

Elemental representation and configural mappings: Combining elemental and configural theories of associative learning

I. P. L. McLaren · C. L. Forrest · R. P. McLaren

© Psychonomic Society, Inc. 2012

Abstract In this article, we present our first attempt at combining an elemental theory designed to model representation development in an associative system (based on McLaren, Kaye, & Mackintosh, 1989) with a configural theory that models associative learning and memory (McLaren, 1993). After considering the possible advantages of such a combination (and some possible pitfalls), we offer a hybrid model that allows both components to produce the phenomena that they are capable of without introducing unwanted interactions. We then successfully apply the model to a range of phenomena, including latent inhibition, perceptual learning, the Espinet effect, and first- and second-order retrospective revaluation. In some cases, we present new data for comparison with our model's predictions. In all cases, the model replicates the pattern observed in our experimental results. We conclude that this line of development is a promising one for arriving at general theories of associative learning and memory.

Keywords Associative learning · Classical conditioning · Comparative cognition · Human learning · Stimulus preexposure · Latent inhibition

Introduction

Our long-term aim is to produce a general model of associative learning and memory that captures the processes that

are common to both humans and infrahumans. This article investigates the feasibility of combining elemental and configural approaches to associative learning and memory as a stepping stone toward that ultimate goal. In doing so, it also addresses a particular computational problem: How we can have representation development at an elemental level while still learning in a holistic fashion? We start by considering the problem in general and by motivating the need to find a solution incorporating both elemental (e.g., Brandon, Vogel, & Wagner, 2000; Estes, 1959; Harris, 2006; McLaren, Kaye, & Mackintosh, 1989; Wagner & Brandon, 2001) and configural (e.g., Honey, 2000; Honey & Ward-Robinson, 2002; McLaren, 1993, 1994; Pearce, 1987, 1994) forms of representation, then move to a specific example of such a combination that attempts to amalgamate the McLaren, Kaye, and Mackintosh (1989) model (henceforth, the MKM model) of representation development with the APECS (Le Pelley & McLaren, 2001; McLaren, 1993, 2011) model of associative learning and memory. To anticipate slightly, the enterprise is a successful one in the sense that the hybrid model is able to reproduce the phenomena that can be simulated using its components (and thus is of wider scope than either of its constituent parts), but this outcome was not achieved without considerable effort and overcoming numerous difficulties. In the course of grappling with this problem, we have developed a new respect for the way in which issues multiply as the complexity of the model increases, and we try to pass on our experience of what will and will not work when synthesizing elemental and configural approaches to learning and memory.

This represents our first attempt at combining a theory of stimulus representation that operates at an elemental level (that due to McLaren, Kaye, & Mackintosh, 1989; further elaborated in McLaren & Mackintosh, 2000) with a theory of associative learning and memory that is clearly of a

I. P. L. McLaren (✉) · C. L. Forrest · R. P. McLaren
School of Psychology, University of Exeter,
Washington Singer Building, Perry Road,
Exeter EX4 4QG, UK
e-mail: I.P.L.McLaren@exeter.ac.uk

configural nature (APECS; Le Pelley & McLaren, 2001; McLaren, 1993, 1994, 2011). While the benchmark elemental model of associative learning over the last 40 years has been the Rescorla and Wagner (1972) model, more recently, results such as those from retrospective revaluation studies and from experiments on latent inhibition and perceptual learning have suggested that the Rescorla and Wagner model can no longer accommodate important findings in the human (see, e.g., Dickinson & Burke, 1996, and Larkin, Aitken, & Dickinson, 1998, on retrospective revaluation; McLaren, 1997, McLaren, Leavers, & Mackintosh, 1994, and Wills, Suret, & McLaren, 2004, on perceptual learning) and animal (see, e.g., Matzel, Schactman, & Miller, 1985, and Matzel, Shuster, & Miller, 1987, on retrospective revaluation; McLaren, Bennett, Plaisted, Aitken, & Mackintosh, 1994, on latent inhibition; Aitken, Bennett, McLaren, & Mackintosh, 1996, on perceptual learning) literature. Miller and colleagues have taken this further by assessing the Rescorla–Wagner model against what is currently known about associative learning and, while they find it a useful benchmark, noting that there are a number of phenomena that it cannot accommodate (Miller, Barnet, & Grahame, 1995). Given this, our strategy is to consider in detail two problem domains that are appropriate test beds for the elemental and configural model classes (we will also refer briefly to a number of other phenomena to illustrate the generality of our approach). One of these problem domains encompasses retrospective revaluation, where both elemental and configural theories vie to explain the data. We will argue that a configural approach is the more successful here. The other focuses on representation development—in particular, the role of stimulus–stimulus associations in latent inhibition and perceptual learning. We attempt to characterize what each approach can bring, computationally, to these problem domains and then use our hybrid MKM–APECS model as a means of illustrating the benefits to be derived from their amalgamation. We begin by introducing the problem domains themselves.

Problem domain 1

Here, we are concerned with representation development, as exemplified by phenomena such as perceptual learning, latent inhibition, and the Espinet effect. We argue in this article that the proper role for elemental models lies in providing the input to configural systems, so that these configural systems can then associate inputs to outcomes and store the result in memory. Thus, the elemental contribution is one of representation development that takes place over time and as a consequence of experience with stimuli, and it will come as no surprise that this is one area that we will focus on in this article as a test domain for any attempt

to combine these two classes of theory. Our view is that configural theories struggle to provide the mechanisms for such basic phenomena as latent inhibition and perceptual learning, brought about as a consequence of preexposure to a stimulus or stimuli. Even so, configural representations still have something to offer (e.g., in explaining recovery effects and context effects) and can enhance our ability to explain the full range of stimulus exposure phenomena. We must be wary of the possibility, however, that the attempt to combine configural and elemental theories, rather than simply delivering the sum of what each model class can do (which would be an entirely acceptable outcome), instead turns out to cause them to interact in such a fashion as to introduce more problems than the combination solves. Thus, our challenge will be, in some sense, to combine what we see as a successful theory of representation development with a configural approach to learning and memory without either introducing inappropriate phenomena or losing the ability to generate appropriate effects.

In this problem domain, the basic phenomena are well known, and there are numerous models capable of explaining them. Thus, preexposure to a stimulus will, other things being equal, retard learning to that stimulus in the same, but not a different, context (e.g., Lovibond, Preston, & Mackintosh, 1984), unless the context is itself familiar (McLaren, Bennett, et al., 1994). Preexposure to stimuli that subsequently have to be discriminated will, in some circumstances, facilitate rather than retard acquisition of the discrimination (see Hall, 1980, for an early review, and McLaren & Mackintosh, 2000, for a later one). In this article, we will focus on the relationship between these two effects, since taken together, they pose a challenge for any unitary explanation of the effects of stimulus exposure on learning. Typically, models will explain one or the other of these effects (e.g., Gibson's (1969) explanation of perceptual learning; Pearce & Hall's (1980) explanation of latent inhibition), and if they attempt to explain both, will appeal to different processes for latent inhibition and perceptual learning (e.g., Saksida's (1999) model of perceptual learning, which in effect combines Pearce and Hall's alpha modulation with a nonassociative connectionist model based on competitive learning that implements Gibson's ideas). McLaren, Kaye, and Mackintosh's (1989) model of representation development is different in this regard, in that it uses the salience reduction consequent on stimulus exposure that causes latent inhibition as one of the mechanisms that drives perceptual learning. The differential latent inhibition of common elements mechanism for perceptual learning relies on the better predicted and more often encountered shared features of a discrimination becoming relatively less salient than those features unique to each stimulus that serve as the basis for successful discrimination. This approach to perceptual learning makes some strong predictions. Clearly, the relationship between latent inhibition and

perceptual learning should, under some circumstances, be directly demonstrable. And we do have Rodrigo, Chamizo, McLaren, and Mackintosh's (1994) demonstration of both latent inhibition and perceptual learning as a consequence of preexposure in the radial maze (see also Prados, Chamizo, & Mackintosh, 1999; Sansa, Chamizo, & Mackintosh, 1996) and Trobalon's as well as Bennett and Tremain's work, as discussed in McLaren and Mackintosh (2000, pp. 230–232), to support this analysis in rats. In the former experiments, manipulation of preexposure to involve the unique (landmarks at end of maze arms) or common (landmarks in between maze arms) elements of the discrimination led to latent inhibition in the first case and perceptual learning in the second. In the latter experiment, preexposure produced latent inhibition that resulted in perceptual learning, but extended preexposure produced more severe (i.e., near asymptotic) latent inhibition that then eroded the perceptual learning effect. There is also ample evidence that perceptual learning requires a difficult discrimination such that the discriminanda are sufficiently similar to one another (e.g., Oswalt, 1972) and that, if this requirement is not met, preexposure can instead lead to a deficit in learning (e.g., Trobalon, Sansa, Chamizo, & Mackintosh, 1991). All these results (which are predicted by the MKM model) suggest that there is an intimate relationship between latent inhibition, on the one hand, and perceptual learning, on the other.

Another prediction that follows from this model, and one which we will focus on in this article, is that preexposure to a category that is defined in terms of a prototype should enhance the ability to discriminate among the members (exemplars) of that category. We have previously shown (McLaren, 1997; McLaren, Leavers, & Mackintosh, 1994) that exposure to exemplars drawn from a category defined by a prototype leads to perceptual learning, as evidenced by an enhanced ability to discriminate between category exemplars after preexposure. And in pigeons, we have been able to show that exposure to the prototype alone can have effects similar to those predicted by the MKM model (Aitken et al., 1996). Now, we are able to extend this result to show that exposure to a single, prototypical stimulus has a similar effect for humans, in that it results in faster acquisition of a discrimination between exemplars drawn from that category (see later in this article). This result is important because it establishes that, in some circumstances at least, perceptual learning is not contingent on the opportunity to compare stimuli (so as to discover the diagnostic features required for later discrimination), ruling out models that see this process as both necessary and sufficient for the demonstration of perceptual learning. It also links Gibson, Walk, Pick, and Tighe's (1958) result in rats and Attneave's (1957) finding in humans, that distortions of a preexposed stimulus are easier to discriminate between than distortions of a novel

stimulus, to the literature on how familiarization with a prototype-defined category can enable better discrimination between members of that category (e.g., McLaren, 1997). Configural theories struggle (because of their holistic nature) to cope with data of this type, because they strongly suggest that some features become advantaged relative to others as a consequence of simple preexposure. As we have indicated, the challenge we face in this problem domain is whether the type of elemental, associative mechanism for latent inhibition and perceptual learning instantiated in the MKM model can be combined with the configural account of associative learning and memory in such a way that the desirable features of the elemental theory are preserved and, in addition, the conditional properties of configural models are brought into play.

Problem domain 2

Here, we focus on studies of retrospective revaluation (RR). In a typical RR experiment with human participants, compound stimuli AB and CD are presented with reinforcing outcomes (i.e., AB + CD+). Following this, A is presented alone and followed by reinforcement, whereas C is presented alone without reinforcement. In the final test phase, the causality ratings for B and D are measured, and typically, the ratings for B are found to be less than those for D. It therefore seems that changes to the associations of B and D with the outcome must occur during their absence in order to account for the differences in their ratings. The effect for B is known as *backward blocking*, and for D as *unovershadowing*. We note in passing that these effects are not always obtained, at least in animals, since sometimes mediated extinction has been the reported consequence of following CD + with C – that is, responding to D decreased (e.g., Ward-Robinson & Hall, 1996). This is, in itself, an interesting result and one that we hope to return to in the future, but for present purposes, we confine our analysis to the RR phenomenon. The Rescorla and Wagner (1972) model has no mechanism to allow learning about absent stimuli, but modifications of the rule have been proposed (e.g., Van Hamme & Wasserman, 1994), and modifications of other theories have also been suggested (e.g., a modification of SOP, Wagner, 1981; put forward by Dickinson & Burke, 1996) to deal with this problem, and these have met with some success. But we do not believe that these elemental models are viable as explanations of RR, for reasons that we will now discuss.

One consideration that led to this decision concerned a result—known as the *Espinet effect*—that is thought to be mediated by inhibitory associations developed as a consequence of preexposure (Bennett, Scahill, Griffiths, & Mackintosh, 1999; Espinet, Iraola, Bennett, & Mackintosh,

1995; Graham, 1999). Espinet et al. took thirsty rats and preexposed them to distinct compound flavours, AX and BX, on alternate days. After a single conditioning trial in which a solution containing A alone was followed by LiCl injection (to induce illness), the rats were slower to learn an aversion to solution B, when solution B was subsequently paired with LiCl injection. The authors went on to show that solution B now acted as a conditioned inhibitor and signaled the absence of illness, as a result of solution A being paired with illness.

In the RR paradigm, Dickinson and Burke (1996) had shown that reliable co-occurrence of A and B during the initial stage, which would result in the formation of excitatory associations between them, was vital if RR effects were to be obtained. But Espinet et al. (1995) showed that alternating preexposure to AX and BX, which would result in inhibitory associations between unique A and B flavors (because the presence of A signals the absence of B and vice versa), produced an effect such that conditioning A to illness either weakened B's association with illness or made B an inhibitor. Both results challenged standard associative theories of conditioning, but the combination of these two results has proved an even greater challenge to accommodate in any one theory. Clearly, a simple negative alpha account will not do, because if we assume that activation of a stimulus representation due to excitatory associations will result in learning to that stimulus, but with the opposite sign to that obtained when the stimulus is physically present (the defining characteristic of these elemental accounts), then activation due to inhibitory associations should have the opposite effect again—that is, produce learning in the same way as that which would be supported by physical presentation of the stimulus. But in both backward blocking and the Espinet effect, the change in associative strength is in the same direction (i.e., downward). Until recently, only Graham (1999) had been able to construct a model that could accommodate both these results. And, to the best of our knowledge, there is no reliable evidence for inhibitory activation of a stimulus producing learning to that stimulus analogous to that obtained when the stimulus was physically present, as the negative alpha or modified SOP accounts would require. For this and other reasons, then, they seem to be contraindicated.

Furthermore, even Graham's (1999) ingenious variation on Wagner (1978) will not accommodate the second-order RR results that we report later on in this article. Given this, we take the position that elemental theories of RR struggle to accommodate the data in both humans and other animals. We also note that Le Pelley and McLaren (2001) and, later, McLaren (2011) were able to show that a configural theory of learning and memory—APECS—was able to model these RR results in a manner that is entirely independent of any need to postulate the within-stimulus associations

that lead to conflict between findings such as the Espinet effect and RR. It did this by interpreting RR as an effect in memory, rather than in learning. Training either A + or A – after initial AB + learning had the effect of altering the retrievability of the initial configural memory for the earlier learning to AB +, and this produced the revaluation effects. Given that elemental theories (e.g., McLaren & Mackintosh, 2000, 2002) are able to explain the Espinet effect, we were motivated to ask whether by combining these two theories, one configural, the other elemental, we could arrive at a theory that encompasses the full set of phenomena. Hence, experiments involving the revaluation of stimuli that have already been paired in the past, as is the case for studies of both RR and the Espinet effect, would seem to be a useful problem domain to consider. This domain also has the advantage of linking back to the first problem domain we identified, because the Espinet effect is also, in some sense, an effect of preexposure.

We now give a brief description of each of our model candidates (elemental and configural) in turn, with a quick summary of what they are able to do singly. Then we briefly consider different methods of combination and the results that the hybrid models produce.

MKM elemental theory

We began with the basic equations taken from McLaren, Kaye, and Mackintosh (1989) that instantiate the version of the delta rule used in that model that lends itself to salience modulation. Our model architecture was to have a representation layer that received input through fixed one-to-one weights, thus providing external input to these units. In our first version, every unit on this input representation layer was linked to every other unit on the representation layer by modifiable links. These links followed the MKM algorithm in attempting to equate the internal input with the external input to each unit in the representation layer; that is, the links change such that the error score for a unit is minimized. The activation of units in the representation layer was computed from the input it received both externally and internally from other units within the layer. Modulation of unit activation was implemented by multiplying the external input to that unit by 10 times the error value for the unit if that error value was positive. This representation layer (which is essentially MKM) was then linked via a hidden unit layer to an output layer on which the target outputs are set in the same way as for a backpropagation network. The model was trained with 0.9 for an output unit that was on and 0.5 for an output unit that was off. The activation of the output layer is calculated using the standard logistic activation function, and the weights from input to hidden and hidden to output layers changed using the generalized delta rule (Rumelhart, Hinton, & Williams, 1986). Thus, with

experience, the network would be able to use the input it receives to refine its representations at input and, meanwhile, learn to link these representations to the correct output representation.

In essence, this implements the rather simplistic approach of grafting the MKM model onto the input of a simple feedforward backpropagation network. This basic system is nevertheless capable of modeling a range of phenomena, including (but not restricted to) simple acquisition and extinction, discrimination learning, blocking, overshadowing, overexpectation, latent inhibition and its context specificity, recovery from latent inhibition, and perceptual learning. These results, especially those involving perceptual learning and latent inhibition, are not surprising, since the MKM model was specifically developed to give an account of these phenomena. Our next step was to introduce the configural APECS component into the model by modifying the backpropagation component of the model to produce the architecture shown in the top panel of Fig. 1. Before we discuss this hybrid architecture, we will introduce APECS and briefly motivate the need for this configural component to fully capture what we know about associative learning and memory in humans and infrahumans.

APECS configural theory

The APECS configural theory of associative learning and memory was introduced in McLaren (1993), further developed in the discussion contained in McLaren (1994), and then applied to a wide range of associative learning phenomena in Le Pelley and McLaren (2001) before its most recent refinement in McLaren (2011). The last is the version of APECS that we will use here. Le Pelley and McLaren had already shown that APECS could give a good account of first-order RR, in that it produces a marked unovershadowing effect and a somewhat weaker backward-blocking effect, a result that was in line with the behavioral data reported in that and other articles (e.g., Larkin et al., 1998). The explanation given in Le Pelley and McLaren for the unovershadowing effect was that training A – after AB + training led to the hidden unit carrying the AB + mapping becoming more easily activated, because its bias became less negative; in effect, memory for AB + training became more accessible, more readily retrieved. This occurred because after A – training, when A was no longer presented, the equilibrium state of the network had to adjust so that the outcome was not negatively predicted; that is, the unit representing the outcome did not have a negative error score because of inappropriate inhibitory activation from the hidden unit mediating the A – mapping. In doing this, the easiest solution for the network was to raise the bias (make it less negative) of the hidden unit representing the AB +

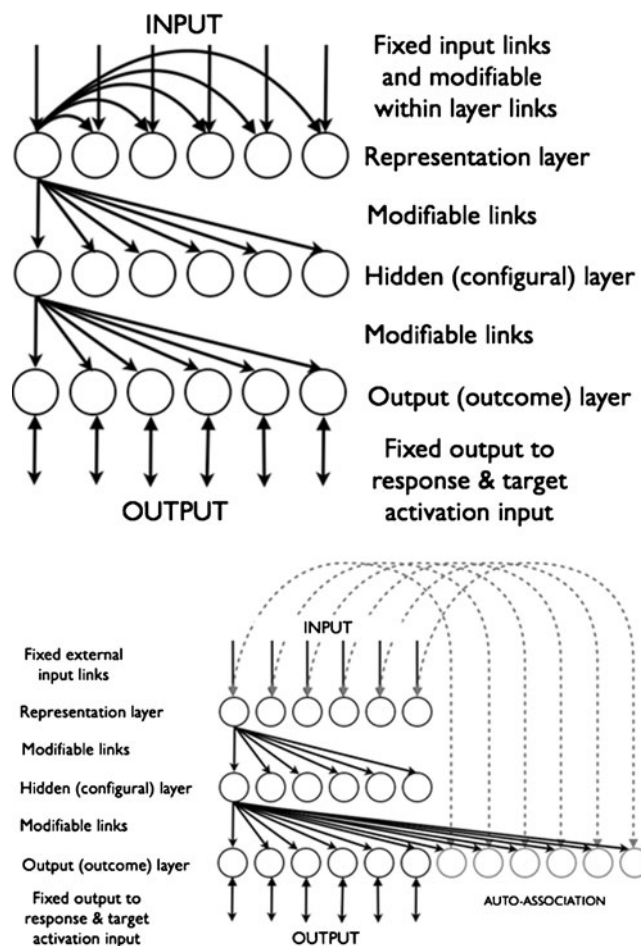


Fig. 1 Top panel: A simple combination of an elemental MKM system acting as the input to an APECS feed-forward network. The input units in the representation layer are completely connected to one another (only some links are shown for clarity), and the links between units are used to generate error scores (the difference between the external input supplied and the internal input from other input units), which then control or modulate the salience (activity) of each unit. These are then used to learn the input to output mappings via the hidden units. Bottom panel: A quite different autoassociative architecture. Once again, only some links are shown for clarity. The input units are not connected to one another but, instead, have counterparts on the output layer that can be associated to and that can then feed input back to the input units they correspond to. This is done with a fixed weighting of 0.4 times the difference between the corresponding output unit activation due to internal input and its resting level (0.5) in the simulations reported here. The error score for the appropriate output unit now controls modulation of the input unit, and this is done by increasing the external input to the unit by a factor of 10 times the corresponding output unit error score if that error score is positive. All links from input to output via the hidden or configural layer are learned via APECS

mapping. This meant that later on, presentation of B on its own was more effective in activating this hidden unit and, hence, activating the unit representing the outcome. The analysis of backward blocking similarly required consideration of the networks equilibrium state after training (see Le Pelley & McLaren, 2001, for a full discussion), and these

analyses continue to hold for the version of APECS used here and the hybrid model we are about to consider. The 2001 version of APECS was also capable of simulating learning to a partially reinforced CS, and predicted that there would be no excitatory learning between two associatively activated representations in memory, a result reported in Le Pelley and McLaren, although contradicted by the work of Dwyer, Mackintosh, and Boakes (1998). Successful simulations of Dickinson and Burke's (1996) demonstration that RR in a standard AB + | A + vs. CD + | C – design was observed only when the CSs were consistently paired were also reported. The last is noteworthy since, up to this point, the Dickinson and Burke data were taken to indicate that RR could occur only when there were strong between-CS associations (brought about by their consistent pairing), but our simulations with APECS did not require the existence of these between-CS associations to generate the result. APECS was also able to simulate the phenomenon of backward conditioned inhibition first reported by Chapman (1991). This uses an AB – | A + vs. CD – | C – design, and the first treatment in terms of APECS was given in Le Pelley, Cutler, and McLaren (2000). Larkin et al. (1998) also showed that consistent CS pairing was a necessary condition for this phenomenon, a result that APECS was able to generate as well. Thus, APECS was already able to model a wide range of RR phenomena, but subsequent developments suggested that it lacked the ability to give a full account of RR (second-order effects), and the McLaren (2011) version was developed to deal with this issue.

The computational principles governing the newest version of APECS are the following:

- A new hidden unit is recruited to represent each novel mapping of input to output. This is an automatic consequence of the architecture of the model and the algorithm. As an example, following AB+, A+, A – training, typically three hidden units will have been recruited—one carrying an AB + mapping, one an A + mapping, and the other an A – mapping. The selected units' learning rate parameters are set high: 0.8 for connections to and from any active hidden unit, to 0.1 for the bias (the bias is the weight to a hidden unit from an input unit that is always on). Unselected units default to .0005 for the first two parameters and .005 for the bias. Thus, a hidden unit that has not yet been recruited to carry a mapping has all learning rates set to near zero; that is, it effectively takes little or no part in the learning process (but can adjust its bias very gradually). These rules have the following exceptions:
- Hidden units that have been recruited previously and have a negative error score as a result, have their bias parameter set high (i.e., 0.1), but not the other learning parameters, which remain at .0005. The fact that the unit

was recruited previously is determined from inspection of the individual components of error propagated back to that unit. In this case, given an overall negative error score for the hidden unit, an individual contribution from an output unit that is more negative (i.e., lower) than – .0025 is taken as the criterion.

- If a unit that has a positive error was previously recruited (criterion of a positive back-propagated error component greater than .0025 and no substantial negative error component due to previous learning, defined as before), then the bias is set high at .1, but not the other learning rate parameters, which remain at .0005.
- Each trial is now followed by a posttrial learning phase in which only biases are allowed to vary. This concept goes beyond the simple idea that the network should free-run during the intertrial interval (ITI). It implies control of learning such that the network cycles between phases of learning mappings from input to output and then adjusting biases with no input or output so that it reaches equilibrium.
- Each trial and each ITI involves 200 learning cycles (the minimum that seems to be effective). All weights are initially set to small random values.

With these changes, we can confirm that the APECS component of the model is capable of giving simple first-order RR effects, as well as more complex first- and second-order effects in multiphase designs of the type we will consider in a moment. We defer our explanation of how it does so until then. We now evaluate the hybrid model produced by combining this with the elemental MKM model described earlier to see whether the hybrid is then capable of encompassing a wider range of phenomena than either model on its own.

The hybrid model: MKM–APECS

We begin by considering the most obvious and straightforward combination of these two modeling approaches. We have already noted that by simply using our MKM model to provide input to an APECS network, we arrive at the architecture shown in the top panel of Fig. 1, in which all input units in the representation layer are connected to one another and also connected to all the units in the hidden layer, which then connects (again completely) to the output layer. This strategy, which certainly seemed worth pursuing, since each component model offers something that the other lacks, does not, as far as we are able to establish, succeed. Once we had constructed this model, we then explored whether the hybrid network would be capable of generating the required RR effects and the other phenomena that we cover in this article. We experienced some success with this

architecture, and by varying assumptions and parameters, we were able at one time or another to solve all the problems presented to the model. But, ultimately, we were unable to find a single set of parameters that would allow us to simulate all of the effects needed for a comprehensive solution to the problem space we are considering in this article. While we are not in a position to rule out this approach entirely, we can say that it has not proved the most productive in our work in attempting to combine elemental stimulus–stimulus models of representation development with configural models of stimulus–outcome learning. As one example of the difficulties we encountered with this architecture, it typically had a strong tendency to generalize between mappings so that, rather than an unovershadowing $AB + |A -$ design producing stronger responding to B, it often actually caused it to decline, relative to control conditions. This was directly attributable to the stimulus–stimulus associations formed by the MKM component of the model, since, if these were turned off, the model was able to demonstrate unovershadowing. This was not entirely unexpected; while we had hoped that stimulus–stimulus associations would help as far as second-order RR effects were concerned (and they did), it had always seemed to us that it might have problematic implications for the first-order effects. Our experience in working with this architecture was that if we managed to find parameters such that we got either of the necessary first- and second-order effects, then this was at the expense of losing the elemental salience modulation properties of the hybrid, which defeated the object of the exercise.

Accordingly, we moved to the second version of our hybrid model architecture, shown in the bottom panel of Fig. 1. Instead of simply bolting MKM onto the front end of APECS, this approach required more integration between the two models at a conceptual and algorithmic level. The concept underpinning this architecture is that the model is a combination of the standard stimulus–outcome mapping architecture used in APECS with an autoassociative component that, in effect, allows for stimulus–stimulus associations. Changes in these associations are driven by the APECS algorithm in the same way as for other associations (i.e., via the hidden [configural] layer), but they give rise to associatively activated input on the representation layer and are also used to generate the error term that controls modulation of the input units' salience (i.e., the units' activation). The figure attempts to illustrate this computational arrangement explicitly by showing that the representation layer connects to a section of the output layer that simply replicates the input layer and uses the inputs to set target activations. But note that it could just as easily (and more elegantly) depict this architecture as simply involving links back from the hidden layer to the representation layer, as long as the computations were handled in the same way. We

hope that the “unpacked” illustration provided helps readers understand how the computations are done, but the more concise recurrent architecture (shown later in Fig. 6) is the one that motivated this approach. The error at these outputs is then used, on a one-to-one basis, to control modulation of the input units in the representation layer, and the activation of the output unit corresponding to a given input unit is used to determine the internal input delivered to the input unit. The last is responsible for the associative activation of input units even when no external input is applied to that particular unit. We have found that this system works well and seems to incorporate the best features of both models. We now demonstrate how it deals with a variety of problems that we have tested it on and, in particular, how it fares when asked to cope with our data on RR, latent inhibition, and perceptual learning.

Simulations and experiments

Model and parameters

For details on model implementation, please consult the primer in the online repository that accompanies this article. The issue of what are “free” parameters in the model is itself an interesting one. In some sense, none of them are free. They vary adaptively, yes, but once set, this adaptation is carried out by the model. We used the same parameter settings throughout our simulations, rather than changing them to produce a better fit to a given problem. If, however, we were asked how many parameters could meaningfully be varied to create different versions of this model, then, including constraints such as the number of input, hidden, and output units, learning rates, criteria, and so forth, the answer would be about 12. The residual uncertainty in this estimate stems from difficulty in deciding what should be counted as a parameter in the model. For example, we used an architecture with 10 input units, 20 hidden units, and 14 output units, of which 10 corresponded to the input units. How many of these (fairly arbitrary) decisions count as “free parameters”? In what follows, we present some of the simulation results obtained with this model. Space constraints prevent us from considering all the simulations we have run, but we do attempt to indicate other results where possible.

Retrospective reevaluation

We begin with RR, since we have already indicated that this proved problematic for earlier hybrids. The first thing we established was that the model had no difficulty in producing unovershadowing and (to a lesser extent) backward blocking (these first-order RR effects are contained within the data presented when we consider second-order effects).

Thus, it did not suffer from the problems attendant on our initial attempt at hybridization. Then we moved on to consider first- and second-order RR effects in combination. To illustrate how we did this, we will introduce the results of a recent experiment in some detail and then go on to show how the model copes when asked to simulate them.

In our experiment, the participants are presented with four experimental problems and four filler problems. In each problem, they are shown whether Mr. X has an “allergic reaction” after eating particular foods or if he “feels fine.” There are three food/food pairs within each problem, and they are presented in three distinct learning phases such that all the presentations from one phase are made before moving on to the next (although the participants were unaware of this division into phases). The design is shown in Table 1 below. The first two problems (1 and 2) are what we call *forward* designs. The last two (3 and 4) are the RR problems. The fillers were chosen to balance up the positive (+ = “allergic reaction”) and negative (− = “feels fine”) outcomes at each stage. Each learning phase presented each of the problems (and fillers) 6 times. Meals were presented one at a time, and the participants had to decide whether an allergic reaction would occur or not. Feedback was then given so that they were able to learn the correct response for each meal by trial and error. At the end of all the learning phases, there was a final rating phase in which each food was evaluated individually on an 11-point scale for its likeliness to provoke an allergic reaction, with 0 being *very unlikely to do so* and 10 being *very likely*.

The results of this study with 32 participants for the RR problems (3 and 4) are shown in the top panel of Fig. 2.

We can see that, in contrast to the results reported by De Houwer and Beckers (2002) and by Melchers, Lachnit, and Shanks (2004), the second-order effect here for food C is in the same direction as that for the first-order effect for food B. We have investigated this matter further to ascertain the conditions under which we are able to obtain their results and have discovered that running essentially the same experiment using a questionnaire-based protocol, with exactly the same problems but with all the phases for a given problem

available on the same page (thus minimizing memory load for our participants), would allow us to obtain the result reported by De Houwer and Beckers and by Melchers et al. That is, the A − condition (unovershadowing, problem 1) leads to a higher rating for B than does the A + condition (backward blocking, problem 2), but now this effect reverses for the second-order cue C. Our conclusion is that, in these circumstances, when memory load is low and all components of the problem are available for inspection, rational inference produces this pattern of results. But in cases where memory load is a real factor (e.g., the computer-based version of the experiment described earlier) and ratings are taken at the end after the learning phases, the second-order result is quite different. This is clearly an important finding for our purposes, because it implies that explaining this second-order effect should not fall within the scope of an associative theory, and we would argue that the results reported by De Houwer and Beckers and by Melchers et al. may well have been produced by something other than an associative mechanism and, so, should not be modeled by an associative mechanism. Instead, this demonstration of second-order RR can be captured by a theory that posits a process of symbolic inference on the part of the participants. Taking the BC + | AB + | A + problem, if they reason that, since A is paired with and predictive of the outcome on its own, then, starting from this point, they can chain back and (assuming perfect memory for the other compounds) infer that B does not have to be (since A was there), generating a first-order effect when tested, and that C (other things being equal) has at least as good a chance of being predictive of the outcome as B. If we now consider the BC + | AB + | A − problem, A is now definitely not predictive of the outcome, so B is; hence, C is less likely to be. Thus, the second-order effect observed by De Houwer and Beckers and by ourselves can be generated by these heuristics (and our participants explicitly claimed to be using them), which produce the correct pattern of effects observed in the data. Our requirement here, however, is for MKM-APECS to be able to model the type of second-

Table 1 The design of the allergy prediction experiment demonstrating first- and second-order retrospective revaluation effects. Training took place in three distinct phases to four problems and was by means of trial (participants predicted the outcome) and error (they then

received feedback). Filler problems were used to equate the occurrence of compound and singleton cues and the two outcomes (“allergic reaction” and “feels fine”). Ratings were taken at the end of training phase 3

Learning	Phase 1		Phase 2		Phase 3	
Problem	Food codes	Reaction	Food codes	Reaction	Food codes	Reaction
1	A	−	AB	+	BC	+
2	D	+	DE	+	EF	+
3	MN	+	OM	+	O	+
4	PQ	+	RP	+	R	−

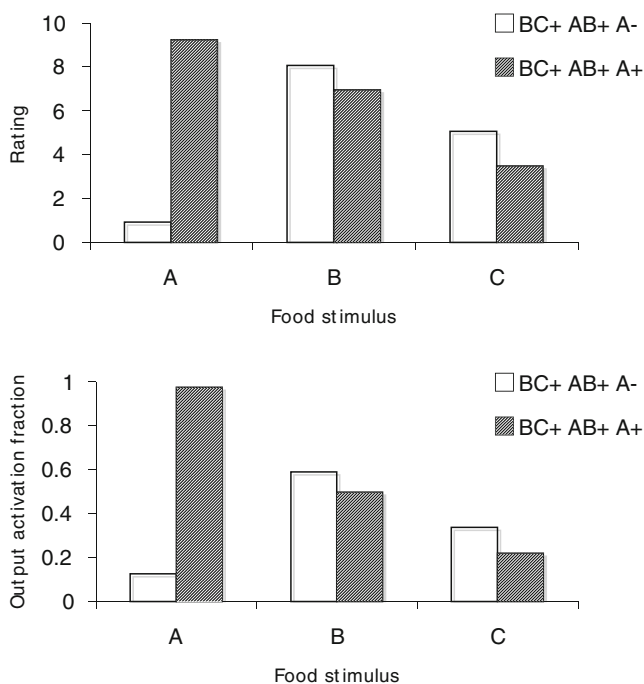


Fig. 2 Ratings data (see text) for our second-order retrospective revaluation design (top panel) and the simulation using the hybrid model (bottom panel). The top panel shows that the rating (high = outcome more likely, maximum value = 10) for A + is much higher than that for A − (low value = outcome less likely, minimum value = 0) and that the first-order RR effect on B is for its rating in the A + condition (backward blocking) to be lower than in the A − condition (unovershadowing). The second-order effect on C is similar, although numerically larger. The simulation results give scores that represent the fraction of the total possible activation of the output unit representing "allergic reaction," using the resting state as a baseline. Thus 0 implies no learning, and 1 is the maximum score. The pattern of results parallels those of the empirical data

order RR effect found in the computer-based high memory load version of the experiment, as well as the first-order effects already considered.

The bottom panel of Fig. 2 gives our simulation results for this design. The pattern is very much the same as that in our data, with the output unit activation (the model's equivalent of a rating) highest for the unovershadowing condition for both B (first-order effect) and C (second-order effect). In both cases, the RR effect is significant, smallest $F(1, 31) = 4.15$, $p < .05$, with, if anything, the larger effect for the second-order case, although this does seem to be a somewhat parameter-dependent result. As McLaren (2011) noted, the last finding (which we have replicated) would call into question theories and models that rely on chained associative activation of stimulus representations to drive retrospective revaluation, because if this were the case, the activation of B by presentation of A must inevitably be stronger than that of C and, so, the effect should be greater for B than for C, which we do not observe to be the case. An explanation of how APECS can produce this pattern of results is already

available in McLaren (2011), so we will give only a brief characterization here. We start by noting that the addition of the MKM component to this model does not now seem to have hindered its ability to continue to produce this effect. In essence, APECS produces revaluation effects by first setting up configural representations of BC + and AB + and then altering their accessibility as a result of experiencing either A + or A −. If A + occurs, then the effect is for the unit carrying the mapping from A and B to + to become harder to activate—in effect, the memory of that learning has become harder to retrieve. This has the consequence that the ability of B to activate this unit and so activate the US representation is reduced. This first-order effect (backward blocking, in this case) is paralleled by a similar second-order effect. The unit carrying the mapping for BC + is also rendered less accessible, and so the effect for C is in the same direction as that for B.

The Espinet effect

One challenge that we posed for our model in the Introduction was to succeed in generating appropriate RR effects but, also, to produce an Espinet effect. Our next step was to present exactly this problem to the model. We used a design in which stimuli AX and BX were preexposed, either all AX before BX or vice versa. Then we conditioned A and tested B. We used this blocked design (even though it is a weaker variant than the original intermixed design) because the results are more informative, but note that we do obtain an Espinet effect with intermixed presentations of AX and BX. The results of this simulation are shown in Fig. 3 (bottom panel). The AX followed by BX condition produced significantly negative responding (i.e., an inhibitory effect) when B was presented after A had been conditioned. Control (A conditioned after no preexposure) and BX exposed before AX conditions produced little or no effect. This is very much in line with our and others' work on the Espinet effect designed to look at the issue of training order (for examples, see Espinet et al., 1995; Graham, 1999) and suggests that we have captured this phenomenon in the model. Our explanation of this effect is based on preexposure setting up a negative link between representations of B and A such that B inhibits A. If A is conditioned, the negative input to the representation of A when B is subsequently presented lowers its activation, which in turn lowers input to the unit representing the outcome, allowing it to take activation below its normal resting level.

Perceptual learning

The top panel of Fig. 3 shows the complementary results obtained for preexposure to AX and BX either in the blocked fashion used for the Espinet effect simulations

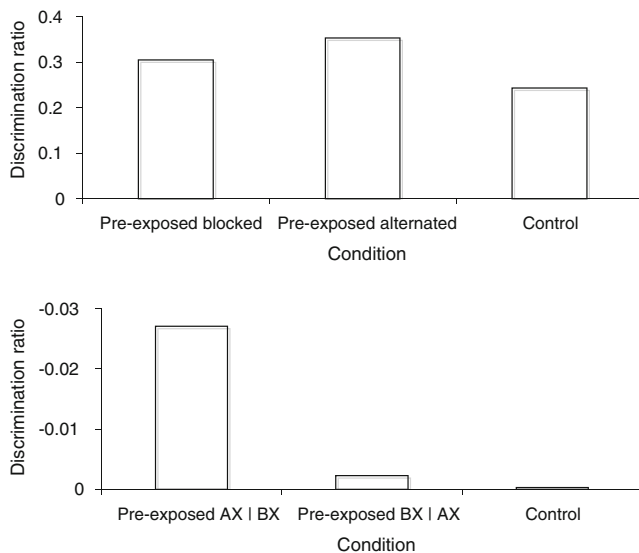


Fig. 3 Top panel: Results of a simulation of preexposure to AX and BX in either blocks or alternation, followed by conditioning of AX and testing of BX. This time the response measure given is a discrimination ratio so that higher scores indicate better performance: 0 means no learning, and 1 is the maximum possible. Controls were simply conditioned to AX without preexposure. Both preexposure conditions show higher (better) scores than does the control, but alternated preexposure is significantly superior to blocked. Bottom panel: Results of a simulation of the Espinet effect. AX and BX are preexposed (either AX | BX or BX | AX), and then A is paired with the outcome. Testing to B reveals no learning for the control (nonpreexposed) and little for BX | AX conditions but a weak (although significant) negative discrimination ratio for the AX | BX condition, indicating that, in this case, B has become inhibitory

(the results given are averaged across the AX followed by BX and the BX followed by AX conditions) or using alternated preexposure to these two stimuli (as well as a control condition that had no preexposure), but with preexposure followed by conditioning to AX rather than just A, and testing to AX and BX. The result is that blocked preexposure produces better performance in discriminating between AX and BX than is obtained in the control condition but alternated preexposure leads to better discrimination still. We believe this to be the first demonstration of such an effect by simulation, and it fits well with demonstrations of such an effect (e.g., Hall, Blair, & Artigas, 2006; Mitchell, Nash, & Hall, 2008). The mechanism here seems to be one that could explain the finding reported by Hall and Rodriguez (2009), characterized by Hall (2003, 2009) as alternated preexposure leading to associative activation of the unique components of the stimuli, which allows for some restoration of the loss in salience to these components that would otherwise have occurred. The outcome is that X suffers from greater differential latent inhibition (relative to A and B), and the discrimination between AX and BX is more easily acquired.

We must acknowledge, however, that there are other theories of perceptual learning—typically, those based on the principles put forward by Gibson (1969) and involving an appeal to comparison processes (see, e.g., Mundy, Honey, & Dwyer, 2007)—that already exist within a configural learning framework and can also account for these basic phenomena. Our response is to offer some new data that we believe require salience reduction mechanisms for their explanation, in conjunction with a demonstration of our model's ability to simulate them. The procedure used with our human participants was simple enough. They were preexposed to a novel checkerboard and then later required to discriminate between two new distortions of that now familiar checkerboard. Performance on this discrimination was compared with performance on two novel distortions of an unfamiliar checkerboard that were actually the experimental stimuli for one of the other participants in the experiment. We ran 32 participants in this experiment, and their ability to learn to discriminate between checkerboards was assessed by means of putting pairs of checkerboards on screen and asking them to learn which was the S+ by trial and error. The results are shown in the top panel of Fig. 4. We can see that there is an advantage for the preexposed

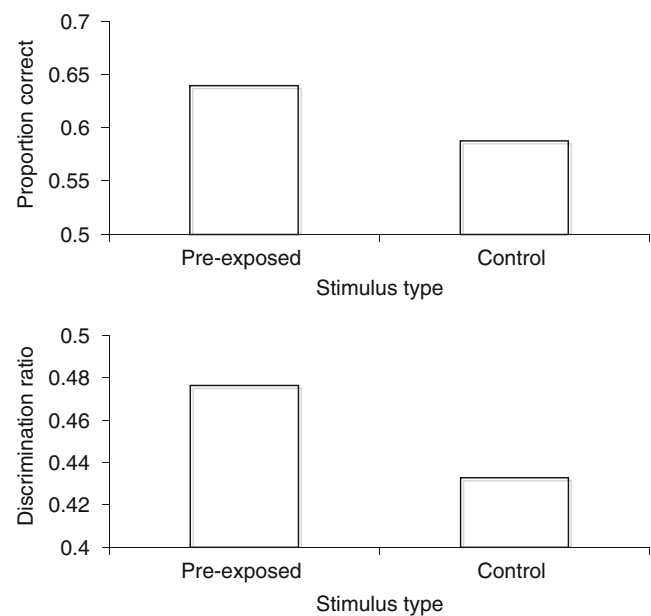


Fig. 4 Top panel: Results of the checkerboard preexposure experiment reported in the text. Discrimination between the two distortions of this preexposed checkerboard leads to a higher proportion of correct responses (the measure shown, chance = 0.5, 1 = perfect) than with the control pair of stimuli derived from a novel checkerboard. Simulation of this result was done by preexposing to AX, then training a discrimination between BX and CX. Bottom panel: The results. Again, a discrimination ratio is used such that 0 denotes no learning and 1 perfect acquisition

checkerboards in acquisition of the discrimination, which reaches significance, $F(1, 31) = 3.1$, $p(\text{one-tailed}) < .05$. The result is important because, since the participants were exposed to only one checkerboard from the hypothetical category that the test exemplars were drawn from, it is difficult to see how some comparison-based mechanism for perceptual learning could produce this result. The result is analogous to that obtained by Aitken et al. (1996) with pigeons and by Mundy et al. in one of their experiments with faces. The bottom panel of Fig. 4 shows our simulation of this study. In this simple simulation, we attempted to capture the essential nature of the checkerboard design by preexposing to AX, then training a discrimination of BX versus CX. The pattern is very much the same, with an advantage for the preexposed stimuli. We are able to claim, then, that our model can demonstrate perceptual learning in situations in which there has been no basis for comparison between stimuli drawn from the to-be-discriminated class. As far as we are aware, this class of model is the only one that can generate this result for this particular experiment, and it does so by reducing the salience of the common X element between BX and CX as a result of the earlier AX preexposure.

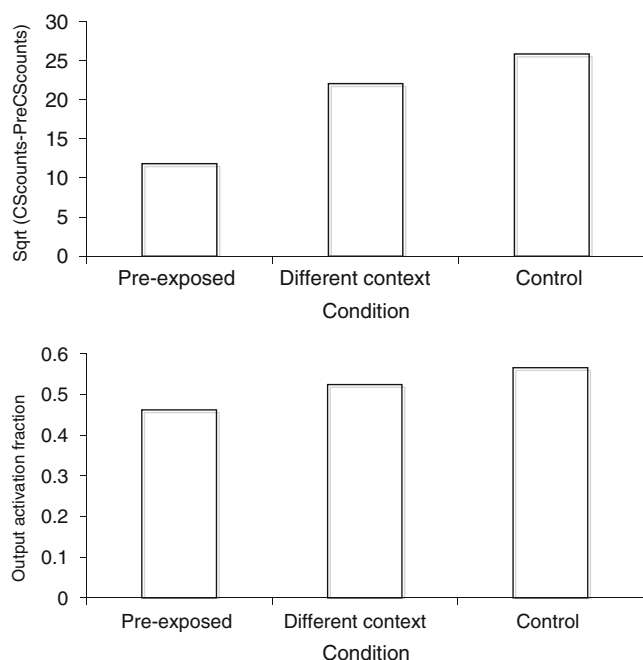


Fig. 5 Top panel: Results of a latent inhibition experiment carried out by McLaren (1990), in which stimuli were preexposed in one context and then conditioned in the same (preexposed) or different (different context) context. Controls received no preexposure. The measure shown is CS–PreCS photocell counts for magazine entry in a licking procedure. Bottom panel: Simulation of this experiment. The scores represent the fraction of the total possible activation of the output unit representing reward (0 = no learning, 1 = maximum). The preexposed condition shows latent inhibition in both empirical and simulated data; a change of context disrupts this (although not completely)

Latent inhibition

Our final simulation is of preexposure causing latent inhibition, and the effect of a change of context on this preparation. The design is simple: Stimulus A is preexposed in a context S and then conditioned either in the same context (AS|AS+) or in a different context (AS|AD+) that is equally familiar (preexposed on its own). The control condition simply removes the preexposure phase. This version of a latent inhibition experiment has the advantage of allowing us to analyze what the model is doing. Simulation results are shown in the bottom panel of Fig. 5 and indicate that conditioning is slower when preexposure and conditioning take place in the same context, but that this effect is greatly ameliorated by a change in context, although there is still a small but detectable retardation in acquisition, relative to controls. Our explanation for this effect is that the autoassociative learning of the representation of the stimulus in a particular context leads to salience reduction (less activation) for that representation, which leads to slower learning of the stimulus–outcome association when it is trained. Changing context means that the hidden unit responsible for carrying the autoassociative mapping is now no longer activated to the same extent; hence, the salience reduction is less, and learning of the stimulus–outcome mapping proceeds more rapidly. The top panel of Fig. 5 shows some lick-conditioning data from

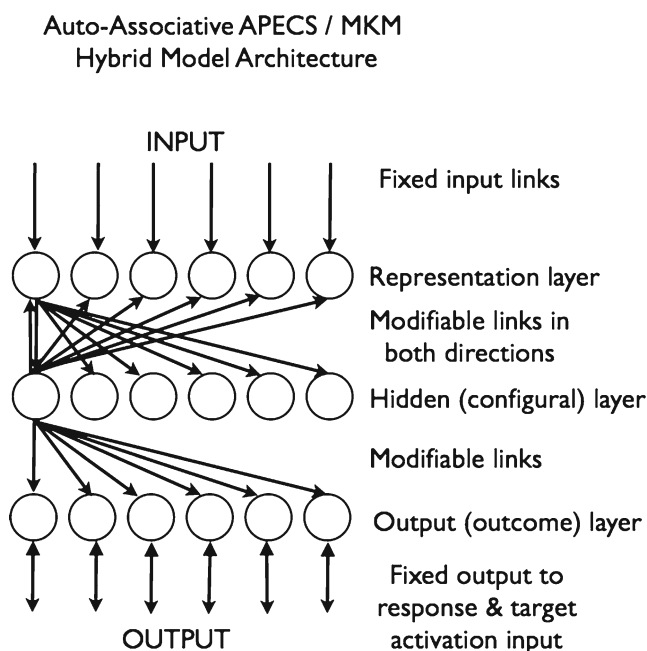


Fig. 6 The final MKM–APECS hybrid model architecture. Only some connections are shown for clarity, but all input units connect to all hidden units, and the hidden units connect back to all the input units via separate modifiable connections. Thus, the model auto-associates to the input layer but also allows for separate associative links between input and output

rats that were collected by McLaren (1990) in a similar design. The animals had been preexposed to tones and lights in one context and were then conditioned to lick for water to the CS, in either the same or a different context. Control animals were not preexposed to the stimuli. The conditioning measure is the square root of the difference between pre-CS and CS entries to the magazine. The pattern is similar to that in the simulation (especially given that there has been no attempt to "fit" the data in any of these simulations).

Conclusions

Our primary conclusion is that it is possible to combine the principles behind MKM and APECS successfully. A corollary is that this was not as straightforward a task as might initially have appeared to be the case! The most obvious combination failed to deliver the performance needed, and only by moving toward a more sophisticated model that integrated the conceptual principles of both components in a different (but ultimately elegant) fashion were we able to finally overcome these difficulties. There may well be a lesson for us here, that a simple, additive approach to modeling is not the way forward, but instead the challenge will be to construct an appropriate framework that allows us to make use of tried and tested ideas in new ways.

The architecture that we finish with as a result of our investigations is the one shown in Fig. 6. This is the autoassociative MKM–APECS hybrid given in the bottom panel of Fig. 1, expressed in its most elegant form. It simply requires an additional set of reciprocal connections from the hidden layer back to the input layer to function and delivers the stimulus–stimulus based representational capabilities of MKM along with the ability to learn stimulus–outcome associations in the way that APECS can. The computations required remain exactly as described already, and so there is a need to distinguish between external input and "internal" input to the input units, but, with this proviso, the system is a simple one and the architecture is straightforward to implement.

We have been able to show that this hybrid model is capable of replicating some of the notable successes of MKM (latent inhibition and perceptual learning, the Espinet effect) and APECS (first- and second-order RR), with the advantage that these phenomena can now be explained by means of one model rather than two. The challenge for us now is to both predict new phenomena by means of new simulations and to develop the model so as to accommodate the effects of stimulus history (we intend to add an implementation of Mackintosh's [1975] alpha model to this in much the way that Suret and McLaren [2005] did to MKM; see McLaren & Dickinson, 1990, for a discussion of this point) and to make it real-time rather than trial driven, as it is

at present. Another area that will need investigation is whether the representational approach taken in McLaren and Mackintosh (2002) and further developed in Livesey and McLaren (2007, 2009, 2011) is required to allow us to explain the representation of stimuli that show dimensional variation leading to effects such as peak shift. We are currently unsure how this investigation will turn out. On the one hand, the computational techniques used in McLaren and Mackintosh (2002) allow for a relatively assumption-free approach to the issue of how to represent dimensional variation and allow us to maximize the representational resources available to an elemental model. But we are conscious that some of this machinery may be redundant when taken in the context of a hybrid model that has configural capabilities. A careful assessment will be needed to arrive at a solution to this issue that will preserve representational power, forestall unwanted interactions between model components, and meet our criteria for elegance and simplicity of integration that have proven useful guides in getting us to this point.

Acknowledgements This research was supported by an Exeter Graduate Fellowship to C. L. Forrest, an Exeter Associate Research Fellowship to R.P. McLaren, and an ESRC award to I.P.L. McLaren.

References

- Aitken, M. R. F., Bennett, C. H., McLaren, I. P. L., & Mackintosh, N. J. (1996). Perceptual differentiation during categorization learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 43–50.
- Attneave, F. (1957). Transfer of experience with a class-schema to identification-learning of patterns and shapes. *Journal of Experimental Psychology*, 54, 81–88.
- Bennett, C. H., Scatill, V. L., Griffiths, D. P., & Mackintosh, N. J. (1999). The role of inhibitory associations in perceptual learning. *Learning & Behavior*, 27, 333–345.
- Brandon, S. E., Vogel, E. H., & Wagner, A. R. (2000). A componential view of configural cues in generalization and discrimination in Pavlovian conditioning. *Behavioral Brain Research*, 110, 67–72.
- Chapman, G. B. (1991). Trial-order affects cue interaction in contingency judgement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 837–854.
- De Houwer, J., & Beckers, T. (2002). 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.
- Dwyer, D., Mackintosh, N. J., & Boakes, R. A. (1998). Simultaneous activation of the representations of absent cues results in strengthening of an excitatory association between them. *Journal of Experimental Psychology: Animal Behavior Processes*, 24, 1–9.
- Espinet, A., Iraola, J. A., Bennett, C. H., & Mackintosh, N. J. (1995). Inhibitory associations between neutral stimuli in flavor-aversion conditioning. *Animal Learning & Behavior*, 23, 361–368.
- Estes, W. K. (1959). The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 2, pp. 380–491). New York: McGraw-Hill.

- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.
- Gibson, E. J., Walk, R. D., Pick, H. L., & Tighe, T. J. (1958). The effect of prolonged exposure to visual patterns on learning to discriminate similar and different patterns. *Journal of Comparative and Physiological Psychology*, 51, 584–587.
- Graham, S. (1999). Retrospective revaluation and inhibitory associations: Does perceptual learning modulate our perception of the contingencies between events? *Quarterly Journal of Experimental Psychology*, 52B, 159–185.
- Hall, G. (1980). Exposure learning in animals. *Psychological Bulletin*, 88, 535–550.
- Hall, G. (2003). Learned changes in the sensitivity of stimulus representations: Associative and nonassociative mechanisms. *Quarterly Journal of Experimental Psychology*, 56B, 43–55.
- Hall, G. (2009). Perceptual learning in human and nonhuman animals: A search for common ground. *Learning & Behavior*, 37, 133–140.
- Hall, G., Blair, C. A. J., & Artigas, A. A. (2006). Associative activation of stimulus representations restores lost salience: Implications for perceptual learning. *Journal of Experimental Psychology. Animal Behavior Processes*, 32, 145–155.
- Hall, G., & Rodriguez, G. (2009). Factors determining the effects of associative activation on habituation. *Journal of Experimental Psychology. Animal Behavior Processes*, 35, 266–270.
- Harris, J. A. (2006). Elemental representations of stimuli in associative learning. *Psychological Review*, 113, 584–605.
- Honey, R. C. (2000). The Experimental Psychology Society Prize Lecture: Associative priming in Pavlovian conditioning. *Quarterly Journal of Experimental Psychology*, 53B, 1–23.
- Honey, R. C., & Ward-Robinson, J. (2002). Acquired equivalence and distinctiveness of cues: I. Exploring a neural network approach. *Journal of Experimental Psychology. Animal Behavior Processes*, 28, 378–387.
- Larkin, M. J. W., Aitken, M. R. F., & Dickinson, A. (1998). Retrospective revaluation of causal judgements under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1331–1352.
- Le Pelley, M. E., Cutler, D. L., & McLaren, I. P. L. (2000). Retrospective effects in human causality judgment. *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 782–787). Hillsdale: Erlbaum.
- Le Pelley, M. E., & McLaren, I. P. L. (2001). Retrospective revaluation in humans: Learning or memory? *Quarterly Journal of Experimental Psychology*, 54B, 311–352.
- Livesey, E. J., & McLaren, I. P. L. (2007). Elemental associability changes in human discrimination learning. *Journal of Experimental Psychology. Animal Behavior Processes*, 33, 148–159.
- Livesey, E. J., & McLaren, I. P. L. (2009). Discrimination and generalization along a simple dimension: Peak shift and rule-governed responding. *Journal of Experimental Psychology. Animal Behavior Processes*, 35, 554–565.
- Livesey, E. J., & McLaren, I. P. L. (2011). An elemental model of associative learning and memory. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 153–172). Cambridge: Cambridge University Press.
- Lovibond, P. F., Preston, G. C., & Mackintosh, N. J. (1984). Context specificity of conditioning, extinction and latent inhibition. *Journal of Experimental Psychology. Animal Behavior Processes*, 10, 360–375.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276–298.
- 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.
- Matzel, L. D., Shuster, K., & Miller, R. R. (1987). Covariation in conditioned response strength between stimuli trained in compound. *Animal Learning & Behavior*, 15, 439–447.
- McLaren, I. P. L. (1990). *Representation and association*. Unpublished thesis Cambridge University.
- McLaren, I. P. L. (1993). APECS: A solution to the sequential learning problem. In *Proceedings of the XVth Annual Convention of the Cognitive Science Society* (pp. 717–722). New York: Psychology Press.
- McLaren, I. P. L. (1994). Representation development in associative systems. In J. A. Hogan & J. J. Bolhuis (Eds.), *Causal mechanisms of behavioural development* (pp. 377–402). Cambridge: Cambridge University Press.
- McLaren, I. P. L. (1997). Categorisation and perceptual learning: An analogue of the face inversion effect. *Quarterly Journal of Experimental Psychology*, 50A, 257–273.
- McLaren, I. P. L. (2011). APECS: An adaptively parameterised model of associative learning and memory. In E. Alonso & E. Mondragón (Eds.), *Computational neuroscience for advancing artificial intelligence: Models, methods and applications* (pp. 145–164). Hershey: IGI Global.
- McLaren, I. P. L., Bennett, C., Plaisted, K., Aitken, M., & Mackintosh, N. J. (1994a). Latent inhibition, context specificity, and context familiarity. *Quarterly Journal of Experimental Psychology*, 47B, 387–400.
- McLaren, I. P. L., & Dickinson, A. (1990). The conditioning connection. *Philosophical Transactions of the Royal Society B*, 329, 179–186.
- McLaren, I. P. L., Kaye, H., & Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 102–130). Oxford: Oxford University Press.
- McLaren, I. P. L., Leavers, H. L., & Mackintosh, N. J. (1994b). Recognition, categorisation and perceptual learning. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 889–909). Cambridge: MIT Press.
- McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, 38, 211–246.
- McLaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior*, 30, 177–200.
- 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.
- Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, 117(3), 363–386.
- Mitchell, C., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 237–242.
- Mundy, M. E., Honey, R. C., & Dwyer, D. M. (2007). Simultaneous presentation of similar stimuli produces perceptual learning in human picture processing. *Journal of Experimental Psychology. Animal Behavior Processes*, 33, 124–138.
- Oswalt, R. M. (1972). Relationship between level of visual pattern difficulty during rearing and subsequent discrimination in rats. *Journal of Comparative and Physiological Psychology*, 81, 122–125.
- Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94, 61–73.
- 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 conditioning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532–552.
- Prados, J., Chamizo, V. D., & Mackintosh, N. J. (1999). Latent inhibition and perceptual learning in a swimming pool navigation task. *Journal of Experimental Psychology: Animal Behavior Processes*, 25, 37–44.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Rodrigo, T., Chamizo, V. D., McLaren, I. P. L., & Mackintosh, N. J. (1994). Effects of pre-exposure to the same or different pattern of extra-maze cues on subsequent extra-maze discrimination. *Quarterly Journal of Experimental Psychology*, 47B, 15–26.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing (Vol. 1)* (pp. 318–362). Cambridge: MIT Press, Bradford Books.
- Saksida, L. M. (1999). Effects of similarity and experience on discrimination learning: A nonassociative connectionist model of perceptual learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 25, 308–323.
- Sansa, J., Chamizo, V. D., & Mackintosh, N. J. (1996). Aprendizaje perceptivo en discriminaciones espaciales. *Psicología*, 17, 279–295.
- Suret, M. B., & McLaren, I. P. L. (2005). Elemental representation and associability: An integrated model. In A. J. Wills (Ed.), *New directions in human associative learning*. Mahwah: Erlbaum.
- Trobalon, J. B., Sansa, J., Chamizo, V. D., & Mackintosh, N. J. (1991). Perceptual learning in maze discriminations. *Quarterly Journal of Experimental Psychology*, 43B, 389–402.
- 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. (1978). Expectancies and the priming of STM. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 177–209). Hillsdale: Erlbaum.
- Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behaviour. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale: Erlbaum.
- Wagner, A. R., & Brandon, S. E. (2001). A componential theory of Pavlovian conditioning. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 23–64). Mahwah: Erlbaum.
- Ward-Robinson, J., & Hall, G. (1996). Backward sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 395–404.
- Wills, A. J., Suret, M. B., & McLaren, I. P. L. (2004). The role of category structure in determining the effects of stimulus preexposure on categorization accuracy. *Quarterly Journal of Experimental Psychology*, 57B, 79–88.