Associations and Propositions:

The case for a dual-process account of learning in humans

I.P.L. McLaren[1], C.L.D. Forrest [1] R.P. McLaren[1], F.W. Jones[1,2], M.R.F. Aitken[3], and N.J. Mackintosh[3]
[1]University of Exeter, [2]Canterbury Christ Church University and [3]University of Cambridge

Citation: I.P.L. McLaren, C.L.D. Forrest, R.P. McLaren, F.W. Jones, M.R.F. Aitken, N.J. Mackintosh (2104). Associations and Propositions: The case for a dual-process account of learning in humans. *Neurobiology of Learning and Memory*, 108, 185-95.

Abstract

We review evidence that supports the conclusion that people can and do learn in two distinct ways – one associative, the other propositional. No one disputes that we solve problems by testing hypotheses and inducing underlying rules, so the issue amounts to deciding whether there is evidence that we (and other animals) also rely on a simpler, associative system, that detects the frequency of occurrence of different events in our environment and the contingencies between them. There is neuroscientific evidence that associative learning occurs in at least some animals (e.g., *Aplysia californica*), so it must be the case that associative learning has evolved. Since both associative and propositional theories can in principle account for many instances of successful learning, the problem is then to show that there are at least some cases where the two classes of theory predict different outcomes. We offer a demonstration of cue competition effects in humans under incidental conditions as evidence against the argument that all such effects are based on cognitive inference. The latter supposition would imply that if the necessary information is unavailable to inference then no cue competition should occur. We then discuss the case of unblocking by reinforcer omission, where associative theory predicts an irrational solution to the problem, and consider the phenomenon of the Perruchet effect, in which conscious expectancy and conditioned response dissociate. Further discussion makes use of evidence that people will sometimes provide one solution to a problem when it is presented to them in summary form, and another when they are presented in rapid succession with trial-by trial information. We also demonstrate that people trained on a discrimination may show a peak shift (predicted by associative theory), but given the time and opportunity to detect the relationships between S+ and S-, show rule-based behaviour instead. Finally, we conclude by presenting evidence that research on individual differences suggests that variation in intelligence and explicit problem solving ability are quite unrelated to variation in implicit (associative) learning, and briefly consider the computational implications of our argument, by asking how both associative and propositional processes can be accommodated within a single framework for cognition.

Introduction

In 1994 McLaren, Green and Mackintosh published a paper on "Animal Learning and the Implicit/Explicit Distinction" as a chapter in the book "Implicit and Explicit Learning of Languages" edited by N.C. Ellis. In it we argued for a dual process account of learning in humans, but against the use of awareness as a criterion for distinguishing between implicit and explicit learning and memory. Instead, we proposed a distinction between associative and cognitive processes, and took a comparative stance that equated associative processes across species. We characterised associative learning as that which involved the establishment of links between representations (and here a representation can be as simple as the activation of a specific unit or neurone though often it will correspond to a pattern of activity over units), and cognitive processes were seen as allowing the induction and application of rules or propositions in a flexible manner not available to associative processes. Executive control was cited as possibly the purest example of the application of rule-based processing. We took the view that the cognitive system dealt in propositional knowledge and would therefore have beliefs, and reasons for those beliefs that could be verbalized, whereas the associative system would extract the statistical regularities from experience. Subsequently, Mackintosh (1997), starting from a rather different position, also argued for the importance of associative processes in human learning.

We are certainly not alone in suggesting that people may have two rather different sets of processes for learning and problem solving. Related ideas, albeit couched in different terminology, have been proposed by, among others, Evans (2008), Kahneman (2003, 2012), Stanovich and West (2000), and in the related area of memory by Jacoby (1991). But they have been strongly resisted by some. For example, Mitchell, De Houwer and Lovibond (2009) and Lovibond and Shanks (2002), have argued that the case for a dual process account is not proven (at best), or non-existent (at worst). Both have suggested that all human learning is to be explained by a single process propositional account, there being no evidence for associative processes. We believe there is a need for proponents of a dual process account of human learning to respond to these analyses, and, in particular, to make the case for associative learning in humans. Our plan in this paper is to revisit the arguments we made in 1994 and 1997, and see how research to date has or has not substantiated the analyses we offered then. In the 1994 paper we discussed Pavlovian conditioning, peak shift, and the Overtraining Reversal Effect (ORE) to illustrate the application of associative processes, and in the 1997 paper additional evidence from studies of peak shift was used to support his argument. Here we will consider Pavlovian conditioning again (this time including the Perruchet effect), look at more recent evidence on peak shift and the ORE in humans to see if it parallels that in other animals. We will also bring to bear recent research on individual differences that offers a new perspective on the dual process debate. In what follows, we construct our argument in terms of making the case for associative learning, as the case for propositional learning naturally emerges as a counterpoint to this theme. Whilst doing this we are conscious of the need to exercise great care in drawing conclusions based on simple parallels between the animal and human literature on learning. We take the view that if we believe in dual process accounts of learning, then this places a responsibility on us to ensure that when we ascribe an effect to associative learning we take some care to show that it is not driven by means of rules or propositions. At the end of this paper,

we conclude with a few thoughts on how associative and propositional processes might co-exist and jointly contribute to human learning and behaviour.

The case for association– neural mechanisms

We start by asking why we might believe that there is such a thing as associative learning in animals and humans. The idea of an association as the basis for learning has its origins in the writings of philosophers such as Hartley and Hume (even Aristotle). Within experimental psychology, its popularity may be because it provides a simple and straightforward account of the basic phenomena of Pavlovian and instrumental conditioning. This in itself is not a sufficient reason to take it as the explanation for these phenomena; if we allow (as we do) that humans can learn by forming propositions, why could this not also be the case for other animals? A first answer is that it is demonstrably the case that for some animals learning is associative, because it has been shown to be so at a neural level. Neuroscientific studies of the substrate for learning have shown that in at least one invertebrate, *Aplysia californica*, learning is dependent on link formation. Kandel and his colleagues have shown in a series of papers (e.g., Hawkins, Abrams, Carew and Kandel, 1983; Hawkins, Carew and Kandel, 1986; and see Hawkins, 1989 for a discussion) that learning in *Aplysia* takes place by means of either enhancing or degrading the effectiveness of synaptic connections between sensory and motor neurones. Here then, the activities of individual neurones are the representations and the synaptic connections between them are the links. *Aplysia* has also been shown to display many of the standard classical conditioning phenomena (acquisition, extinction, differential conditioning, ability to track contingency, blocking, conditional discrimination; see Hawkins et al, 1986; Colwill, 1985; Colwill, Absher and Roberts, 1988); thus the case for there being an associative basis for this type of learning in at least one animal is very strong indeed. Our first point, then, is that associative learning exists. The corollary must be that if it exists in *Aplysia*, it would be very surprising if it did not also exist in other animals, and not wholly surprising to find it in the laboratory rat or pigeon, (and perhaps even in humans for that matter). The demonstration of the existence of associative learning in *Aplysia* nullifies the single process argument that follows from conceding the existence of propositional learning in humans. It would seem that there are at least two types of process supporting learning in the world. The question becomes whether we are endowed with both of them or not, rather than whether either of them exists.

At this juncture it is worth stating that the evidence also points to an error-correcting form of associative learning in humans and other animals. We have already mentioned that *Aplysia* displays phenomena (e.g. blocking) that are often taken as evidence for error-correction. Whilst the basic neural mechanism for learning in *Aplysia* does not, at first sight, seem to be of an error-correcting nature, both Hawkins et al (1989) and McLaren (1989) have shown that an assembly of the type of neurones found in *Aplysia* will function as an error-correcting computational system. There can be little doubt that the eyeblink response in the rabbit (Thompson, 1965, 1989) is controlled by an error-correcting circuit, and there is also evidence from Wolfram Schultz's laboratory that neural correlates of prediction error can be found in primates (see Schultz, Dayan and Montague, 1997). Behavioral phenomena such as superconditioning and over-expectation (see Miller, Barnet and Grahame, 1995 for a review) support the case for associative learning in animals being driven by some form of error correction. Rescorla and Wagner (1972) have provided the benchmark

model for associative learning based on error correction, but it is worth mentioning that other models such as those of Mackintosh (1975), Pearce and Hall (1980) and Pearce and Mackintosh (2010) that are not often thought of as error-correcting do, in fact, depend on something akin to that process. Adjustment of associability in these models relies on a summed error-term, and this associability parameter controls learning (see McLaren and Dickinson, 1990, for a discussion of this issue, and how associability might be implemented neurally). Even Pearce's (1987) configural model is more error-driven than is commonly considered to be the case: It uses a simple, single error term to control learning between a configural CS and US, but the fact that it allows generalisation from other trained configurations to contribute to that error term effectively allows a summed error term to influence learning. Each of these models is able to account for a wide range of the effects found in the associative learning literature. Thus there is considerable evidence, both empirical and theoretical, for learning being controlled (or at least influenced) by prediction error.

The case for association– Pavlovian conditioning

A second reason for believing that animal learning is associative is that it can be shown to be irrational in some circumstances, and produce behaviour that is not at all what one would expect from a propositional system. It must of course be the case that if it is to be adaptive and enable the animal to survive, in most circumstances the outcome of associative learning should parallel that to be expected from a rational, propositional system. Thus, the ability to track contingencies (Rescorla, 1967; Wasserman, Chatlosh and Neunaber, 1983) is exactly what you would expect from either system. The phenomenon of blocking (Kamin, 1968) is also susceptible to explanation in associative or propositional terms (see Beckers, Miller, De Houwer and Urushihara, 2006, for an example of the propositional approach and Haselgrove, 2010 for the associative response to their case). We will consider the example of overshadowing in more detail here, as this is another example of cue competition where both classes of explanation can apply. Overshadowing occurs if two quite distinct, equally salient cues, A and B, are trained in compound to predict a US, and then responding to either A or B is less than would be seen if that cue had been trained in isolation. If one cue, say A, is more salient than B, then it tends to dominate learning when they are trained in compound, and relatively little accrues to B (see Mackintosh, 1976, for just such an experiment). This result is easily explained by associative theories. According to the Rescorla-Wagner model, the two cues, A and B, share the associative strength to the outcome between them in proportion to their relative salience. Pearce's configural theory arrives at the same result by a different route, arguing that learning about AB generalises only imperfectly to A or B, again to an extent determined by the relative salience of the cue. Associative theories, then, provide good explanations of cue competition phenomena in animals.

Equally, however, there is no doubt that the propositional approach can explain overshadowing, by taking the view that the participants in the experiment are using cognitive inference based on a heuristic of the type "if there are two cues predicting the outcome, then credit for this prediction must be shared between them according to their salience". For example, this heuristic can be used to explain the results of allergy prediction paradigms such as in Le Pelley and McLaren (2001), where a combination of two foods, A and B, predict an allergic reaction in a hypothetical patient, "Mr. X". The result is that the ratings for A and B are less than that for control cues trained on their own to predict the same outcome.

If the results of such experiments are equally well explained by either associative or propositional accounts, how will it be possible to decide between them? In humans, one way may be to use procedures that make it unlikely that participants will be able to employ their propositional system – which we assume has a limited capacity. Le Pelley, Oakeshott and McLaren (2005) argued that using many different trials, presented in a random order, each employing some of a large number of stimuli with different relationships to the available outcomes, should make it hard for participants to keep explicit track of the contingent relationships in the experiment. Earlier, Le Pelley and McLaren (2001) were also at pains to use these conditions (high memory load due to using many cues and trial by trial presentation) for similar reasons, so it seems reasonable to argue that the cue competition effects they observed were associative in origin. But in many other cases, where few cues are used and memory load is low, the rating given may well owe more to cognitive inference than associative learning. One version of this inferential explanation for overshadowing requires that the subject, whether animal or human, know which cues predict which outcomes, and then uses this information to generate the appropriate behaviour. We can characterise this version of the propositional account of overshadowing as reliant on explicit memory as well as learning. This inferential explanation of overshadowing would be supported by the claim that humans do not show cue competition effects under incidental conditions (Jimenez and Vazquez, 2011). Incidental conditions are precisely those that make it likely that participants do not have access to the necessary explicit cue-outcome information required for cognitive inference to be brought to bear and so produce overshadowing. In the absence of such information, no overshadowing is to be expected in this version of the propositional account. The critical question, therefore, is whether it is right to claim that overshadowing is not observed when people are trained under incidental conditions. Our next section offers what we believe is the first evidence for cue competition in humans under incidental conditions.

Cue competition in an incidental learning paradigm

We have already indicated that demonstrations of overshadowing using the allergy prediction paradigm, whilst robust, are susceptible to the complaint that they may be propositionally driven rather than associatively mediated. A second issue is that the stimuli that serve as the CSs in these experiments may be too similar in kind, in that they are both foods. The analogy would be to an animal experiment in which the overshadowing was demonstrated to two tones of different pitch, rather than a tone and a light. The former might give rise to concerns that the two tones when played together interacted in some way so as to change their stimulus quality, and that this interaction was lost when presented individually, so that the reduction in rating that occurred on test could be explained by some change in the perceived stimulus. Clearly no such process would apply when the stimuli were trained alone. It would be better if the two CSs were different in kind so that this type of potential confound could be avoided. Our two classes of cue were chosen to have quite distinct characteristics to avoid this problem. We employed a basic SRT paradigm similar to that of Willingham, Nissen and Bullemer (1989), in which there were two circles that defined two stimulus locations, left and right. At the start of a trial the circles are outlines, then one of them fills, and the corresponding key has to be pressed. Unknown to the participants, in those groups that were given sequential information, there was a 2/3 chance of a trial being predicted by the two preceding trials. The rule was that if the

two preceding trials were both the same, then the next trial was likely to be a left, whereas if they were different, it was likely to be a right, with these response assignments counterbalanced across participants. Thus, the first type of cue was provided by the sequence of locations that occurred / responses required. The second cue type was provided by a colored square that flashed up just before the circle filled in, presented at fixation between the two circles. Participants for whom color information was relevant had a 3/4 chance that the color would predict the response location on half the trials. On the other half of trials different colors were used that were not predictive and so could be used as color control trials. We settled on these parameters for the tasks as a result of extensive piloting and prior work, to ensure that both the sequential information and the color information were capable of supporting learning under incidental conditions (see Jones and McLaren, 2009 for more on the sequences, and Yeates, Jones, Wills, Aitken and McLaren, 2012, 2013 for details on the colour task). Table 1 gives the details of stimulus construction for each group.

Table 1 about here please

Group Dual had both sequence and color information programmed in. Group Color had the same type of color information as Group Dual, and Group Sequence had the same type of sequence information as Group Dual. Group Sequence were still shown a colored square just before the response location was indicated, but the color bore no relation to that location; equally Group Color experienced sequences of trials in just the same way as Group Dual, but there was no predictive relationship between them.

Another point worth raising in this context is that in all the experiments (that we are aware of) on overshadowing in humans that come close to meeting our first two conditions, the comparison has been between CSs trained in compound and tested individually, and a group or groups trained with the individual CSs and then tested. The problem with this procedure is that one group experiences a major change from training to test (the compound group) whereas the other does not. This, on its own, may be enough to depress responding in the compound group if they come to believe that circumstances have changed and deliberately and strategically alter their responses as a consequence (something that seems intuitively less likely to be the case in a rat or a pigeon). Note that this is not the same as a generalization decrement account of overshadowing that would, for example, follow from Pearce's (1987) configural model. It is rather an appeal to a strategic decision based on changing circumstances during the course of the experiment, and we avoided this in our design by making sure that the transition from training to test was unsignalled and unlikely to be noticeable. Thus, all groups experienced a fast-paced sequence of trials cued by a colored square during both training and test, the only difference being the contingencies that applied during test (when all contingencies were 50:50, and sequence and color information were uncorrelated for all groups).

Figure 1 about here please

Figure 1 shows the test results for sequence learning (left panels) and color learning (right panels) on test. The data shown in Figure 1 are the mean difference between trained and untrained sequences or colors in RTs (top) and errors (bottom) for each of the three groups. Higher scores indicate more learning (chance is zero), and, starting with the panels shown left, it is clear that both the Sequence and Dual groups showed

good evidence of sequence learning, whereas there was little evidence of sequence learning in the Color group (as should be the case). Given that the test phase was, in effect, an extinction treatment, the evidence for sustained performance on the basis of what had been learned during training in Group Sequence and Group Dual is noteworthy and implies strong learning of the sequence information available during training. We conclude that Group Dual learned something about the sequence structure in addition to whatever it learned about the ability of the colors to predict the next response location. We now turn to the difference scores obtained by comparing performance for the predictive colors with the control colors for RTs and errors on test and shown on the right of the figure. Unsurprisingly, Group Sequence shows no evidence of having learned anything about the colors on test (there was nothing to learn). Group Dual also shows no evidence of learning about the colors either on the RT or the error measure, but Group Color's RT performance is significant, and is significantly better than that of the other two groups.

It would appear, then, that the Dual group learned about the sequences, but did not learn the color information available to them, even though Group Color shows that this was eminently possible. This is what would be expected if the sequence cues had overshadowed the color cues in the Dual group (but not vice-versa). Thus, it is possible to demonstrate cue competition effects in humans trained under incidental conditions, and the parallel with animals learning under similar conditions would seem to have some validity. Were the participants in this experiment aware of the sequence or color rules (if so, this would undermine our results as evidence that cue competition effects can occur when the basis for cognitive inference is unavailable to participants)? Previous experiments and pilot work suggest that this should not be the case, and post-experiment interviews established that participants were unable to give any accurate information about the sequences, or say which colors were predictive. Crucially, there was no difference between Dual and Color groups in terms of their ability to guess which colors were predictive (44% and 45% respectively), and both values are numerically below chance (50%). Equally, there was no reliable difference in the proportion of participants asserting that the sequences were random (using a conservative criterion in scoring this) in any of the Color, Dual or Sequence groups (50%, 57% and 70%), although numerically the Sequence group was actually more likely to think that their sequences were random. Given this, we can now reject the argument that cue competition in humans is only observed under intentional learning conditions, and its corollary that this is because it relies on cognitive inference to manifest. It certainly does not seem to be the case that explicit cognitive inference is required to generate overshadowing in humans.

Irrational cue competition effects

Further evidence against a propositional account of human learning might be provided by an experiment that passed what we shall term the "participant-as-experimenter test[1]". The essence of this is to imagine what the reaction of participants would be if, instead of running them in the experiment, you were simply to describe its design to them. In many cases, of course, you would expect the same result from both procedures. For example, when you describe a blocking design to people, they will

[1] The " participant-as-experimenter test" is based on a conversation with David Shanks in 1990. We imagine what the reaction of a participant will be if, instead of running them in the experiment, we were to describe its design to them. Shanks pointed out to IPLM that in many cases you would expect the same result from both procedures.

suggest a blocking type result as the most likely outcome, and give reasons why this would be so. If the design is "A+ followed by AB+" then they will likely say that the pre-training to A means that when A and B occur in compound A already predicts the outcome so B is redundant and consequently they will give it less credit or be less confident about attributing the outcome to B. In order to differentiate between associative and propositional accounts and pass our test the experiment has to be such as to produce divergent results under these two conditions (i.e. actually experiencing the contingencies in the experiment or being asked to predict the results from the design). Are there more complex cue competition designs that can create problems for the propositional account by passing the participant-as-experimenter test? One possibility is unblocking (Dickinson, Hall and Mackintosh, 1976, Dickinson and Mackintosh, 1979). The design of the first of these experiments is shown in Table 2.

Table 2 about here please

After training rats that A predicts a shock (A-Sh), and then that AB predicts two shocks, one after the other (AB-Sh-Sh), testing B now reveals good conditioning compared to a blocking control group (A-Sh-Sh, AB-Sh-Sh). This may not seem surprising, as the total amount of US experienced has increased in the experimental group, but the same effect can also be obtained if initial training with A-Sh-Sh is followed by AB-Sh. This experimental group also shows stronger conditioning to B compared to a control group experiencing A-Sh followed by AB-Sh. In these circumstances the unblocking effect is not so easily explained by the propositional approach, which would tend to reason that if A and B in compound are followed by one less shock than A on its own, then clearly B is responsible. Thus, rational inference would suggest that B should, if anything, become an inhibitor for shock, and certainly not show an unblocking effect. It is hard to see how a propositional account would have predicted this result, but the Mackintosh (1975) theory did, and was the inspiration for this experiment. McLaren and Dickinson (1990) discuss the neural mechanisms needed to produce this result and conclude that an associability parameter of the type envisaged in Mackintosh (1975) that is specific to both the CS and to the type of outcome it is paired with is required. It is worth noting that Le Pelley, Oakeshott and McLaren (2005) have produced an analogous result in humans using an allergy prediction paradigm. They also demonstrated that if humans were given the design in a manner similar to that shown in Table 2 so that cognitive inference was engaged, then no unblocking was observed in the condition analogous to Dickinson et al's experiment. Thus, where participants were able to use the propositional system to solve the problem, they did not show an irrational unblocking effect (as we would expect), but where they had to rely on associative processes, they did. Another set of experiments that support this thesis using a similar manipulation (though this time with second order retrospective revaluation as the effect in question) can be found in McLaren, Forrest and McLaren (2012).

The Perruchet effect

Our final example taken from simple Pavlovian conditioning is the Perruchet effect (Perruchet, 1985) in which conscious expectancy dissociates from conditioned responding in humans in a partial reinforcement procedure. The original demonstration used an air puff paradigm, and has been replicated with additional controls by Weidemann, Tangen, Lovibond and Mitchell, 2009), but here we will focus on a recent demonstration by McAndrew, Jones, McLaren and McLaren (2012)

that used an electrodermal paradigm. In their procedure these authors paired a single visual CS with an electrodermal shock 50% of the time. As a consequence, there were times when a run of CS presentations would be followed by shock, and times when a run of extinction trials would occur. On each trial they recorded the skin conductance response (SCR) to the CS, and also asked participants to rate their expectancy (during the CS presentation) of a shock occurring on that trial (if it did occur – it would occur in the last 500 msec of the CS). The results were clear: expectancy of shock increased over a run of extinction trials and decreased over a run of reinforced trials. But the SCR (i.e. the CR to the CS) showed the opposite pattern, decreasing over a run of extinction trials and increasing over a run of reinforced trials. This SCR pattern is exactly what associative theory would predict, and cannot easily be explained by the expectancy data: if learning is due to the formation of a proposition that the CS is followed by shock, a stronger expectation should result in a stronger CR. Lovibond and Shanks (2002) have suggested that this is the explanation for differential conditioning using these electrodermal procedures, so to deny this would undermine their explanation for a large corpus of data on differential electrodermal conditioning. We are left with the conclusion that dual processes are needed to explain this effect, associative processes for the SCR pattern and propositional processes to generate the expectancy ratings. Mitchell et al (2009) have conceded that this phenomenon is one of the best pieces of evidence for a dual process account of learning.

The case for association– discrimination learning

The Overtraining Reversal Effect

Experiments on discrimination learning provide further examples of phenomena not predicted by a propositional analysis. If rats are trained and then reversed on a visual discrimination problem (A+, B-, followed by B+, A-), they typically learn the reversal more rapidly if overtrained on the initial discrimination – the so-called overtraining reversal effect (ORE). It is difficult to see how a rational analysis could predict that additional training on a discrimination would actually make it easier to reverse that discrimination. The ORE is a well-established effect in animal learning (see Mackintosh, 1969; Sutherland and Mackintosh, 1971), and a version of it has been demonstrated in humans by Suret and McLaren (2003). They used a dimension (actually four of them) constructed by morphing from one face (face 1) to another (face 11) in 10 stages. The dimension used is illustrated at the bottom of Figure 2. It was essentially a transfer along a continuum (TAC, see Lawrence, 1952) experiment in which pre-training on an easy problem (morph 3 vs. 9) facilitated later learning of a hard problem (morph 5 vs. 7). Suret and McLaren showed (see graph in Figure 2) that extended (but not relatively brief) pre-training on 3 vs. 9 facilitated learning of 5 vs. 7 even when the response assignments were reversed at transfer (i.e. 3+ vs. 9- was followed by 5- vs. 7+). Learning in this condition was superior to that observed when participants were trained on 5 vs. 7 for the same total amount of time.

Figure 2 about here please

This result also passes the participant-as-experimenter test referred to earlier. We acknowledge that a basic transfer along a continuum experiment, where training is on the easy problem and then participants are transferred to the hard problem (without reversal of the response assignments) seems like a plausible method for enhancing acquisition of the hard discrimination (and indeed both humans and other animals

show such an effect). This makes it hard to distinguish between an associative and a propositional account with this design, because human participants at least may have become aware of the experimental design, and will simply deliver the expected (by them) result. But, when the intuition based on the design is *not* reflected in the behavioral result, then the case for a propositional explanation is weakened. The reversal version of this experiment involves a design in which the intuition is that extended training will make learning the reversal *harder* for the participant, which turns out not to be the case, and so passes our test. It is not easy to see how this could be predicted on a propositional basis. For example, it might be argued that extra training on the original 3+ vs. 9- discrimination helped participants learn and remember the trained stimuli and the required responses. Because this was the case, they then found it easier to a) transfer to 5 vs 7 because they could see that 5 was more like 3 and 7 more like 9, and b) swap the response mappings. But Suret and McLaren (2003) were able to demonstrate transfer from easy to hard (without reversing the response mappings) after standard (not extended) training, training that was ineffective in producing rapid reversal. Thus, based on these data and this analysis, participants were already able to equate 5 with 3 and 7 with 9 perfectly well, leaving their inability to reverse the mapping from stimulus to response as something of a mystery. An associative account explains why these two sets of results can be obtained, by postulating that only overtraining allows the associability of the trained cues to get to the point where it will be maintained during reversal, and so deliver an advantage for the easy condition.

Figure 3 about here please

Transposition and Peak Shift

Our next example illustrating the need to posit associative as opposed to cognitive processes in learning comes from consideration of peak shift and the related phenomenon of transposition. Transposition was first studied by Kohler (1917), who trained animals to select the brighter of two stimuli, and then tested them with their original S+ and an even brighter stimulus (the gray rectangles depicted in Figure 3, they would be trained on those labelled + and -, then tested on + vs. the rectangle on its immediate left). He took their choice of the even brighter stimulus on this test as evidence that they had solved the initial discrimination by responding to the relationship (brighter, darker) between S+ and S-, rather than to their absolute values. Spence (1937), however, showed how interacting gradients of generalization round S+ and S- could in principle predict transposition (see Figure 3), and the demonstration of a peak shift in pigeons by Hanson (1957), seemed to prove the plausibility of Spence's analysis. Figure 3 illustrates this point. Spence argued that training on an S+ and S- created two gradients, one of excitation shown as solid black in the figure, one of inhibition shown in gray. These gradients could correspond to the two rectangles labelled as + and – in the figure if we wished to apply this analysis to Kohler's experiment, or to the corresponding positions on a color continuum if we wished to apply it to Hanson's experiment. Responding is governed by the resultant gradient obtained by subtracting inhibition from excitation. Because the difference between the two gradients is maximal for a luminance or color that is (in this diagram) to the left of S+ (i.e. a point on the continuum not at S+ and further away

from S- than S+), then we would expect maximal responding here as in Hanson's experiment, and the brighter stimulus to be chosen over S+ as in Kohler's.

The critical difference between transposition due to relational learning and the peak shift is that, as can be seen in Figure 3, Spence's interacting gradients predict that although responding may be greater to a stimulus just beyond S+ than to S+ itself, if you move further away from the original S+ and S-, responding will eventually fall off, with the consequence that responding will be greater to the stimulus closer to S+ than to the further stimulus. Relational learning might predict some eventual decline in responding as you move further away from the training stimuli, but it could never predict this reversal. Wills and Mackintosh (1999) were able to show that their pigeons' and people's behaviour followed the pattern predicted by Spence's analysis.

Table 3 about here please

Wills and Mackintosh (1998) used an artificial dimension (see Table 3) in which the constituent parts (icons) were constrained to behave according to the principles of Blough's (1975) elemental theory, to demonstrate peak shift. Thus, after training on the S+ and S- designated in Table 3, they were able to demonstrate that responding to Near+ was greater than to S+, but that it had declined by the time Far+ was reached. Their success in doing this constitutes an existence proof that the theory is capable of generating a peak shift, and makes plausible the assertion that peak shift on a natural dimension is due to this type of representation coupled with associative learning. It is difficult to give an account of these results based on any kind of propositional analysis, because the stimuli simply do not easily lend themselves to rule induction. In a similar vein researchers have also used face dimensions (like those used above for ORE effects) to demonstrate peak shift in humans (McLaren and Mackintosh, 2002; Lewis and Johnson, 1999; Spetch, Cheng and Clifford, 2004).

The evidence that discrimination between two values on a dimension can be either associatively-mediated or rule-based has accumulated since our discussion in 1994 and 1997. Jones and McLaren (1999), using patches of green colour which varied along a dimension of brightness, showed that relatively short training led to discrimination that produced the typical peak shift pattern in humans, whereas lengthier training resulted in behaviour indicative of rule use: i.e. increasingly accurate performance with increasing distance from the decision boundary. They also showed that the participants producing this pattern were aware of the rule in play (they could explain the response mappings required for light vs. dark shades of green), whereas those producing peak shift were not. We have recently replicated this result using a technique based on that employed in Livesey and McLaren (2009), whereby participants are classified as rule users or not by means of a post-experimental interview. It seems reasonable to assume that if participants have recognised this simple rule they would be able to verbalise it (and many participants did). Figure 4 shows the results of applying this technique with the stimuli originally used by Jones and McLaren (1999).

Figure 4 about here please

Overall performance on this task was relatively poor as can be seen from the graph (the training stimuli were very hard to discriminate), but the performance of rule users was respectable, and significantly better than that of those participants who did not induce the rule. The rule users showed a monotonically increasing trend from the Training stimuli (which are fairly similar to one another) to the Near stimuli (which are further from the decision criterion for light vs. dark and so less confusable), and then again to the Far stimuli (which are quite bright and quite dark respectively). These people stated that the rule was "left key for dark green, right key for bright green" (or v.v. as this was counterbalanced), and it is reasonable to assume that this knowledge was responsible for their superior performance and the transposition effect they demonstrated. To explain the monotonic trend, we simply have to point out that distinguishing between bright and dark greens becomes progressively easier the brighter and darker they are. The non-rule users showed a quite different pattern. Their overall performance was poor, with only performance on the Near stimuli being significantly above chance. Performance on these stimuli was reliably superior to that on either the Far or Training stimuli. Thus these participants exhibited a classic peak shift, which we take as a sign of associative processing, by analogy with the pattern of responding obtained with pigeons in similar tasks. Livesey and McLaren (2009) were able to track the transition from associatively-mediated performance as indexed by peak shift, to rule-based performance in their studies, and our results confirm the correlation between knowing the rule that applies to the discrimination and the pattern of responding obtained on the task itself (which seems best explained by propositional processing in the rule use case and associative processes in the other). A propositional theorist might offer the thought that some other rule is being used by those we classify as non-rule users, one that is not picked up by our questionnaire. If this is the case, we have been unable to find any evidence for any consistent verbal response (or any response at all) from these participants, which is somewhat surprising given the consistency of report available from those we classify as rule users if the same propositional system is deemed to be at work in both groups.

Aitken (1996) relied on a different technique to dissociate peak shift from relational transposition in human participants. The difficult discrimination required was one between short, fat wedge-shaped stimuli and long thin ones. When these stimuli were presented one at a time with ample time for responding, and feedback given after every trial, testing with stimuli progressively further removed from the training stimuli revealed progressively more accurate performance, with no sign of any decline with the furthest stimuli, let alone any reversal. He took this as evidence of transposition and relational learning. In order to demonstrate a peak shift, the procedure was completely changed. On each trial, one of two circles on the screen was briefly illuminated, and the participants' task was to respond as rapidly as possible to the illuminated circle. A variety of different stimuli were also shown on the screen on every trial, most of them wholly irrelevant to the task at hand. But on some trials, one of these stimuli was either a short, fat wedge or a long, thin wedge, and these reliably predicted which circle would be illuminated at the end of the trial. That participants had learned about these contingencies was evident from the fact that their reaction times on these trials were faster than on those when there was no predictive stimulus present, but only a very few participants were able to articulate these contingencies when asked after the experiment. When tested with stimuli progressively further away from the wedges used in training, the 'unaware' participants showed a classic peak shift – initially responding more accurately to test

stimuli not too far away from the training stimuli, but significantly less accurately as the test stimuli moved even further away. This decline in performance to the furthest stimuli was *not* seen in the small number of participants who were able to articulate the contingencies during the course of training on the initial discrimination.

The fact that Jones and McLaren (1999) and Aitken (1996) used quite different procedures to produce either a peak shift or relational transposition is important, for it surely makes any artifactual explanation of their results markedly less plausible. We conclude that these data on discrimination, and other data like them, support the case for two dissociable sets of processes involved in learning, one associative, the other rule-based.

The case for association - individual differences

For our final domain of evidence we consider studies of a correlational nature that exploit the natural variation in human abilities. If we take an IQ score as a (no doubt imperfect) measure of how well someone thinks, reasons or solves complex problems using rule-based processes, then we know that people differ quite widely in the 'efficiency' of their propositional system(s). A final potential way of decoupling an associative from a propositional system would be to see if people also differed in the efficiency of their associative learning, but this was quite independent of their IQ.

IQ scores are at least moderately correlated with a surprisingly wide variety of other measures of human performance – at least some of which might reasonably be thought to depend on an associative system. Although, for example, Jensen (1998) argued that there was only a very modest correlation between measures of IQ and paired associate learning, and although some studies have indeed supported Jensen's claim, others have not. Alexander and Smales (1997) reported a correlation of .56 between a composite measure of paired associate learning and verbal IQ, while Williams and Pearlberg (2006) have also reported correlations of about .50 between one verbal learning task and IQ.

It will come as no surprise that we should suggest that a 'purer' measure of the associative system might be provided by incidental learning tasks. Reber, Walkenfeld and Hernstadt (1991) were the first to demonstrate that performance on one implicit learning task, artificial grammar learning, was unrelated to IQ. They reported a correlation of only .25 (ns) between artificial grammar learning and WAIS IQ in a sample of 20 college students, but at the same time a correlation of .69 between IQ and an explicit series completion task. The two correlations were significantly different. Reber et al.'s results were replicated by McGeorge, et al. (1997) in a larger study of over 100 adults ranging in age from 18 to 77. They reported correlations of .12 and .67 between WAIS IQ and artificial grammar and series completion tasks respectively.

Performance on a second incidental learning task, the serial reaction time task, has also shown a negligible correlation with IQ. In a study of over 400 schoolchildren Feldman, Kerr and Streissguth (1995) reported a correlation of .05 with WISC IQ, and similar results have been reported by Unsworth and Engle (2005), and Kaufman et al (2010).

Correlations between incidental learning and IQ have been small – but not usually zero, and in at least one or two studies have indeed been significant: for example, Danner et al. (2011) reported a significant correlation of .32 between artificial grammar learning and measures of fluid intelligence in a sample of 173 adults – although this compared with a correlation of .86 between the same IQ measures and a problem solving task. Salthouse et al (1999) also found a significant correlation between serial reaction time and IQ.

One explanation of this is that explicit cognition may sometimes contribute to performance on a nominally 'implicit' task (just as it surely contributes to ordinary paired associate learning). Thus in the serial reaction time task, one measure of learning is that participants' reaction times are faster to the sequence they have been trained on than to a novel sequence. But another measure is to ask them to predict what the next item will be. Feldman et al. (1995) found that although the former measure did not correlate with IQ (see above), the latter did ($r = .28$).

A better test of our argument might be to compare performance on various 'implicit' tasks under standard incidental instructions and under 'explicit' instructions, when participants are told in advance that the letter strings or sequence of stimuli have been constructed in accordance with a set of rules and that it is their job to detect these underlying rules, with hints on how to go about this. In a study of some 400 German schoolchildren, Gebauer and Mackintosh (2007) reported nonsignificant correlations ranging from .07 to .11 between different components of IQ and performance under incidental instructions, while under explicit instructions the now significant correlations ranged from .23 to .37. In a rather smaller study, Yang and Li (2012) observed a correlation of .56 between working memory and artificial grammar learning under explicit instructions, but no correlation under incidental instructions.

There have been a few exceptions, but the weight of the evidence strongly suggests that differences in incidental learning are at best only weakly related to conventional measures of intelligence – and certainly less strongly related than are measures of explicit problem solving. Given that experimental psychologists tend to be somewhat skeptical of correlational evidence, it is worth reinforcing our argument by appealing to neuroscience as well. It has long been known from functional imaging studies that explicit and implicit learning engage different regions of the brain, the former resulting in activation of medial temporal regions, the latter striatal-frontal circuits (see, for example, Cohen, Eichenbaum, Deacedo and Corkin, 1985; Knowlton, Mangels and Squire, 1996; for a recent review, see Dennis and Cabeza, 2011). Moreover, in performance on any task there is a negative relationship between the activation of one system and of the other (Poldrack and Packard, 2003). It seems clear that there are two relatively distinct systems in the human brain, one engaged by implicit tasks, the other by explicit tasks.

We would argue, of course, that these findings support the conclusion that there is an important distinction between implicit (which we define as associative) and explicit (cognitive, rule-based) learning. But is implicit learning merely something observed under artificial laboratory conditions, or does it contribute to the way people actually learn about and adapt to the world? In the domain of individual differences in intelligence, Sternberg and his colleagues have been arguing for many years that standard IQ tests fail to capture individual differences in 'practical intelligence',

which is based on 'tacit knowledge' of the world, informal, implicit procedural. If Sternberg is right, implicitly acquired knowledge is of wide importance.

One problem with this suggestion is that different implicit learning tasks are clearly measuring, at least in part, rather distinct abilities or skills. Gebauer and Mackintosh (2007) found only weak correlations between artificial grammar learning, serial reaction time, and process control tasks. By giving a wider variety of tasks and a number of different measures of each, however, Gebauer (2003) was able to show some small but significant relations between different tasks. German schoolchildren (N= 195) were given different measures of IQ, as well as different artificial grammar and process control tasks, and a serial reaction time task. A preliminary factor analysis yielded distinguishable factors for each of these tests, but a second order analysis yielded two superordinate factors: the IQ tests and an explicit problem solving task loaded onto one factor, while the artificial grammar and process control tasks loaded onto the second (serial reaction time loaded onto both). There was, therefore, good evidence for distinct explicit and implicit factors.

The study also obtained information on the students' grades for Mathematics and English (a foreign language for German students). The correlations between the Explicit and Implicit second-order factor scores, and Maths and English grades are shown in Table 4. As can be seen, IQ correlated modestly and significantly with both Maths and English, while the correlation between the Implicit factor and English (but not Maths) was of the same order of magnitude. The correlation of the Implicit factor with English remained significant when Explicit factor scores were partialled out; that with Maths did not.

Table 4 about here please

In a study of English 6th form students, Kaufman et al (2010) also reported significant correlations between implicit learning and GCSE grades in French and German (GCSE are public exams taken at age 16 in English schools). Implicit learning was measured by a probabilistic serial reaction time task, and scores on this task correlated .27 with the students' French grades, and .29 with their German grades. These correlations were unaffected by controlling for students' scores on IQ tests, a working memory task, and explicit associative learning.

As Kaufman (2011) has reviewed, there is good evidence that scores on laboratory-based implicit learning tasks do seem to predict other things of importance. For our purposes, the important message is that such tasks seem to engage an associative system that is independent of the cognitive, rule-based system that people undoubtedly possess.

Cognition and association

The case for associative processing in animals is extremely strong. We feel that there is strong evidence for this type of processing in humans as well, but at the same time would want to acknowledge that there is equally strong evidence for rule-based processes in humans. A number of questions quite naturally arise from these two propositions. What do we mean by allowing two distinctly different types of processing? Are we implying separate systems running independently side-by-side?

How do propositional and associative processes interact? In this section we address these questions to the extent to which we are currently in a position to do so, and point towards the research that is needed to make progress on these issues.

Many theorists have made the simplifying assumption that what we term associative and propositional processes run in parallel and in an independent fashion with both contributing to behavior. An example here would be Jacoby's distinction between automatic and intentional influences on memory as operationalized in the Process Dissociation Procedure (Jacoby, 1991). We have taken this approach with respect to learning in the past (e.g. the SARAH model in Spiegel and McLaren, 2003) and it has proven useful (e.g. in allowing us to explain the quite remarkable parallels between predictions made by the SRN and human sequence learning in Spiegel and McLaren, 2006; whilst still allowing us to explain the quite different pattern of results generated by rule-based generalization in Spiegel and McLaren, 2003), but ultimately we would concede that this state of affairs is very unlikely to reflect reality. Instead, our view is better expressed by asserting that cognition is controlled association. By this we mean that associative processes are fundamental – and when low levels of control are in play then their operation is transparent and their associative nature obvious. Our view is that more cognitive and ultimately rule-based symbolic processing is made possible by employing associative processes within a complex recurrent architecture (the simplest possible example of this would be the SRN, Elman, 1990) that can be controlled (i.e. parameterised and modulated) on the fly (examples of this would be the APECS SRN, Jones, Le Pelley and McLaren, 2002; and the RASRN, Yeates et al., 2013). From this perspective, our distinction between Cognitive and Associative processes is more akin to a continuum, with associative and propositional processes as the endpoints of this continuum and the degree of control the factor that varies across it. Another perspective afforded by this approach emphasises the scale at which computation is considered, in that locally our model of cognition is inevitably associative, but globally symbolic processes emerge from the model. More detail on this approach to the dual system account of learning and cognitive control can be found in a paper currently in preparation by Verbruggen, McLaren and Chambers (available on request).

What does our position predict as to the nature of the interactions between propositional and associative processes? Clearly this becomes a matter of specifying the degree of control involved in any given behavior and then inferring from that the extent to which learning or performance will seem, to the researcher, to be governed by associative or rule-based processes. One corollary of our position is that associative learning will always occur – but whether it is expressed in a simple form is quite a different matter. Given that the degree of control exerted will vary from task to task, we cannot say that propositional processes will always be involved in any learning episode, but it seems likely that, to the extent that sufficient control is in play to engage this mode of processing, then the expression of any simple associative learning will be either masked or suppressed. The prediction, then, is that associative processing will be automatic and "run in the background" even when more controlled, propositional processes dominate, but that the converse will not be the case. The idea that the expression of any associative learning might depend on the degree of cognitive control is consistent with the arguments made by Jones and McLaren (2009) on the basis of their investigation of sequence learning in humans.

Conclusions

It will come as no surprise to the reader to learn that we believe that the case for dual systems supporting learning in humans is strong, and that the evidence accumulated since 1994 increasingly points in this direction. To our mind, there can be little doubt that much (if not all) of learning in infra-humans is associatively based, though we would not rule out at least the precursors of symbolic thought in some species (e.g. chimpanzees). We realize that this conclusion is somewhat at variance with that of recent reviews (e.g. Mitchell et al, 2009) that argue for propositional processes and against associations as a mechanism for learning in humans, but we would argue that these treatments do not adequately deal with the dual-process position. As a critique of pure associationism they are entirely convincing, but once we grant the existence of propositional processes in humans, the arguments deployed in these reviews lose a great deal of their force. Proving that people can learn propositionally is not sufficient to prove that they never learn associatively.

But we would like to go a step further by way of conclusion, and point out that the best way forward now is to start from the standpoint of a dual-process account of learning, and to design experiments explicitly geared to investigating the interaction between the two systems instead of spending our time trying to adduce evidence for one system or the other. We believe that this will lead to the discovery of novel phenomena, a better understanding of human and animal learning and the relationship between them, and quite naturally result in the provision of all the evidence anyone could want for a dual-process account. It would be a mistake to continue fighting old battles if we wish to make progress in understanding human and animal learning, memory, and cognition.

Acknowledgements
The research in this paper was supported by a grant from the ESRC to Ian McLaren and Fergal Jones, and an Exeter Graduate Fellowship to Charlotte Forest.

References

Aitken, M.R.F. (1996). *Peak shift in pigeon and human categorisation.* Unpublished PhD Thesis, University of Cambridge.

Alexander, J. R. M. and Smales, S. (1997). Intelligence, learning and long term memory. *Personality and Individual Differences,* 23, 815–825.

Beckers, T., Miller, R., De Houwer, J. and Urushihara, K. (2006). Reasoning Rats: Forward Blocking in Pavlovian Animal Conditioning Is Sensitive to Constraints of Causal Inference. *Journal of Experimental Psychology: General*. 135, 92-102.

Blough, D. S. (1975). Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 1*, 3-21.

Cohen, N.J., Eichenbaum, H., Deacedo, B.S., and Corkin, S. (1985). Different memory systems underlying acquisition of procedural and declarative knowledge. *Annals of the New York Academy of Sciences*, 444, 54–71.

Colwill, R.M. (1985). Context conditioning in Aplysia Californica. *Soc. Neuroscience Abstracts*, 11, 796.

Colwill, R.M., Absher, R.A. and Roberts, M.L. (1988). Conditional discrimination learning in Aplysia Californica. *Journal of Neuroscience*, 8, 4440-4.

Danner, D., Hagemann, D., Schankin, A., Hager, M., and Funke, J. (2011). Beyond IQ: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. Intelligence, 39, 329-334.

Dennis,N.A., and Cabeza,R. (2011). Age-related dedifferentiation of learning systems: an fMRI study of implicit and explicit learning, *Neurobiology of Aging*, 32,2318.e17-2318.e30.

Dickinson, A., Hall, G., & Mackintosh, N. J. (1976). Surprise and the attenuation of blocking. *Journal of Experimental Psychology: Animal Behavior Processes, 2,* 313–322.

Dickinson, A., & Mackintosh, N. J. (1979). Reinforcer specificity in the enhancement of conditioning by posttrial surprise. *Journal of Experimental Psychology: Animal Behavior Processes, 5,* 162–177.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Evans, JStBT. (2003). In two minds: dual process accounts of reasoning. Trends in Cognitive. Science, 7, 454–59.

Feldman, J., Kerr, B., and Streissguth, A. P. (1995). Correlational analyses of procedural and declarative learning performance. *Intelligence,* 20, 87–114.

Gebauer, G. F. (2003). Implicit learning and intelligence. Unpublished Ph.D.Thesis, University of Cambridge.

Gebauer, G. F. and Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 33, 34–54.

Hanson, H. M. (1957). Discrimination Training Effect on Stimulus Generalization Gradient for Spectrum Stimuli. *Science, 125*, 888-889.

Haselgrove, M. (2010). Reasoning rats or associative animals? A common-element analysis of the effects of additive and sub-additive pre-training on blocking. *Journal of Experimental Psychology: Animal Behavior Processes, 36*, 296-306.

Hawkins, R.D. (1989). A biologically realistic neural network model for higher-order features of classical conditioning. In R.G.M. Morris (Ed.) *Parallel Distributed Processing - Implications for Psychology and Neurobiology*. Oxford. OUP.

Hawkins, R.D., Abrams, T.W., Carew, T.J. and Kandel, E.R. (1983). A cellular mechanism of classical conditioning in Aplysia: activity-dependent amplification of presynaptic facilitation. *Science*, 219, 400-5.

Hawkins, R.D., Carew, T.J. and Kandel, E.R. (1986). Effects of interstimulus interval and contingency on classical conditioning of the Aplysia siphon withdrawal reflex. *Journal of Neuroscience*, 6, 1695-701.

Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language, 30,* 513-541.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger Publishers.

Jiménez, L., & Vázquez, G. A. (2011). Implicit sequence learning and contextual cueing do not compete for central cognitive resources. *Journal of Experimental Psychology:Human Perception and Performance*, 37, 222–235.

Jones, F.W., Le Pelley, M. E., & McLaren, I. P. L. (2002). The APECS-SRN: Towards a model of SRT sequence learning. *Proceedings of the World Congress on Computational Intelligence*, Honolulu, Hawaii.

Jones, F.W. and McLaren, I.P.L. (2009). Human Sequence Learning Under Incidental and Intentional Conditions. *Journal of Experimental Psychology: Animal Behavior Processes*, 35, 538-553.

Jones, F., and McLaren, I.P.L. **(**1999). Rules and Associations. *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*, NJ, LEA.

Kaufman, S. B. (2011). Intelligence and the cognitive unconscious. In *Cambridge handbook of intelligence* (eds R. J. Sternberg and S. B. Kaufman). Cambridge: Cambridge University Press.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., *et al.* (2010). Implicit learning as an ability. *Cognition,* 116, 321–340.

Kahneman, D. (2011) *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kamin, L.J. (1968). 'Attention-like' processes in classical conditioning. In M.R. Jones (Ed.) *Miami symposium on the prediction of behaviour: Aversive stimulation*. Univ. of Miami Press. pp. 9-33.

Knowlton, B.J., Mangels, J.A., and Squire, L.R., (1996). A neostriatal habit learning system in humans. *Science*, 273, 1399–1402.

Kohler, W. (1917). *The mentality of apes*, transl. from the 2nd German edition by Ella Winter (1925). London: Kegan, Trench and New York: Harcourt, Brace and World.

Le Pelley, M. E., & McLaren, I. P. L. (2001). Retrospective revaluation in humans: Learning or memory? *Quarterly Journal of Experimental Psychology,* 54B, 311-352.

Le Pelley, M. E., Oakeshott, S. M. and McLaren, I. P. L. (2005). Blocking and unblocking in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 31.

Lewis, M.B. & Johnson, R.A. (1999) Are Caricatures Special? Evidence of Peak Shift in Face Recognition. *European Journal of Cognitive Psychology, 11(1),* 105-117.

Livesey, E.J. and 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.

Lovibond, P.F., & Shanks, D.R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes, 28,* 3-26.

Mackintosh, N.J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276-98.

Mackintosh, N.J. (1976). Overshadowing and stimulus intensity. *Animal Learning and Behavior*, 4, 186-92.

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

Mackintosh, N.J. & Little, L. (1970). An analysis of transfer along a continuum. *Canadian Journal of Psychology*, 24, 363-369.

McAndrew, A., Jones, F.W., McLaren, R.P. and McLaren, I.P.L. (2012). Dissociating Expectancy of Shock and Changes in Skin Conductance: An Investigation of the Perruchet Effect Using an Electrodermal Paradigm. Journal of Experimental Psychology: Animal Behavior Processes, Vol 38, 203-208.

McGeorge, P., Crawford, J. R., and Kelly, S. W. 1997). The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 239-245.*

McLaren, I.P.L. (1989). The computational unit as an assembly of neurones: an implementation of an error correcting learning algorithm. In R. M. Durbin, R.C. Miall and G. Mitchison (Eds.) *The Computing Neuron*. Amsterdam. Addison-Wesley.

McLaren, I.P.L., Kaye, H. and 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*. Oxford. OUP.

McLaren, I.P.L. and Dickinson, A. (1990). The conditioning connection. *Phil. Trans. R. Soc. Lond.* B, 329, 179-186

McLaren, I.P.L., Green, R.E.A. and Mackintosh, N.J. (1994). Animal learning and the implicit/explicit distinction. In N.C. Ellis (Ed.) *Implicit and explicit learning of languages*. Academic Press.

McLaren, I. P. L., Forrest, C.L., and McLaren, R.P. (2012). Elemental representation and configural mappings: Combining elemental and configural theories of associative learning. *Learning and Behavior*, 40 (3), 320-33.

McLaren, I.P.L. and Mackintosh, N.J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning and Behavior, 38(3),* 211-246.

McLaren, I. P. L. & Mackintosh, N. J. (2002).Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior*, 30, 3, 177-200.

Miller, R. R., Barnet, R. C., and Grahame, N. J. (1995). Assessment of the Rescorla-Wagner Model. *Psychological Bulletin,* 117 (3): 363–386.

Mitchell, C.J., De Houwer, J., & Lovibond, P.F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32,* 183-246.

Pearce, J. M. (1987). A model of stimulus generalization for Pavlovian conditioning. *Psychological Review*, 94, 61-73.

Pearce, J.M. and Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532-52.

Pearce, J. M. and Mackintosh, N. J. (2010). Two theories of attention: a review and a possible integration. In: Mitchell, C. and Le Pelley, M. eds. *Attention and Associative Learning: From Brain to Behaviour*. Oxford: Oxford University Press, pp. 11-39.

Poldrack, R.A. and Packard, M.G. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*, 41, 245–251.

Reber, A. S., Walkenfeld, F. F., and Hernstadt, R. (1991). Implicit and explicit learning: Individual diff erences and IQ. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 17, 888–896.

Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological Review*, 74, 71–80.

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.

Salthouse, T. A., McGuthry, K. E., and Hambrick, D. Z. (1999). A framework for analysing and interpreting differential aging patterns: Application to three measures of implicit learning. Aging, Neuropsychology, and Cognition, 6, 1-18.

Schultz W, Dayan, P, Montague RR. (1997). A neural substrate of prediction and reward. *Science* 275: 1593-1599.

Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review, 44*, 430-444.

Spetch, M.L., Cheng, K. and Clifford, C.W.G. (2004) Peak shift but not range effects in recognition of faces. *Learning and Motivation, 35,* 221-241.

Spiegel, R. & McLaren, I.P.L. (2003). Abstract and associatively-based representations in human sequence learning. Phil Trans. Roy. Soc B., 358.

Spiegel, R. and McLaren, I.P.L. (2006). Associative Sequence Learning in Humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 32, 2, 156-63.

Stanovich, K.E. and West, R.F. (2000) Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23*, 645–726.

Suret, M. B., & McLaren, I. P. L. (2003). Representation and discrimination on an artificial dimension. *Quarterly Journal of Experimental Psychology*, 56B, pp. 30 – 42.

Thompson, R.F. (1965). The neural basis of stimulus generalisation. In D.I. Mostofsky (Ed.) *Stimulus Generalisation*. Stanford Univ. Press.

Thompson, R.F. (1989). The essential memory trace circuit and the essential reinforcement system for a basic form of associative learning. In T. Archer and L.G. Nilssom (Eds.) *Aversion, Avoidance and Anxiety: Perspective on Aversively Motivated Behaviour*. Hillsdale, N.J. LEA, pp251-264.

Unsworth, N., and Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory and Cognition*, 33, 213-220.

Wasserman, E.J., Chatlosh, D.L. and Neunaber, D.J. (1983). Perception of causal relations in humans: Factors affecting judgments of response–outcome contingencies under free-operant procedures. *Learning and Motivation*, 14, 406-432.

Weidemann, G., Tangen, J.M., Lovibond, P.F., and Mitchell, C.J. (2009). Is Perruchet's dissociation between eyelid conditioned responding and outcome expectancy evidence for two learning systems? *Journal of Experimental Psychology: Animal Behavior Processes*, 35, 169-176.
Williams, B. A. and Pearlberg, S. L. (2006). Learning of three-term contingencies correlates with Raven scores, but not with measures of cognitive processing. *Intelligence,* 34, 177–191.

Willingham D.B., Nissen M.J. and Bullemer P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 1047-1060.

Wills, S., & Mackintosh, N. J. (1998). Peak shift on an artificial dimension. *Quarterly Journal of Experimental Psychology Section B- Comparative and Physiological Psychology, 51*, 1-32.

Wills, S.J. & Mackintosh, N.J. (1999). Relational learning in pigeons? *Quarterly Journal of Experimental Psychology*, 52B, 31-52.

Yang, J., and Li, P. (2012). Brain networks of explicit and implicit learning. PLoS One, 7, e42993. doi:10:1371

Yeates, F., Jones, F.W., Wills, A.J., McLaren, R.P. and McLaren, I.P.L. (2013). Modelling human sequence learning under incidental conditions. *Journal of Experimental Psychology: Animal Behavior Processes*.

Yeates, F., Jones, F.W., Wills, A.J., Aitken, M.R.F. and McLaren, I.P.L. (2012). Implicit Learning: A Demonstration and a Novel SRT Paradigm. In: *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, Sapporo, Japan.

Yeates, F., Jones, F.W., Wills, A.J., Aitken, M.R.F. and McLaren, I.P.L. (2013). Implicit Learning: A Demonstration and a Revision of a Novel SRT Paradigm. In: *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Berlin, Germany.

Table and Figure Legends

Table 1: This shows the construction of the stimulus sequences and contingencies for the three groups in an idealised form to convey the relationship between the groups. S1-S8 are the sequences that were concatenated to construct blocks, and the mapping between responses and colors is shown for each sequence. Sequences and mappings were randomised / counterbalanced where appropriate. The letters (X, Y) stand for left/right responses, and the numbers (1, 2, 3, 4) for colors. Stimuli shown in bold are those changed with respect to Group Color. In Group Color, colors 1 and 2 are predictive, and 3 and 4 act as controls. All eight sequence triplets are shown that were used to construct the pseudorandom trial order. In Group Sequence no color is predictive, but only four sequence triplets are used so that e.g., XX is typically followed by X. In Group Dual colors 1 and 2 are once again predictive, and so are the sequences.

Table 2: The design used in Dickinson et al (1976). T=tone, L=light, Sh=shock. The subjects in this experiment were rats.

Table 3: The table shows how icon stimuli are used to construct a stimulus dimension resulting in a stimulus of the type shown in the top left corner of the table.

Table 4: Correlations between Maths and English grades with second-order factor scores. See text for details.

Figure 1: The top panels display the mean differences in RT (msec) and their standard error analysed by sequence (left panel: untrained sequences–trained sequences), and by color (right panel: untrained colors–trained colors). The corresponding mean error differences and their standard errors are shown underneath.

Figure 2: The morphed face dimension used in Suret and McLaren's (2003) experiment is shown at the bottom of the figure, and the results (mean and standard error) are shown above this continuum. Group Hard were trained on 5+ vs. 7- throughout, Group Easy Reversed on 3- vs. 9+ then transferred to 5+ vs. 7- (response assignments were counterbalanced across subjects). Note that analysis was by means of non-parametric tests that confirmed that performance on these two graphs differed significantly.

Figure 3: This figure gives a version of Spence's (1937) account of transposition (and peak shift) in terms of interacting excitatory and inhibitory generalization gradients. The black solid line is the excitatory gradient and the gray one the inhibitory gradient. The dotted line is the resultant. The dimensions shown at the bottom illustrate how this analysis could be applied to luminance or color as in Hanson's (1957) study.

Figure 4: Test performance (means and standard errors) on the Training (T), Near (N) and Far (F) stimuli on the lighter/darker shades of green dimension split by rule use vs. non-rule use. Note that 0=chance for these scores which collapse across right and left responses, 1=perfect classification and -1=perfect responding with the wrong response assignments.

Table 1

GROUP	S1	S2	S3	S4	S5	S6	S7	S8
COLOR	XXX 4 3 1	XXY 2 4 2	XYX 3 2 1	XYY 3 3 2	YXX 1 4 1	YXY 4 1 2	YYX 4 3 1	YYY 3 4 2
SEQUENCE	XXX 4 3 1	XX**X** 2 4 2	XY**Y** 3 2 1	XYY 3 3 2	YX**Y** 1 4 1	YXY 4 1 2	YYX 4 3 1	YY**X** 3 4 2
DUAL	XXX 4 3 1	XX**X** 2 4 **1**	XY**Y** 3 2 **2**	XYY 3 3 2	YX**Y** 1 4 **2**	YXY 4 1 2	YYX 4 3 1	YY**X** 3 4 **1**

Table 2

Group	Phase 1	Phase 2	Test
S-S	T - Sh	TL - Sh	L
S-D	T - Sh	TL – Sh-Sh	L
D-S	T – Sh-Sh	TL - Sh	L
D-D	T – Sh-Sh	TL – Sh-Sh	L

Table 3

	Icons							
	B	C	D	E	F	G	H	I
Stimuli								
Far+	3	6	3					
Near+		3	6	3				
S+			3	6	3			
S-				3	6	3		
Near-					3	6	3	
Far-						3	6	3

Table 4

Subject / Factor	Maths	English
Explicit-IQ	0.30	0.26
Implicit	0.14	0.26

Figure 1

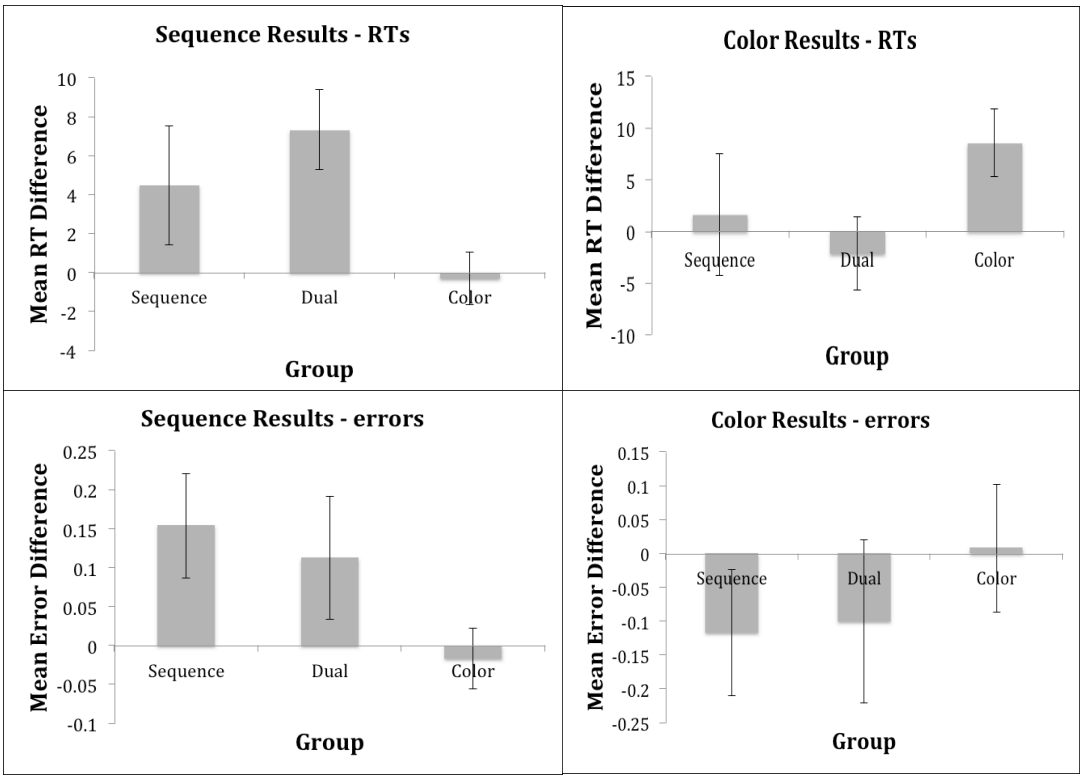

Sequence Results - RTs
Mean RT Difference
10
8
6
4
2
0
-2
-4
Sequence
Dual
Color
Group
Color Results - RTs
Mean RT Difference
15
10
5
0
-5
-10
Sequence
Dual
Color
Group
Sequence Results - errors
Mean Error Difference
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
Sequence
Dual
Color
Group
Color Results - errors
Mean Error Difference
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
Sequence
Dual
Color
Group

Figure 2

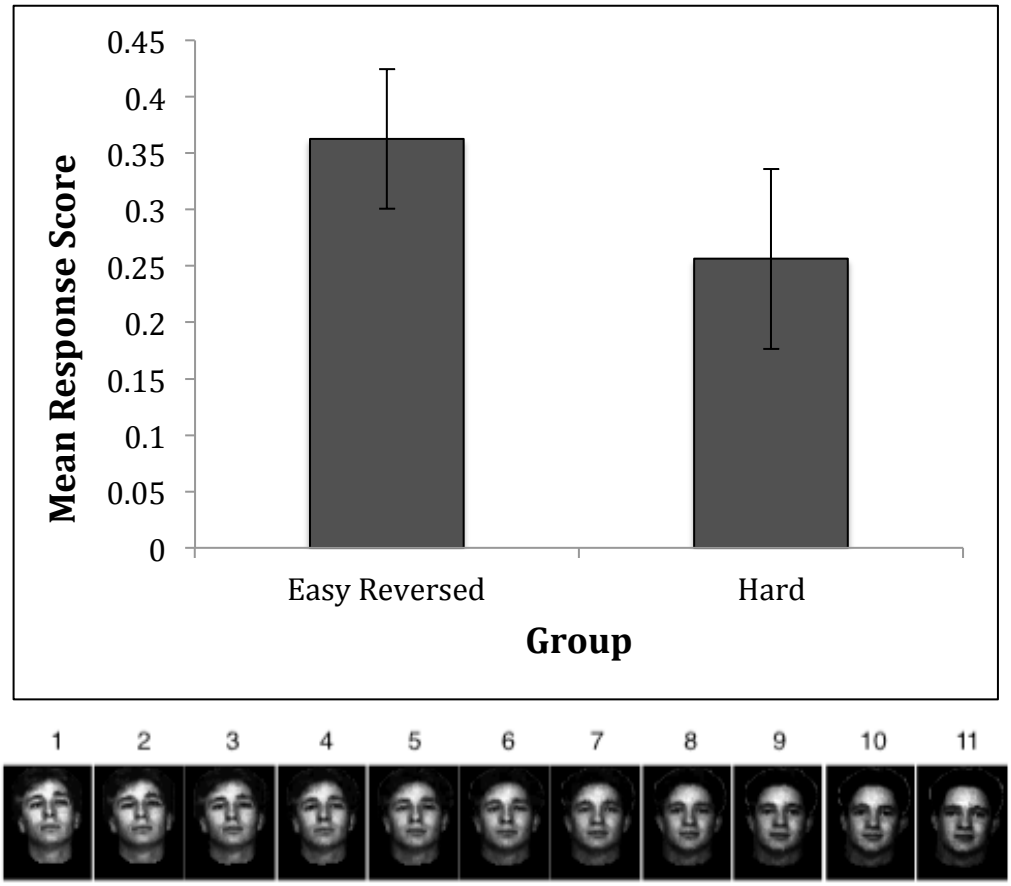

0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Mean Response Score
Easy Reversed
Hard
Group
1
2
3
4
5
6
7
8
9
10
11

Figure 3

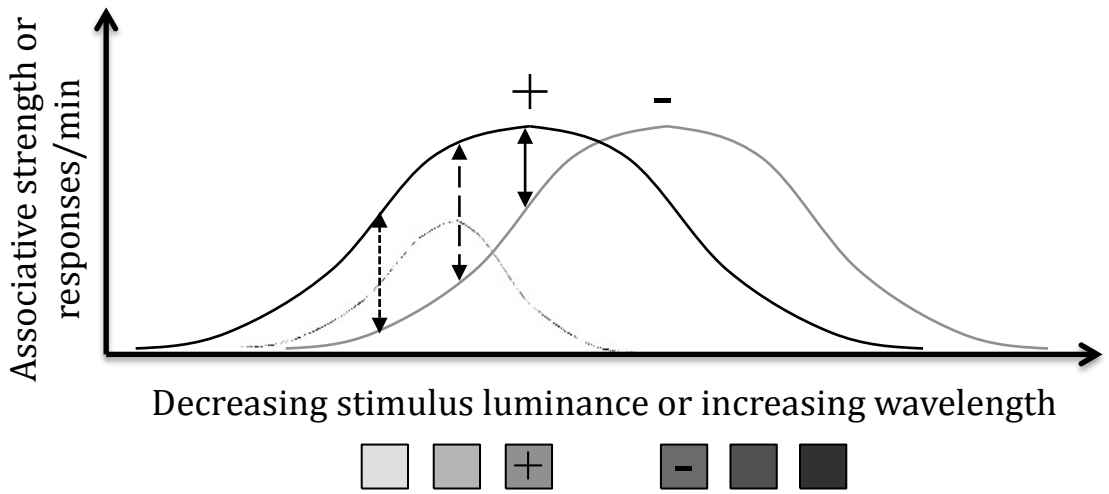
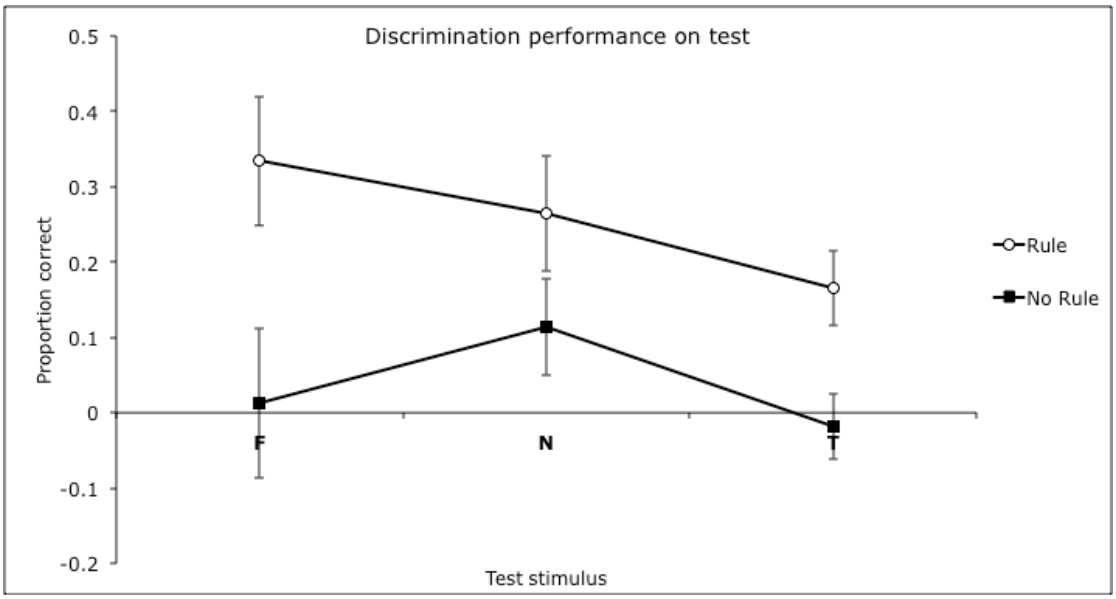

Associative strength or responses/min
+
-
Decreasing stimulus luminance or increasing wavelength

Figure 4


Discrimination performance on test
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
Proportion correct
F
N
T
Test stimulus
Rule
No Rule