The role of within-compound associations in learning about absent cues

James E. Witnauer · Ralph R. Miller

Abstract When two cues are reinforced together (in compound), most associative models assume that animals learn an associative network that includes direct cue–outcome associations and a within-compound association. All models of associative learning subscribe to the importance of cue–outcome associations, but most models assume that within-compound associations are irrelevant to each cue's subsequent behavioral control. In the present article, we present an extension of Van Hamme and Wasserman's (Learning and Motivation 25:127–151, 1994) model of retrospective revaluation based on learning about absent cues that are retrieved through within-compound associations. The model was compared with a model lacking retrieval through within-compound associations. Simulations showed that within-compound associations are necessary for the model to explain higher-order retrospective revaluation and the observed greater retrospective revaluation after partial reinforcement than after continuous reinforcement alone. These simulations suggest that the associability of an absent stimulus is determined by the extent to which the stimulus is activated through the within-compound association.

Keywords Computational modeling · Cue competition · Fear conditioning · Retrospective revaluation · Within-compound associations

J. E. Witnauer · R. R. Miller (✉)
Department of Psychology, State University of New York,
Binghamton, NY 13902-6000, USA
e-mail: rmiller@binghamton.edu

Introduction

Retrospective revaluation occurs when the behavioral control of a target cue (X) changes as a result of posttraining associative changes in associated nontarget stimuli. Retrospective revaluation is problematic for traditional accounts of Pavlovian conditioning (e.g., Rescorla & Wagner, 1972) and human contingency learning (Shanks, 1985) that were designed to explain blocking (Kamin, 1968). Blocking refers to Kamin's discovery that, when a target cue (X) is paired with an unconditioned stimulus (US) in the presence of a well-established signal (A) for the US, X's behavioral control is reduced, relative to when A is replaced with an associatively neutral cue. Blocking is readily explained by total error reduction (TER) models. For example, the Rescorla–Wagner model was developed in large part to account for blocking. This model asserts that the discrepancy between the observed outcome and the predicted outcome based on all cues present (i.e., total error) drives learning. Specifically, predictions are based on the algebraic sum of the associative strengths of all cues present on a given trial. The Rescorla–Wagner model explains the blocking effect by assuming that the presence of A during X–US trials increases expectation of the US, which attenuates total error on AX–US trials, consequently decreasing learning about X. Interestingly, this model assumes that subjects learn X–US, A–US, X–A, and A–X associations but that only direct conditioned stimulus (CS)–US associations are relevant in Pavlovian paradigms. So-called within-compound associations are ordinarily irrelevant. This view was challenged by retrospective revaluation. Most conventional accounts of retrospective revaluation assume that the

associations between a target cue and a nontarget cue (within-compound associations) contribute to retrospective revaluation. In the present article, we discuss the potential role of within-compound associations in learning of direct CS–US associations in retrospective revaluation situations.

The observation that behavioral control by X can change as a result of posttraining associative changes in associated nontarget stimuli prompted researchers to consider the role of within-compound associations in Pavlovian conditioning, as well as human contingency and causality learning situations. Retrospective revaluation was first reported by Kaufman and Bolles (1981). In their critical experiment, rats received compound training consisting of AX–US trials, which was shown to attenuate responding to the target cue X, relative to elemental X–US trials (i.e., overshadowing occurred). Subsequently, subjects received either nonreinforced presentations of A (i.e., extinction; associative deflation) or presentations of an irrelevant cue. When testing X, Kaufman and Bolles observed stronger responding by subjects that received extinction of A than by control subjects. That is, they observed a recovery from overshadowing. The most important aspect of this phenomenon is that behavioral control by X changed despite X not being presented in subsequent treatment of the companion cue. Backward blocking is related to recovery from overshadowing because it constitutes changes in behavioral control by an absent target stimulus as a result of reinforcement of an associate of the target (Shanks, 1985). In a human contingency learning paradigm, Shanks's subjects received compound training consisting of AX–outcome trials with X and A being of similar salience, which prevented appreciable overshadowing of X. Subsequently, subjects received either A–outcome trials or control trials in which A was nonreinforced. Behavioral control by X was attenuated when A–outcome trials were administered, relative to the behavioral control exhibited by subjects that received A-alone trials following AX–outcome trials. Importantly, backward blocking is procedurally similar to Kamin's, (1968) forward-blocking effect, except that the order of elemental and compound trials is reversed in backward blocking. Despite the procedural similarity between backward and forward blocking, backward blocking and, more generally, retrospective revaluation , challenges the TER view of learning as originally construed (e.g., Rescorla & Wagner, 1972) because changes in the associative status of A should not affect learning about X if they occur subsequent to training with X.

In response to the observation of retrospective revaluation, several theories of associative learning have been proposed. One of the most influential accounts of retrospective revaluation based on the TER approach of Rescorla and Wagner (1972) is the view that the associability of an absent target stimulus during revaluation trials is proportional to the additive inverse of the salience of the target stimulus when it is present (Van Hamme & Wasserman, 1994; Wasserman & Berglan, 1998; Wasserman & Castro, 2005). This model, which otherwise recapitulates the Rescorla–Wagner model, explains recovery from overshadowing by assuming that, during nonreinforced presentations of A, subjects expect the outcome on the basis of A. The omission of the outcome should result in a negative total error term (i.e., an overexpected outcome), based on the magnitude of US expectation exceeding the magnitude of the US administered. Those stimuli that are present during A-alone trials (e.g., A and the context of revaluation) will undergo decreases in associative strength (i.e., associative extinction). In contrast, associative changes with respect to X (the absent cue) are expected to be in the opposite direction of A's change in associative strength, due to X's negative associability when it is absent and A is present. Thus, X's association with the US should increase in strength during A-alone trials. The same principle applies to backward blocking, except that associative changes to X are in the direction opposite to recovery from overshadowing because the US is presented on the retrospective revaluation trial. Pairing A with the US should result in a positive total error term, because A should not fully signal the US after compound training in which X and A presumably competed for associative strength with the US. Thus, the magnitude of US expectation should be weaker than the magnitude of the US actually delivered to subjects on the A–US trials. In this situation, the strength of the associations between the present cue (i.e., A) and the US should increase, and the strength of the associations between the absent stimulus (i.e., X) and the US should decrease. By assuming that absent stimuli undergo associative changes opposite to that of present stimuli, Van Hamme and Wasserman's model explains both backward blocking and recovery from overshadowing.

Critically, the Van Hamme and Wasserman (1994) model assumes that the representations of absent stimuli are activated by presented stimuli through within-compound associations (Wasserman & Berglan, 1998). For example, in recovery from overshadowing (i.e., Phase 1, AX–US; Phase 2, A−), activation of the absent target cue's representation occurs because presentation of the overshadowing cue during A-alone trials activates it through associative retrieval. Mathematically, this is accomplished by assuming that the associability of an absent but associatively activated (expected) stimulus is opposite to the associability of the same stimulus, were it present. Van Hamme and Wasserman did not assume that the strength of within-compound associations is important for retrospective revaluation, short of asserting that they are necessary (i.e., they must have some nonzero value) for the effect to occur.

That is, the specific strength of such within-compound associations is not assumed to influence the associability of absent stimuli.

Since the Van Hamme and Wasserman (1994) model of associative learning was developed, considerable research has been amassed concerning the determinants of retrospective revaluation. Some research (e.g., DeHouwer & Beckers, 2001) appears inconsistent with Van Hamme and Wasserman's approach to within-compound associations. For instance, DeHouwer and Beckers observed that, after AB–outcome, BC–outcome, CD–outcome training, associative inflation of Stimulus A caused decrements in behavioral control by Stimuli B and D (consistent with Van Hamme & Wasserman, 1994) and increments in behavioral control by Stimulus C (inconsistent with Van Hamme & Wasserman, 1994), relative to a condition that received deflation of Stimulus A. Moreover, recent studies in Pavlovian fear conditioning suggest that posttraining context deflation is more effective after treatments that strengthen the context–target-stimulus within-compound association (e.g., Laborda, Witnauer, & Miller, in press; Miguez, Witnauer, & Miller, 2010). Simulations of some of these studies indicate that the strength of within-compound associations contributes to the associability of absent stimuli in retrospective revaluation. The purposes of the present simulations are threefold: (1) to determine whether some phenomena are problematic for the original Van Hamme and Wasserman model, (2) to develop an alternative but similarly conceived model of retrospective revaluation (the within-compound model) that better captures the view that within-compound associations drive the associability of absent stimuli, and (3) to determine whether the within-compound model is able to explain the phenomena that are problematic for Van Hamme and Wasserman's model.

Models and simulation methods

Van Hamme and Wasserman (1994) asserted that the associability of a stimulus that is absent but expected on the basis of within-compound associations is proportional to the salience of the stimulus when it is present but of negative value. In their model, the strength of within-compound associations is not considered, except to assume that some degree of retrieval through within-compound associations is needed for activation of absent cues. In other words, regardless of whether the within-compound association between absent and present cues is at full strength or degraded, retrieval of the absent cue should occur, and updating its associative strength should proceed solely on the basis of the CS–US association, CS salience, and a free parameter relating the associability of an absent cue to that of a present cue. Thus, when

simulating their model, we assumed that, when a stimulus (Stim 1) was present,

$$\alpha_{\text{Stim1}} = \text{Salience}_{\text{Stim1}} \tag{1}$$

where Stim 1 is an antecedent stimulus (i.e., a cue). When Stim 1 was absent and the sum of the expectation of Stim 1 was not equal to zero,

$$\alpha_{\text{Stim1}} = k2 * \text{Salience}_{\text{Stim1}}, \tag{2}$$

where $k2$ is a free parameter ($-1 < k2 < 0$) that computes the associability of absent but expected stimuli.

In the present simulations, we slightly altered the Van Hamme and Wasserman (1994) model to create the within-compound model, which aims to better capture the psychological intuition that absent stimuli are retrieved through within-compound associations. In other words, the associability of an absent cue is a function of the extent to which the absent cue is retrieved by the present cue, which is a function of the sum of within-compound associations between the present cues and the absent target cue. Thus, in the present simulations of the within-compound model, we replaced Equation 2, when Stim1 is absent, with

$$\alpha_{\text{Stim1}} = k2 * \text{Salience}_{\text{Stim1}} * \Sigma V_{i-\text{Stim1}}, \tag{3}$$

where $k2$ is a free parameter ($-1 < k2 < 0$) that represents the diminished (and negative) activation of a retrieved but absent stimulus. $\Sigma V_{i\text{-Stim1}}$ reflects the activation of the representation of the absent stimulus that is induced by present cues through within-compound associations. In all other respects, the within-compound and the Van Hamme and Wasserman models are identical and are based on the Rescorla and Wagner (1972) model.

Associative changes in both the Van Hamme and Wasserman (1994) and within-compound models were simulated as follows. First, changes in the strength of all associations (ΔV) were modeled using the following equation for each stimulus dyad on each trial:

$$\Delta V_{\text{Stim1}-\text{Stim2}} = \alpha_{\text{Stim1}} * \beta_{\text{Stim2}} * (\lambda_{\text{Stim2}} - \Sigma V_{i-\text{Stim2}}), \tag{4}$$

where Stim1 and Stim2 represent two different stimuli, α_{Stim1} represents the associability of Stim1, β_{Stim2} represents the associability of Stim 2, and ($\lambda_{\text{Stim2}} - \Sigma V_{i\text{-Stim2}}$) represents the discrepancy between the magnitude of Stim2 that occurred (λ_{Stim2}) and the magnitude of Stim2 that was expected on the basis of all the cues present other than Stim2 itself ($\Sigma V_{i\text{-Stim2}}$). To reduce the number of free parameters, all λs were assumed to be fixed at 1. This assumption might be unrealistic because the maximum amount of associative strength supportable by a stimulus is likely to depend on factors such as stimulus salience. However, alternative assumptions would involve increased model complexity and more parameters. Moreover, this assumption presumably

hurt the within-compound model's fit more than the Van Hamme and Wasserman model's fit because the within-compound model accesses the strength of within-compound associations in computing the associability of absent stimuli. Thus, the values of λ for all cues and the US can potentially influence the models predictions. In contrast, the Van Hamme and Wasserman model assumes that only direct CS–US associations are important for learning. Consequently, Van Hamme and Wasserman's model assumes that only the strength of λ_{US} is directly important for learning. In the present simulations, when Stim2, the outcome or subsequent stimulus (e.g., the US), was present,

$$\beta_{Stim2} = Salience_{Stim2} \qquad (5)$$

and when Stim2 was absent:

$$\beta_{Stim2} = k1*Salience_{Stim2} \qquad (6)$$

where $k1$ is a free parameter $(0 < k1 < 1)$ that represents the diminished associability of absent stimuli, relative to the presence of the same stimuli. This parameter allows the model to anticipate several important phenomena, including the relative stimulus validity effect and the observation that extinction ordinarily requires more trials than does acquisition.

Finally, both models assumed a unitary mapping of associative strength onto response potential (i.e., $R_{X–US} = V_{X–US}$). This is obviously an oversimplification and is contradicted by a vast literature concerning performance factors in Pavlovian situations (see, e.g., Stout & Miller, 2007). However, more realistic assumptions might obscure the critical similarities and differences between the within-compound and Van Hamme and Wasserman models that are highlighted by the present simulations.

Scaling In the present simulations, we were interested in differences between models' predictions and empirical data. Consequently, we needed to scale the predictions of the models to match the scales used in empirical research. The critical measurement in some of the present simulations was the $\log_{10}$ latency to consume water for five cumulative seconds in the presence of the test CS. Following Larrauri and Schmajuk (2008), we assumed that lick suppression multiplied by a scaling factor is equal to the response potential of X (R_X) as predicted by the models. In addition, we assumed that the lick suppression predicted by such proportional scaling is added to the scale's minimum (i.e., $\log_{10} 5$) set by our measuring latency to complete 5 s of drinking in the presence of the CS. Thus, in the present simulations, we approximated the lick suppression (LS) scale, using the equation

$$LS = \log_{10}5 + (Scaling*R_X) + (Scaling*R_{Context}), \qquad (7)$$

where Scaling is a free parameter used to approximate the $\log_{10}$ latency scale measured in lick suppression situations.

To approximate the Kamin (1968) suppression ratio (SR) scale used as the critical measure in some of the present simulations, the outputs of the models used in the present simulations response potential were scaled using the following equation:

$$SR = (Scaling - R_X)$$
$$\div [(Scaling - R_X) + (Scaling - R_{Context})], \qquad (8)$$

where Scaling is a free parameter that was used to represent baseline levels of leverpressing (i.e., in the absence of any behaviorally relevant stimuli). Following Larrauri and Schmajuk (2008), in the present simulations that used either scaling technique, testing was conducted in a context that was not differentially excitatory, so $R_{Context}$ was assumed to be 0 in all of the lick suppression and leverpress suspression ratio simulations. In causal learning studies with human participants, causal ratings were assumed to be proportional to the cue's behavioral control, such that

$$Rating = Scaling*R_X. \qquad (9)$$

Optimization of parameters The predictions of most models are highly dependent on the particular values of free parameters. In our simulations, we sought to compare the fits of models to the data from experiments and to be reasonably confident that differences in the models' fits were driven by differences in the psychological mechanisms instantiated by those models, rather than by differences between models in terms of their parameters. That is, we wanted to control the extent to which models were matched in terms of the quality of their parameters. Thus, we used a hill-climbing algorithm to optimize the parameters of each model that was simulated for the data being simulated. The hill-climbing algorithm used was one of the built-in functions in MATLAB's optimization toolbox. For any given randomly selected start in parameter space, this algorithm simulates the model, using parameter combinations corresponding to the immediately adjacent parameters, and, once the search space is exhausted, moves the search space in the direction that causes the greatest reduction in error. In the present simulations, the sum of the squared error (*SSE*) between the predicted and observed group means was the measurement of error. Of note, we elected not to include any significance tests on the differences between the fits of the models to the data, because our conclusions are centrally focused on differences between the ordinal predictions of the models.

Simulation 1: First-order retrospective revaluation. Simple recovery from overshadowing

First-order retrospective revaluation is the simplest form of retrospective revaluation and includes recovery from

overshadowing. Overshadowing, arguably the simplest form of cue competition, is explained by TER models of cue competition, (e.g., Rescorla & Wagner, 1972) by assuming that A and X compete for a limited amount of associative strength determined by the US when they are trained in compound. The TER view was seriously challenged by the observation that posttraining associative deflation (through extinction) of the overshadowing cue results in stronger responding to the target stimulus than when an irrelevant control cue is extinguished (e.g., Kaufman & Bolles, 1981; Matzel, Schachtman, & Miller, 1985). According to traditional TER models, posttraining associative deflation of the overshadowing cue should not affect the target cue's behavioral control, because this should not change on trials on which the target cue is absent. Recovery from overshadowing was the first and is the most readily observed form of retrospective revaluation in the associative literature.

Toward establishing the basic recovery from overshadowing effect, Matzel et al. (1985) conducted the experiments summarized in Table 1 in a fear-conditioning preparation with rats as subjects. In their Experiment 1, subjects were assigned to one of three groups. Group Elemental received simple elemental X–US pairings in Phase 1, followed by exposure to the training context alone in Phase 2. Group Compound was treated similarly, except that, instead of receiving elemental training, subjects received compound training in which Stimulus A was expected to overshadow Stimulus X, thereby reducing behavioral control by X. Group Extinction received training identical to that of Group Compound, except that instead of receiving context exposure in Phase 2, subjects received 72 A-alone extinction trials. In Experiment 2, the same critical groups were replicated, except that extinction groups received either 4 or 72 nonreinforced presentations of A in Phase 2. Matzel et al. observed the basic overshadowing deficit, which was indexed by reduced suppression in the compound groups, relative to the elemental groups, in both experiments. Critically, suppression was stronger in the extinction

condition, relative to the compound condition, after 72, but not 4, extinction trials (Experiment 2). This constitutes a conceptual replication of recovery from overshadowing driven by extinction of the overshadowing cue (i.e., Kaufman & Bolles, 1981).

The purpose of Simulation 1 was to apply both the within-compound and the Van Hamme and Wasserman (1994) models of retrospective revaluation to the results of Matzel et al. (1985) toward determining how well the two models fit data indicative of basic first-order retrospective revaluation. Ultimately, it would be pointless to show that our revised retrospective revaluation model (i.e., the within-compound model) fits new data without first showing that the revised model fits benchmark phenomena.

Results and discussion

The results of Simulation 1, accompanied by the results of Matzel et al. (1985), are depicted in Fig. 1. The hill-climbing procedure was used to fit the models' predictions to A and X data. Only predictions and observations related to X are reported. Critically, both the within-compound and the Van Hamme and Wasserman (1994) models provided a reasonably good fit to the results of Matzel et al.'s data. Both models anticipated the response reduction observed in the compound condition, relative to the elemental condition (i.e., the overshadowing effect). This resulted from their use of the Rescorla and Wagner (1972) learning rule, which asserts that A and X compete for limited associative strength supported by the US. Concerning the effect of posttraining associative deflation of A, both models explained the response increase to X, which was observed after 72, but not 4, A-alone trials. Recovery from overshadowing is explained by Van Hamme and Wasserman's model by assuming that presentation of A during Phase 2 results in a negative value for the associability of X that is equal to the product of $k2$ and the salience of X. Thus, increments in the strength of the X–US association are
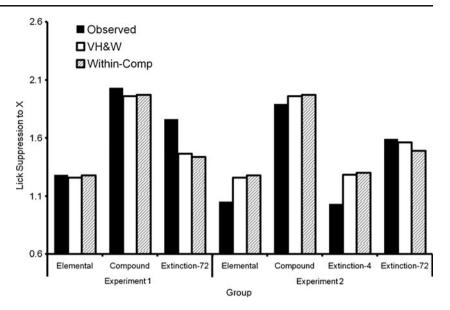
Table 1 Design summary of Matzel, Schachtman, and Miller's (1985) Experiments 1 and 2

Experiment	Group	Phase 1	Phase 2	Test X	Test A
1	Elemental	X+	–	2.03	1.70
1	Compound	AX+	–	1.28	2.40
1	Extinction	AX+	72 A-alone	1.76	1.75
2	Elemental	X+	–	1.89	1.20
2	Compound	AX+	–	1.05	2.10
2	Extinction-4	AX+	4 A-alone	1.03	2.05
2	Extinction-72	AX+	72 A-alone	1.59	1.43

X was a flashing light stimulus. A was a loud tone. + indicates a footshock US. Test values represent observed mean log latencies (in seconds) to complete 5 cumulative seconds of drinking in the presence of the test stimulus

Fig. 1 Predicted and observed conditioned lick suppression to X in Matzel et al. (1985) Experiments 1 and 2. Black bars represent group mean suppression. White bars represent the best fitting predictions of the Van Hamme and Wasserman (1994) model. Striped bars represent the best fitting predictions of the within-compound model. See the text and Table 1 for details

anticipated during A-alone trials because associative changes are assumed to be equal to the product of the CS's negative associability, the US's associability, and a total error term (negatively valued on the basis of non-reinforced presentations of an excitor). The account offered by the within-compound model is similar, except that, instead of assuming that salience and $k2$ alone drive the associability of X, it assumes that retrieval of the representation of X through the within-compound A–X association also influences the associability of X. The *SSE* between observed and predicted responding to X was slightly greater for the within-compound model (*SSE* = 0.25) than for the Van Hamme and Wasserman model (*SSE* = 0.21). The Van Hamme and Wasserman model was relatively well matched to the within-compound model with respect to the proportion of the variance in observed data that could be explained by the models' predictions (Van Hamme and Wasserman's r^2 = .80 and the within-compound model's r^2 = .76). Despite the slight statistical advantage for the fit of the Van Hamme and Wasserman model to the data, critically, both models explained all of the statistically significant differences that were observed by Matzel et al. The results of these simulations confirmed our expectations that both the Van Hamme and Wasserman and the within-compound models would explain first-order retrospective revaluation. The two models required the same number of free parameters to fit the data (see the Table 5).

Simulation 2: Higher-order retrospective revaluation

The most frequently cited problem with the Van Hamme and Wasserman (1994) model (and similar models; e.g.,

Dickinson & Burke, 1996) is its inability to account for higher-order retrospective revaluation, which occurs when changes in the associative status of a higher-order (i.e., beyond first-order) associate of the target stimulus subsequent to target training result in changes in behavioral control by the target stimulus. In first-order retrospective revaluation (e.g., recovery from overshadowing), the associatively inflated or deflated cue is a first-order associate of the target stimulus. Critically, in this situation, the overshadowing cue is directly associated with the target cue through an overshadowing cue–target-cue within-compound association. However, retrospective revaluation does not seem to be restricted to situations that involve a direct association between the directly presented stimulus and the absent stimulus (but see below).

DeHouwer and Beckers (2001) provided the first demonstration of so-called higher-order retrospective revaluation in a human contingency learning experiment. In their Experiment 1 (see Table 2), all the participants received training consisting of AB–outcome/BC–outcome, followed by either associative inflation (through reinforcement) or deflation (through extinction) of Stimulus A. This study was conducted using an allergy task in which cues were meals consumed by a fictitious patient and the outcome was the patient's experiencing an allergic reaction to the meal. Participants were asked to rate the allergenic potentials of A, B, and C (in addition to nontarget filler cues E and F). Not surprisingly, ratings of A were greater in the inflate condition than in the deflate condition, indicating that reinforcement increased ratings to A, relative to extinction. Importantly, ratings to B, which was a first-order associate of A, were attenuated in the inflate condition, relative to the deflate condition.

Table 2 Design Summary of DeHouwer and Beckers's (2001) Experiments 1 and 2

Experiment	Group	Phase 1	Phase 2	A	B	C	D
1	Deflate	AB+/BC+/E+/F−	A−/E+/F−	10	66	26	–
1	Inflate	AB+/BC+/E+/F−	A+/E+/F−	83	35	57	–
2	Deflate	AB+/BC+/CD+/E+/F−	A−/E+/F−	15	61	23	48
2	Inflate	AB+/BC+/CD+/E+/F−	A+/E+/F−	58	14	50	21

Scores under columns A, B, C, and D indicate observed mean ratings of the food cues A through D. + represents the occurrence of an allergic reaction in a fictitious patient. − represents omission of the allergic reaction. Slashes denote interspersed trials

This constitutes a replication of basic first-order retrospective revaluation. However, it is unclear how much of this difference is attributable to recovery from overshadowing (driven by deflation of A), relative to backward blocking (driven by inflation of A). Critically, in DeHouwer and Beckers's experiments, the difference between the inflation and deflation groups was not restricted to the first-order associate of Stimulus A (i.e., Stimulus B). Ratings to C, which was presumably only indirectly associated with Stimulus A through an A–B–C associative linkage, were *elevated* in the inflate condition, relative to the deflate condition. This is a pattern of results opposite to that observed with respect to ratings of Stimulus B. In DeHouwer and Beckers's Experiment 2, participants were treated similarly to those in Experiment 1, except that, during Phase 1, they received CD–outcome trials in addition to AB–outcome and BC–outcome trials. Stimulus D was also affected by posttraining manipulations of Stimulus A; ratings of D were attenuated in the inflate condition, relative to the deflate condition, a pattern similar to that for B and opposite to that for C.

These findings are problematic for the Van Hamme and Wasserman (1994) model because associative inflation or deflation of A seemed to affect stimuli that are only indirectly associated with the target stimuli (i.e., C and D). That is, the A–C and A–D relationships are presumably supported by the A–B–C and A–B–C–D associative structures, respectively. This is problematic for the Van Hamme and Wasserman model because it assumes that an absent target stimulus is eligible for associative changes (i.e., associability is nonzero) in any situation that involves activation of the absent stimulus through a direct within-compound association. Moreover, the model assumes that the degree of activation is irrelevant to the associability of the absent stimulus, except to assume that some nonzero level of activation is needed. The purpose of Simulation 2 was to establish whether or not the Van Hamme and Wasserman or the within-compound model could explain the results of DeHouwer and Beckers concerning first-, second-, and third-order retrospective revaluation.
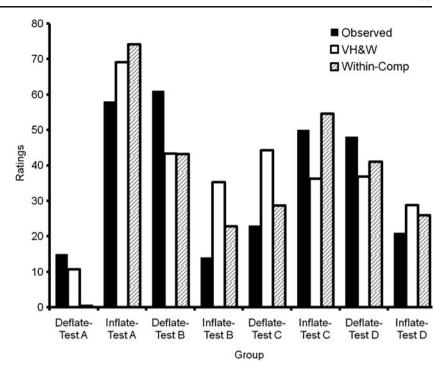
Results and discussion

The results of Simulation 2 are depicted in Fig. 2. The predictions for both models were derived from a hill-climbing exploration of parameter space that minimized differences between observed and predicted group means in both Experiments 1 and 2 of DeHouwer and Beckers (2001), including the filler cues. We included both experiments in the hill-climbing exploration because this would presumably improve the generality of the parameters that were derived. In the interest of simplifying presentation of the data and highlighting critical predictions, *SSE*, r^2, and Fig. 2 are based only on predictions and observations of target stimuli (A, B, C, and D) in DeHouwer and Beckers's Experiment 2. Experiment 2 replicated Experiment 1, with the addition of third-order retrospective revaluation. Our decision to report only the predictions concerning Experiment 2 did not influence our conclusions. Simulation 2 showed that the within-compound model (*SSE* = 995 and r^2 = .72) provided a better fit to DeHouwer and Beckers's results than did the Van Hamme and Wasserman model (*SSE* = 1733 and r^2 = .42). This difference is driven mostly by differences between the predictions of the models concerning ratings to Stimulus C.

On the basis of both models' using the Rescorla–Wagner learning rule, some important predictions are made by each model. With respect to ratings of A, both the Van Hamme and Wasserman (1994) and within-compound models explained the basic difference between the inflate and deflate conditions, indicating, not surprisingly, that both models explain the difference between acquisition and extinction. Both models predicted that, following training, the A–B and A–D associations should be excitatory and the A–C association should be inhibitory. According to both models, AB–outcome and BC–outcome trials should establish excitatory A–B, B–A, B–C, and C–B within-compound associations. On AB–outcome trials, B evokes an expectation of C that exceeds the magnitude of C experienced by participants (zero). Thus, an inhibitory A–C

Fig. 2 Predicted and observed ratings to A, B, C, and D in DeHouwer and Beckers's (2001) Experiment 2. Black bars represent group mean suppression. White bars represent the best fitting predictions of the Van Hamme and Wasserman (1994) model. Striped bars represent the best fitting predictions of the within-compound model. See the text and Table 2 for details

association is expected to develop. AB–outcome/BC–outcome training is similar to the procedures in experiments that have produced evidence of inhibitory within-compound learning (e.g., Espinet, Iraola, Bennet, & Mackintosh, 1995); thus, this account is consistent with the empirical literature concerning within-compound associations. A revision of the McLaren, Kaye, and Mackintosh (1989) model proposed by Graham (1999) can explain inhibitory within-compound learning, but it appears that the model cannot account for the higher-order retrospective revaluation phenomena that are explained by the within-compound model. According to both of the models, the B–D association is inhibitory for the same reason that the A–C association is inhibitory (i.e., the expectation of D on BC trials exceeds the magnitude of D experienced [zero], which drives learning of an inhibitory B–D association). On AB–outcome trials, B evokes an inhibitory expectation of D, which results in under-expectation of D (i.e., magnitude of the actual D stimulus [zero] exceeded that of the expected magnitude), thereby driving increments in the A–D association. The predicted strengths of the within-compound associations fall straightforwardly from the Rescorla–Wagner model as applied to within-compound associations. However, the models differ with respect to how they use the within-compound associations. This is critical in determining the models' predictions concerning retrospective revaluation.

Concerning ratings of B, both models accurately predicted lower ratings to B in the inflate than in the deflate condition. This indicates that both models explain

first-order retrospective revaluation, thereby replicating the central result of Simulation 1. Critically, the predictions of the within-compound model diverge from those of the Van Hamme and Wasserman (1994) model with respect to ratings of Stimulus C. Consistent with the analysis of Van Hamme and Wasserman's model provided by most researchers (e.g., DeHouwer & Beckers, 2001; McConnell & Miller, 2010; McConnell, Wheeler, Urcelay, & Miller, 2009; McLaren, 2009; Melchers, Lachnit, & Shanks, 2004; Witnauer & Miller, 2010), it failed to predict higher ratings of Stimulus C in the inflate than in the deflate condition. That is, the model fails to explain second-order retrospective revaluation. In fact, on the basis of there being nonzero A–B, A–C, and A–D associative strengths (see above), the Van Hamme and Wasserman model predicts the same pattern of behavior across B, C, and D. That is, the Van Hamme and Wasserman model predicts that inflation of A should attenuate ratings of B, C, and D, relative to a deflation treatment. The Van Hamme and Wasserman model correctly predicted higher ratings to Stimulus D in the deflate than in the inflate condition because the A–D within-compound association should be nonzero during Phase 2, which should cause Stimulus D's associability to be negative (see below for an explanation of how Stimulus D presumably gained association with Stimulus A). Thus, Stimulus D is predicted to be similar to Stimulus B with respect to the direction of changes in associative strength.

Surprisingly, the within-compound model correctly predicted higher ratings to C in the inflate than in the deflate condition, thereby correctly predicting second-
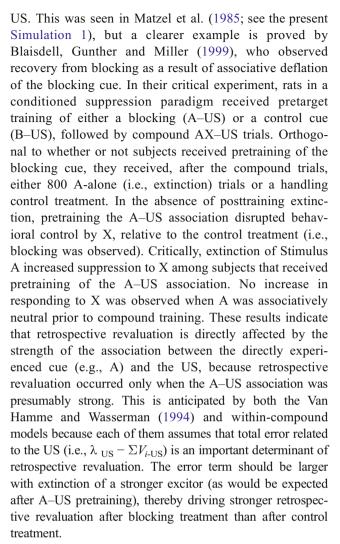
order retrospective revaluation. This was rather unexpected because, like the Van Hamme and Wasserman (1994) model, the within-compound model assumes that only direct within-compound associations to the target cue can contribute to the associability of an absent cue. To better understand the source of this prediction, we consider the potential effects of AB–outcome/BC–outcome/CD–outcome training on direct within-compound associations. Critically, during elemental training of A (i.e., Phase 2), the expectation of C is negative (i.e., inhibitory). The within-compound model assumes that the associability of C is determined by the product of (1) $k2$, (2) the salience of C, and (3) the strength of the A–C association. Because both 1 and 3 are negative and 2 is positive, the model asserts that C's associability is positive during elemental training of A. Thus, the model predicts that changes in the strength of the C–outcome association should be in the same direction as changes in the strength of the A–outcome association (and opposite to that of the B–outcome association), which is consistent with DeHouwer and Beckers's (2001) observations. Concerning ratings of Stimulus D, DeHouwer and Beckers observed higher ratings in the deflate than in the inflate condition. This is predicted because the within-compound model assumes that the A–D association is weakly excitatory following Phase 1. This expectation is also driven by the use of competitive learning of within-compound associations. Thus, the effect of retrospective revaluation on Stimulus D should be similar to the effect on Stimulus B, which is consistent with DeHouwer and Beckers's results.

The results of Simulation 2 suggest that higher-order retrospective revaluation is explained by the within-compound model, and not by the Van Hamme and Wasserman (1994) model. These results also suggest that higher-order retrospective revaluation is better modeled by assuming that the direct (rather than higher-order) relationship between the absent and directly presented cue is a critical determinant of changes in the behavioral control of absent cues. Perhaps the term *higher-order retrospective revaluation* is misleading, because such effects are explained by the within-compound model without reference to higher-order associative structures (but see Stout & Miller, 2007, for an account based on higher-order associative structures).

Simulation 3: Contextual associations in first-order retrospective revaluation

The results of numerous studies indicate that retrospective revaluation is directly affected by the strength of the association between the inflated or deflated cue and the US. This was seen in Matzel et al. (1985; see the present Simulation 1), but a clearer example is proved by Blaisdell, Gunther and Miller (1999), who observed recovery from blocking as a result of associative deflation of the blocking cue. In their critical experiment, rats in a conditioned suppression paradigm received pretarget training of either a blocking (A–US) or a control cue (B–US), followed by compound AX–US trials. Orthogonal to whether or not subjects received pretraining of the blocking cue, they received, after the compound trials, either 800 A-alone (i.e., extinction) trials or a handling control treatment. In the absence of posttraining extinction, pretraining the A–US association disrupted behavioral control by X, relative to the control treatment (i.e., blocking was observed). Critically, extinction of Stimulus A increased suppression to X among subjects that received pretraining of the A–US association. No increase in responding to X was observed when A was associatively neutral prior to compound training. These results indicate that retrospective revaluation is directly affected by the strength of the association between the directly experienced cue (e.g., A) and the US, because retrospective revaluation occurred only when the A–US association was presumably strong. This is anticipated by both the Van Hamme and Wasserman (1994) and within-compound models because each of them assumes that total error related to the US (i.e., $\lambda_{US} - \Sigma V_{i\text{-}US}$) is an important determinant of retrospective revaluation. The error term should be larger with extinction of a stronger excitor (as would be expected after A–US pretraining), thereby driving stronger retrospective revaluation after blocking treatment than after control treatment.

In addition to studies demonstrating that the amount of competing cue–US training is an important determinant of retrospective revaluation, some research suggests that the amount of competing cue–target-cue training also influences retrospective revaluation. For example, Laborda et al. (in press) conducted the experiments summarized in Table 3, which revealed that retrospective revaluation achieved by posttraining context deflation is increased when retrospective revaluation treatment occurs after the target stimulus has been repeatedly presented in the presence of a competing cue (the acquisition context, in this case). Specifically, Laborda et al. were interested in the difference between the effectiveness of extinction in the acquisition context (AAC renewal) and extinction in a neutral context (ABC renewal). A potentially important difference between these two paradigms is that AAC renewal involves many more CS–acquisition-context pairings than does ABC renewal. Hence, the Context A–CS association should be stronger in AAC than in ABC renewal. In their Experiment 1, rats in a lick suppression preparation were used to demonstrate both the basic

Table 3 Design Summary of Laborda et al. (in press) Experiments 1 and 2

Experiment	Group	Phase 1	Phase 2	Phase 3	Test X
1	AAA	X+	X − (Ctx A)	–	1.14 (Ctx A)
1	AAC	X+	X − (Ctx A)	–	1.06 (Ctx C)
1	ABC	X+	X − (Ctx B)	–	1.74 (Ctx C)
1	A-C	X+	–	–	2.07 (Ctx C)
2	AAC-20	X+	X − (Ctx A)	20 min	0.93 (Ctx C)
2	AAC-480	X+	X − (Ctx A)	480 min	1.56 (Ctx C)
2	ABC-20	X+	X − (Ctx B)	20 min	1.71 (Ctx C)
2	ABC-480	X+	X − (Ctx B)	480 min	1.71 (Ctx C)

X was a click train. + indicates a footshock US. Test values represent observed mean log latencies (in seconds) to complete 5 cumulative seconds of drinking in the presence of X. Number of Phase 3 minutes represents the amount of Context A deflation

extinction and renewal phenomena. In Experiment 2, half of the subjects received AAC renewal treatment, in which training and extinction occurred in Context A and testing occurred in Context C. The remaining subjects received ABC renewal treatment, in which training occurred in Context A, extinction in Context B, and testing in Context C. Orthogonally, subjects received either massive post-training associative deflation (480 min) of the acquisition context (Context A) or a handling control treatment (20 min). Critically, deflation of Context A increased responding only in the AAC condition, in which subjects experienced 36 more pairings of the target stimulus with the training context than in the ABC condition. These data, in conjunction with similar studies of latent inhibition (Grahame, Barnet, Gunther, & Miller, 1994), suggest that increasing the strength of the within-compound association between the competing cue (Context A) and the target cue (X) increases retrospective revaluation. The purpose of Simulation 3 was to compare the best-fitting predictions of the within-compound model and the Van Hamme and Wasserman (1994) model with the results of Laborda et al.
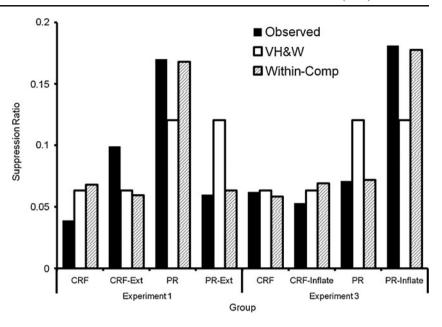
Results and discussion

The results of Simulation 3 are depicted in Fig. 3. Inspection of this figure reveals that both models provided a reasonably good fit to the data. However, despite both models' explaining most of the critical results in these studies, the pattern of behavior in Experiment 2 was better explained by the within-compound than by the Van Hamme and Wasserman (1994) model. Given the best-

fitting parameters, the Van Hamme and Wasserman model failed to predict much of an effect of posttraining acquisition context deflation in Experiment 2, which contrasts with the observation of increased responding in Group AAC-480 relative to Group AAC-20. Moreover, the Van Hamme and Wasserman model predicted a greater difference between Groups ABC-480 and ABC-20 than between Groups AAC-480 and AAC-20, indicating that more retrospective revaluation was expected in the ABC condition than in the AAC condition, which was contradicted by the subjects' behavior. This erroneous pattern is anticipated because, according to the model, context extinction during the Phase 2 intertrial intervals should have attenuated the strength of the Context A–US association in the AAC condition (but not the ABC condition). This should have reduced the surprising nature of the subsequent context extinction treatment, relative to the ABC condition, thereby attenuating retrospective revaluation.

The within-compound model provided a somewhat better (although far from perfect) fit to the results of Laborda et al.'s (in press) Experiment 2 than did Van Hamme and Wasserman's (1994) model. It successfully explained the central observation that retrospective revaluation (i.e., increased responding to X as a result of associative deflation of the acquisition context) was more pronounced in the AAC than in the ABC condition. According to the within-compound model, this occurred because the within-compound association between Context A and the target CS, which is presumably critical for retrospective revaluation, was stronger in the AAC than in the ABC condition. However, the surprise from the context deflation treatment should have been stronger in

the ABC than in the AAC condition, thereby slightly diminishing the size of the difference between AAC and ABC renewal. But, overall, the within-compound model anticipates that the effect of increasing the strength of within-compound association between the target cue and Context A during Phase 2 should be stronger than the effect of context deflation during Phase 2 in the AAC condition. This allowed the model to anticipate the results of Laborda et al. A statistical analysis of Simulation 3 supported our intuitions about the predicted group means. The variance in the observations explained by the within-compound model ($SSE = 0.12$ and $r^2 = .90$) was greater than the variance explained by the Van Hamme and Wasserman model ($SSE = 0.21$ and $r^2 = .81$).

Simulation 4: Contextual associations in first- and second-order retrospective revaluation

The results of Simulation 2 suggest that the within-compound model can explain higher-order retrospective revaluation, whereas the Van Hamme and Wasserman (1994) model fails to account for the effect. The results of Simulation 3 suggest that the target-CS–competing-cue association (in this case, the CS–Context A association) can be a critical determinant of the effect of posttraining associative context deflation on retrospective revaluation.

An important limitation to those simulations is that they were explained using separate sets of parameters. It is possible that the within-compound model's predictions are sufficiently parameter dependent that it would not be able to explain both critical findings with the same set of parameters. However, it would be inappropriate to use the same set of parameters to make predictions concerning human contingency learning and Pavlovian fear conditioning with rats. Thus, we did not apply the hill-climbing procedure to both sets of predictions simultaneously. Instead, the purpose of Simulation 4 was to simulate the results of Miguez et al. (2010), in which both higher-order retrospective revaluation and differential retrospective revaluation as a function of the strength of within-compound associations were demonstrated.

The design of Miguez et al. (2010, Experiments 1 and 3) is summarized in Table 4. The purpose of their experiments was to determine whether the partial-reinforcement (PR) performance deficit is sensitive to retrospective revaluation. The PR performance deficit occurs when nonreinforced presentations of a target CS (X) during intervals between X–US pairings result in less conditioned responding at test to X than when X is continuously reinforced (i.e., the nonreinforced trials are omitted). Procedurally, the PR effect is similar to extinction (i.e., the response decrement caused by nonreinforced trials after X–US pairings) and latent inhibition (i.e., the response decrement is caused by

Table 4 Design Summary of Miguez, Witnauer, and Miller's (2010) Experiments 1 and 3

Experiment	Group	Phase 1	Phase 2	Test X
1	CRF-20	X+	20 min	.04
1	CRF-480	X+	480 min	.10
1	PR-20	X+/X−	20 min	.17
1	PR-480	X+/X−	480 min	.06
3	CRF-Ctrl	X+/A−	B+	.06
3	CRF-Inflate	X+/A−	A+	.05
3	PR-Ctrl	X+/X−/A−	B+	.18
3	PR-Inflate	X+/X−/A−	A+	.07

X was a click train. + indicates a footshock US. Test values represent suppression ratios during X. Lower numbers indicate more suppression. Number of Phase 3 minutes represents the amount of context deflation received. CRF = continuous reinforcement. PR = partial reinforcement

nonreinforced trials before X–US pairings) because it is a response decrement driven by nonreinforced presentations of the CS. Research suggests that latent inhibition (Grahame et al. 1994) and extinction (Witnauer & Miller, 2009) are attenuated by posttraining associative deflation of the acquisition context. Miguez et al. sought to evaluate whether posttraining context deflation would similarly attenuate the PR performance deficit. In their Experiment 1, subjects received either PR or continuous reinforcement (CRF) in Phase 1 and, orthogonally, either 20 or 480 min of Phase 2 context exposure, which was administered to associatively deflate the acquisition context. The effect of associative deflation of the acquisition context was specific to the PR condition; recovery from the PR performance deficit was observed as a consequence of context deflation. The extent to which the PR performance deficit is sensitive to higher-order retrospective revaluation was evaluated in Miguez et al.'s (2010) Experiment 3, the design of which is summarized in Table 4. Subjects received either PR or continuous reinforcement (CRF). Orthogonally, they received inflation of either Stimulus A (a higher-order associate to the target) or an irrelevant control stimulus (B). For all subjects, Stimulus A was a cue that was nonreinforced in the training context. This was intended to establish a training-context–Stimulus A association, which could mediate higher-order retrospective revaluation. Thus, among subjects in the inflate condition, the Stimulus A–US association was second-order with respect to the direct X–US association and could be activated through the X–context–Stimulus A associative linkage. Consequently, the relationship be-

tween X and Stimulus A was comparable to the relationship between Stimulus A and Stimulus C in DeHouwer and Beckers's (2001) inflate condition in their Experiment 1. Consistent with DeHouwer and Beckers's results, Miguez et al. observed increased responding to X as a result of associative inflation of Stimulus A. Moreover, this effect was specific to the condition that received PR in Phase 1. In summary, Miguez et al. replicated higher-order retrospective revaluation (DeHouwer & Beckers, 2001) and the finding that retrospective revaluation is more robust after added X–context nonreinforced trials than with only CRF of X (Laborda et al., in press).We did not simulate Miguez et al.'s Experiment 2, which investigated the effect of context inflation on the PR performance deficit because the experiment used a sensory preconditioning procedure that precludes comparison with first-order conditioning. On the basis of the results of Simulation 2 and 3, we expected that the within-compound model's predictions would provide a superior fit to Miguez et al.'s observations than would those of the Van Hamme and Wasserman (1994) model.
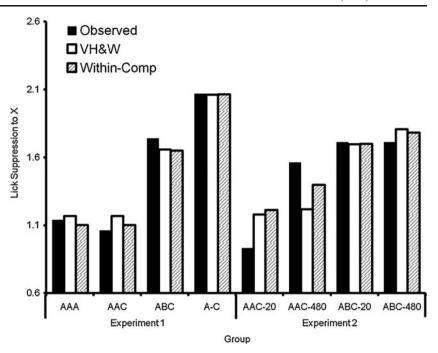
Results and discussion

The results of Simulation 4 are summarized in Fig. 4. Critically, the predictions of the within-compound model were more similar to the observations of Miguez et al. (2010) than were those of the Van Hamme and Wasserman (1994) model. This was supported both by comparisons between the statistical fits of the models and by comparisons between predicted and observed ordinal differences between groups. The within-compound model provided a better statistical fit to the results of Miguez et al. ($SSE = 0.003$ and $r^2 = .87$) than did the Van Hamme and Wasserman model ($SSE = 0.014$ and $r^2 = .32$).

As in Simulation 3, the within-compound model anticipated that 480 min of context deflation should increase responding after PR training, but not after CRF training, because the PR treatment should increase the strength of the training-context–CS association, which should enhance the associability of X during context deflation, thereby increasing retrospective revaluation of X. In contrast, the Van Hamme and Wasserman (1994) model failed to anticipate a difference between the PR and CRF conditions with respect to the amount of recovery that should occur when the context is deflated because the model does not assume that the magnitude of the within-compound association between the context and X influences the associability of the absent Stimulus X. That is, Van Hamme and Wasserman assert that

the associability of X during retrospective revaluation should be equivalent in the PR and CRF conditions, which was contradicted by the observations. Concerning Miguez et al.'s (2010) Experiment 3, the within-compound model explained the response recovery observed for X as a result of inflating Stimulus A. This was anticipated because the within-compound model posits that an inhibitory X–A association should develop as a result of these cues being explicitly unpaired in the same context (i.e., X−/A−) in the PR condition in Phase 1. Then, in Phase 2, the inhibitory expectation of X during A–US trials should cause the associability of X to be positive. Thus, associative changes with respect to X should proceed in the same direction of A's associative changes (i.e., associative increments should occur when A is inflated). These predictions are consistent with the observations of Miguez et al. In contrast, the Van Hamme and Wasserman model failed to predict higher-order retrospective revaluation because it assumed that the associability of X is equivalent in first- and second-order retrospective revaluation.

General discussion

In the present article, we developed the within-compound model, which is identical to the Van Hamme and Wasserman (1994) model with the exception that it asserts

that the value of within-compound associations influences the associability of absent stimuli. The present simulations revealed that the within-compound model explains some critical aspects of retrospective revaluation that elude the original Van Hamme and Wasserman model on which the within-compound model is based. The within-compound model is a logical elaboration of the Van Hamme and Wasserman model and captures the same psychological intuition, that within-compound associations drive retrospective revaluation, which was promoted by Wasserman and his colleagues. Thus, the within-compound model is not an alternative to the Van Hamme and Wasserman model but, rather, a different implementation of the same elegant framework. The within-compound model implemented here explains both higher-order retrospective revaluation (e.g., DeHouwer & Beckers, 2001; Miguez et al., 2010) and more retrospective revaluation after PR or extinction than after CRF (Laborda et al., in press; Miguez et al., 2010).

It is noteworthy that preliminary simulations revealed some limitations of the within-compound model but that the within-compound model always performed at least as well as the Van Hamme and Wasserman (1994) model. First, the model does not explain failures to observe retrospective revaluation. For instance, Holland (1999) observed that, despite extinction of the overshadowing stimulus, associative deflation of the overshadowing cue

does not always cause recovery from overshadowing. Similarly, extinction of the blocking cue following compound training does not always result in recovery from blocking (Blaisdell et al., 1999; Dopson, Pearce, & Haselgrove, 2009; Holland, 1999). Perhaps more problematic than the null results (e.g., Dopson et al., 2009) were studies that revealed less retrospective revaluation after forward blocking (where all elemental trials precede all compound trials) than after interspersed blocking (where the two trial types are interspersed; Blaisdell et al., 1999). In their Experiment 2, Blaisdell et al. administered either phasic (i.e., all A–US before any AX–US trials) or interspersed blocking, followed orthogonally by either 200 nonreinforced presentations of the blocking cue or a control treatment. They observed recovery from blocking only in the condition that received interspersed blocking trials, indicating a strong trial order effect in retrospective revaluation. Neither the within-compound model nor the Van Hamme and Wasserman model is able to explain this difference.

The within-compound model assumes that the associability of an absent stimulus is driven in part by its associability when it is physically presented. This is contradicted by studies of the effect of CS salience on recovery from overshadowing (Liljeholm & Balleine, 2006). In their critical experiment, A was a highly salient cue, and X was a weakly salient cue. After compound AX + training, subjects received posttraining deflation of either A (with subsequent test of X) or X (with subsequent test of A). Liljeholm and Balleine failed to observe an effect of deflation of X but observed reliable recovery from overshadowing after deflation of A. This is incompatible with the assumption that the associability of an absent stimulus is driven by the salience of that stimulus when it is physically presented, which is a central assumption of both the Van Hamme and Wasserman (1994) and within-compound models. However, this observation does not contradict the central claim of the present simulations, which is that within-compound associations contribute to the associability of absent stimuli. We are ultimately not wedded to the present assumption about salience and, in fact, find some reasons to be skeptical of it.

To the authors' knowledge, the only other model that can explain the data reviewed in the present article is the sometimes competing retrieval (SOCR) model (Stout & Miller, 2007). This model assumes that the strength of conditioned performance at test is determined by comparisons between directly activated outcome (e.g., US) representations and outcome representations activated through higher-order associative structures. For example,

in overshadowing, subjects are assumed at test to compare the US representation activated through the X–US association with the US representation indirectly activated through the X–A–US associative linkage. The indirectly activated US representation functions to reduce the magnitude of conditioned excitatory behavior. The model explains recovery from overshadowing by assuming that the deflation treatment reduces the strength of the indirectly activated US representation, thereby increasing the strength of responding. The observation of higher-order retrospective revaluation has often been cited as being uniquely anticipated by SOCR (e.g., Witnauer & Miller, 2009) or at least incompatible with the view that retrospective revaluation is driven by direct within-compound associations (e.g., McLaren, 2009). The present simulations establish an alternative to SOCR that can explain higher-order retrospective revaluation. Although we did not seek to directly compare the within-compound model with SOCR here, indeed, some preliminary comparisons revealed that the within-compound model provided a superior fit to some higher-order retrospective revaluation data than did SOCR. Concerning the effect of posttraining associative inflation of the context, SOCR readily explains the observation that retrospective revaluation is greater after nonreinforced presentations in the acquisition context than after nonreinforced reinforced in a neutral context (Laborda et al., in press). It also explains greater retrospective revaluation after PR than after CRF (Miguez et al., 2010).

The modified sometimes opponent processes (MSOP) model (Dickinson & Burke, 1996) is similar to the within-compound model with respect to assumptions about the role of within-compound associations in determining the associability of absent cues (i.e., A2 activation within MSOP). However, this model is conceptually quite different from the within-compound model because it is based on the sometimes opponent processes model (Wagner, 1981) rather than on the Rescorla and Wagner (1972) model. MSOP assumes that a stimulus is represented by a set of microelements that are activated in one of three different memory states. A1 is a high activation, focal working memory state. A2 is a low activation state. Elements in State I are inactive (i.e., in long-term memory). In the MSOP framework, presenting a stimulus results in a self-generated priming of representational elements from I into A1. Elements in A1 decay into A2, which is of relatively high capacity. A2 elements slowly decay into I. Elements cannot be activated from A2 into A1. MSOP assumes that excitatory associations form between elements conjointly activated in A1. Moreover, excitatory associations between simultaneously activated elements in A2 or in A1 are

learned. According to MSOP, inhibitory associations are learned between elements that are activated in different working memory states. The MSOP model can explain the basic finding that context extinction increases responding more after PR than after CRF training (e.g., Miguez et al., 2010). It asserts that, in the CRF condition, an excitatory X–US association develops as a result of simultaneous activation of X and US elements in A1. At test, presenting X increases the number of US elements activated into A2, which drives excitatory responding. In the PR condition, this effect is attenuated because, during nonreinforced training trials, X elements are activated in A1 and US elements are activated in A2, which encourages inhibitory X–US learning, thereby reducing A2 activation of the US (and excitatory responding) at test. The model predicts stronger retrospective revaluation in the PR than in the CRF condition because PR should establish a stronger excitatory-context–X association, which should increase A2 activation of X during context deflation. Because excitatory learning occurs among simultaneously activated A2 elements, deflation of the context should increase X's behavioral control on the basis of both X and US elements being associatively activated into A2 by the training context. This effect should be enhanced by the nonreinforced context–X pairings of PR treatment, which strengthens the context–X association. Thus, MSOP anticipates Miguez et al.'s Experiment 1 and, using similar mechanisms, Laborda et al.'s (in press) Experiment 2. However, MSOP fails to explain higher-order retrospective revaluation, despite its assumption that within-compound associations drive retrospective revaluation (see Wheeler & Miller, 2008).

According to MSOP, in Miguez et al.'s (2010) Experiment 3, inhibitory X–A associations should be established as a result of explicitly unpaired X– / A − training of Phase 1. In the inflate condition, presentation of CS A during Phase 2 should have little effect on the activation of X elements. That is, the A–X association is expected to be inhibitory, which ordinarily reduces A2 activation of representational elements, which is otherwise driven by retrieval through excitatory associations. This should not allow for net *excitatory* learning with respect to the X–US relationship. Miguez et al. observed an increase in behavioral control by X rather than a decrease, as predicted by MSOP. Thus, MSOP fails to explain higher-order retrospective revaluation as observed by Miguez et al. However, MSOP's use of microelemental stimulus representation and real-time temporal representation precluded formal comparison with the simpler, elemental, trial-wise models simulated in the present research.

The present research is consistent with the observations of Melchers, Lachnit, and Shanks (2004; see also Aitken, Larkin, & Dickinson, 2001; Dickinson & Burke, 1996). In retrospective revaluation in humans, within-compound memory is correlated with the extent to which behavioral control by an absent target stimulus changes during posttraining manipulations of nontarget stimuli. Specifically, subjects received either a within-subjects forward blocking (Phase 1, A+/B-; Phase 2, AX+/BY+) or a within-subjects backward blocking (Phase 1, AX+/BY+; Phase 2, A+/B-). The amount of blocking observed (i.e., ratings to Y minus ratings to X) was correlated with subjects' memory of the within-compound associations. Moreover, this effect occurred only when the target stimulus was retrospectively revaluated (i.e., in backward blocking). This suggests that retrieval of the absent stimulus through within-compound associations is critical for retrospective revaluation. The within-compound model developed in the present report is consistent with these observations.

The present research has two broadly relevant implications. First, these simulations support the view that the specific strength of within-compound associations is important for cue interactions. Recent theoretical work in our laboratory suggests that animals use information accessed through within-compound associations in retrospective revaluation, cue competition, conditioned inhibition, extinction, and counteraction situations (Witnauer & Miller, 2010). The present work furthers this claim. Second, the present simulations are incompatible with the view that higher-order retrospective revaluation is *uniquely* supportive of performance-focused models of associative learning. According to such models, learning is assumed to be driven by relatively simple mechanisms (e.g., spatial-temporal contiguity; Miller & Matzel, 1988). and performance is based on sophisticated response rules (e.g., Stout & Miller, 2007). The present simulations indicate that select acquisition-focused models can also explain higher-order retrospective revaluation. Such models assume that associations encode more than spatial-temporal proximity (contiguity). For example, the Rescorla and Wagner (1972) model assumes that associations depend on both contiguity and total error across a stimulus compound. They also assume that performance is based on a monotonic mapping of associative strength into behavior. The present simulations imply that higher-order retrospective revaluation is driven by within-compound associations, which is assumed by all models that can currently explain the effect. Future research should aim to use other techniques to differentiate between acquisition- and performance-focused accounts of retrospective revaluation (see Wheeler & Miller, 2008).

Appendix

Table 5 Best fitting parameters

Best-fitting parameters were revealed through hill climbing on *SSE*. Sal = Salience. Ctx = Context. Ctx was used to simulate punctate cue C in Simulation 7). k1 = factor for dampening absent beta. K2 = factor for dampening absent alpha. NA = not applicable

Simulation	Sal X (All Cues Sim 2)	Sal A	Sal Ctx	Sal US	$k1$	$k2$	Scaling
Best-fitting parameters for Van Hamme and Wasserman's (1994) model							
1	0.36	1.00	0.00	0.29	0.04	1.00	2.13
2	0.28	NA	NA	0.28	0.70	0.14	78.62
3	0.42	NA	0.51	0.46	0.15	0.11	3.14
4	0.24	0.01	0.08	0.17	0.03	0.67	0.001
Best-fitting parameters for within-compound model							
1	0.33	0.86	0.01	0.33	0.04	1.00	2.17
2	0.37	NA	NA	1.00	0.55	1.00	74.09
3	0.36	NA	0.48	1.00	0.09	1.00	3.16
4	0.01	1.00	0.08	0.17	0.03	0.67	0.004

References

Aitken, M. R. F., Larkin, M. J. W., & Dickinson, A. (2001). Re-examination of the role of within-compound associations in retrospective revaluation of causal judgements. *Quarterly Journal of Experimental Psychology, 54B*, 27–51.

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.

DeHouwer, J., & Beckers, T. (2001). Higher-order retrospective revaluation in human causal learning. *Quarterly Journal of Experimental Psychology, 55B*, 137–151.

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

Dopson, J. C., Pearce, J. M., & Haselgrove, M. (2009). Failure of retrospective revaluation to influence blocking. *Journal of Experimental Psychology: Animal Behavior Processes, 35*, 473–484.

Espinet, A., Iraola, J. A., Bennet, C. H., & Mackintosh, N. J. (1995). Inhibitory-like associations among neutral stimuli in flavor-aversion conditioning. *Animal Learning & Behavior, 23*, 361–368.

Graham, S. (1999). Retrospective revaluation and inhibitory associations: Does perceptual learning modulate our perception f the contingencies between events? *Quarterly Journal of Experimental Psychology, 52B*, 159–185.

Grahame, N., Barnet, R. C., Gunther, L. M., & Miller, R. R. (1994). Latent inhibition as a performance deficit resulting from CS–context associations. *Animal Learning & Behavior, 22*, 395–408.

Holland, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or blocking cues. *Quarterly Journal of Experimental Psychology, 52*, 307–333.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *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.

Laborda, M. A., Witnauer, J. E., & Miller, R. R. (in press). Contrasting AAC and ABC renewal: The role of context associations. *Learning & Behavior*.

Larrauri, J. A., & Schmajuk, N. A. (2008). Attentional, associative, and configural mechanisms in extinction. *Psychological Review, 115*, 640–676.

Liljeholm, M., & Balleine, B. W. (2006). Stimulus salience and retrospective revaluation. *Journal of Experimental Psychology: Animal Behavior Processes, 32*, 481–487.

Matzel, L. D., Schachtman, T. R., & Miller, R. R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowed stimulus. *Learning and Motivation, 19*, 99–121.

McConnell, B., & Miller, R. R. (2010). Protection from extinction provided by a conditioned inhibitor. *Learning & Behavior, 38*, 68–79.

McConnell, B., Wheeler, D., Urcelay, G. P., & Miller, R. R. (2009). Protection from latent inhibition provided by a conditioned inhibitor. *Journal of Experimental Psychology: Animal Behavior Processes, 35*, 498–508.

McLaren, I. P. (2009). *Rules and associations: Dissociating different processes in retrospective revaluation*. Boston, MA: Paper presented at the 50th Annual Meeting of the Psychonomic Society.

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

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

Miguez, G., Witnauer, J. E., & Miller, R. R. (2010). [The role of contextual associations in producing the partial reinforcement acquisition deficit]. Unpublished raw data.

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, CA: Academic Press.

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

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *Quarterly Journal of Experimental Psychology, 37B*, 1–21.

Stout, S. C., & Miller, R. R. (2007). Sometimes competing retrieval (SOCR): A formalization of the extended comparator hypothesis. *Psychological Review, 114*, 759–783.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgements: 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, NJ: Erlbaum.

Wasserman, E. A., & Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causality judgment: The role of within-compound associations. *Quarterly Journal of Experimental Psychology, 51B*, 121138.

Wasserman, E. A., & Castro, L. (2005). Surprise and change: Variations in the strength of present and absent cues in causal learning. *Learning & Behavior, 33*, 131–146.

Wheeler, D. S., & Miller, R. R. (2008). Determinants of cue interactions. *Behavioral Processes, 78*, 191–203.

Witnauer, J. E., & Miller, R. R. (2007). Degraded contingency revisited: Posttraining extinction of a cover stimulus attenuates a target cue's behavioral control. *Journal of Experimental Psychology: Animal Behavior Processes, 33*, 240–250.

Witnauer, J. E., & Miller, R. R. (2009). Contrasting the overexpectation and extinction effects. *Behavioral Processes, 81*, 322–327.

Witnauer, J. E., & Miller, R. R. (2010). Within-compound associations: Models and data. In N. A. Schmajuk (Ed.), *Computational models of conditioning* (pp. 108–149). Cambridge, UK: Cambridge.