984). The strength of the orienting response during Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*, 90–109.
- Killcross, A. S., Dickinson, A., & Robbins, T. W. (1995). The on-baseline latent inhibition effect is not counterconditioning. *Psychopharmacology*, *118*, 104–106.
- Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.
- Kruschke, J. K., & Blair, N. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, *7*, 636–645.

- Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology*, *57B*, 193–243.
- Le Pelley, M. E., Beesley, T., & Griffiths, O. (2011). Overt attention and predictiveness in human contingency learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*, 220–229.
- Le Pelley, M. E., Beesley, T., & Suret, M. B. (2007). Blocking of causal learning involves learned changes in stimulus processing. *Quarterly Journal of Experimental Psychology*, *60*, 1468–1476.
- 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.
- Lin, T.-C. E., & Honey, R. C. (2011). Encoding specific associative memory: Evidence from behavioral and neural manipulations. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*, 317–329.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.
- 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.
- Matzel, L. D., Schachtman, T. R., & Miller, R. R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning and Motivation*, *16*, 398–412.
- Mercier, P., & Baker, A. G. (1985). Latent inhibition, habituation, and sensory preconditioning: A test of priming in short-term memory. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*, 485–501.
- Miller, R. R., Hallam, S. C., & Grahame, N. J. (1990). Inflation of comparator stimuli following CS training. *Animal Learning & Behavior*, *18*, 434–443.
- Miller, R. R., & Matute, H. (1996). Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgment. *Journal of Experimental Psychology: General*, *125*, 370–386.
- Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51–92). San Diego: Academic Press.
- Napier, R. M., Macrae, M., & Kehoe, E. J. (1992). Rapid reacquisition in conditioning of the rabbit's nictitating membrane response. *Journal of Experimental Psychology: Animal Behavior Processes*, *18*, 182–192.
- Navarro, J. I., Hallam, S. C., Matzel, L. D., & Miller, R. R. (1989). Superconditioning and overshadowing. *Learning and Motivation*, *20*, 130–152.
- Pavlov, I. P. (1927). *Conditioned reflexes*. London: Clarendon Press.
- Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, *101*, 587–607.
- 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., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. In C. Mitchell & M. E. LePelley (Eds.), *Attention and learning* (pp. 11–39). Oxford: Oxford University Press.
- Pearce, J. M., Redhead, E. S., & Aydin, A. (1997). Partial reinforcement in appetitive Pavlovian conditioning with rats. *Quarterly Journal of Experimental Psychology*, *50B*, 273–294.
- Reed, P., Anderson, A., & Foster, C. (1999). Modality of stimulus effects in compound preexposure procedures: Associative influences in enhanced latent inhibition. *Learning and Motivation*, *30*, 35–52.
- Rescorla, R. A. (1970). Reduction in the effectiveness of reinforcement after prior excitatory conditioning. *Learning and Motivation*, *1*, 372–381.
- Rescorla, R. A., & Heth, C. D. (1975). Reinstatement of fear to an extinguished conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, *1*, 88–96.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton.
- Robbins, S. J. (1990). Mechanisms underlying spontaneous recovery in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*, 235–249.
- Rodriguez, G., & Hall, G. (2008). Potentiation of latent inhibition. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*, 352–360.
- Schmajuk, N., & Kutlu, M. G. (2011). Latent inhibition and compound conditioning: A reply to Holmes and Harris (2009). *Journal of Experimental Psychology: Animal Behavior Processes*, *37*, 254–260.
- Schmajuk, N. A., Lam, Y. W., & Gray, J. A. (1996). Latent inhibition: A neural network approach. *Journal of Experimental Psychology: Animal Behavior Processes*, *22*, 321–349.
- Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *Quarterly Journal of Experimental Psychology*, *37B*, 1–21.
- Sokolov, E. N. (1963). *Perception and the conditioned reflex*. New York: Mc Millan.
- Stout, S., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review*, *114*, 759–783.
- Urcelay, G. P., Witnauer, J. E., & Miller, R. R. (2012). The dual role of the context in postpeak performance decrements resulting from extended training. *Learning & Behavior*, *40*, 476–493.
- Urushihara, K., Wheeler, D. S., Pineño, O., & Miller, R. R. (2005). An extended comparator hypothesis account of superconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*, 184–198.
- Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*, 127–151.
- Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale: Erlbaum.
- Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, *76*, 171–180.