To be published in *Behavioral and Brain Sciences* (in press)

The propositional nature of human associative learning

Chris J Mitchell
School of Psychology
University of New South Wales
Kensington 2052 Australia
Chris.mitchell@unsw.edu.au
http://www.psy.unsw.edu.au/profiles/cmitchell.html

Jan De Houwer
Department of Psychology
Ghent University
Henri Dunantlaan 2
B-9000 Ghent Belgium
jan.dehouwer@ugent.be
http://users.ugent.be/~jdhouwer/

Peter F Lovibond
School of Psychology
University of New South Wales
Kensington 2052 Australia
p.lovibond@unsw.edu.au
http://www.psy.unsw.edu.au/profiles/plovibond.html

Abstract: The past 50 years have seen an accumulation of evidence suggesting that associative learning depends on high-level cognitive processes that give rise to propositional knowledge. Yet many learning theorists maintain a belief in a learning mechanism in which links between mental representations are formed automatically. We characterize and highlight the differences between the propositional and link approaches, and review the relevant empirical evidence. We conclude that learning is the consequence of propositional reasoning processes that cooperate with the unconscious processes involved in memory retrieval and perception. We argue that this new conceptual framework allows many of the important recent advances in associative learning research to be retained, but recast in a model that provides a firmer foundation for both immediate application and future research.

Keywords: Conditioning, associative link, association, human associative learning, dual-system, awareness, automatic, propositional, controlled

1. Introduction

The idea that behavior is determined by two independent and potentially competing systems has been used repeatedly in psychology (see Evans, 2008 for a recent review of some of these ideas). The diversity of research areas in which this idea has been reproduced is striking. It includes, for example, fear learning (e.g. Öhman & Mineka, 2001), memory (e.g. Schacter, 1987), reasoning (e.g. Evans 2003), decision making (e.g. Kahneman & Frederick, 2002) and the activation of attitudes (e.g. Wilson, Lindsey & Schooler, 2000). In each case, one system is generally characterised as conscious, cold and calculating, and the other as unconscious, affective and intuitive. In this paper we reconsider (and reject) one of the oldest and most deeply entrenched dual-system theories in the behavioral sciences, namely the traditional view of associative learning as an unconscious, automatic process that is divorced from higher-order cognition.

The classic empirical demonstration of associative learning comes from Pavlov (1927). He presented his dogs with a ringing bell followed by food delivery. As a consequence, the dogs would salivate on hearing the sound of the bell, even in the absence of food. This shows that Pavlov's dogs learned to associate the bell with the presentation of food. The biologically neutral bell is usually referred to as a conditioned stimulus (CS) and the biologically relevant food (to a hungry dog) is referred to as an unconditioned stimulus (US). Most contemporary animal learning theorists now consider that the dogs salivated on hearing the bell because a link formed between the mental representations of the bell (CS) and food (US). This link allowed the presentation of the bell to activate the mental representation of food (see Figure 1) and, therefore, produce salivation in much the same way as would actual presentation of the US itself.

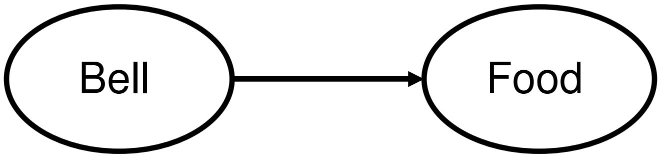

Figure 1. Elipses indicate metal representations (of the bell and the food). The arrow between the two elipses indicates the mental link formed as a consequence of bell-food pairings. The bell ringing produces salivation because it activates the mental representation of food, which, in turn, produces salivation.

It is clear from this description of Pavlov's (1927) hugely influential work, that the term "associative learning" has two meanings. These meanings are often confused. The first refers to a phenomenon – the capacity possessed by a broad range of organisms to learn that two or more events in the world are related to one another. That is, one event may refer to, signal, or cause the other. This meaning of the term associative learning is silent as to the psychological mechanism responsible for learning. The second meaning of the term associative learning does specify a psychological mechanism. This mechanism is the formation of links between mental representations of physical stimuli as illustrated in Figure 1. The links are said to be formed passively and automatically as a direct consequence of contiguous (with some restrictions) pairings of those physical stimuli. These mental links then allow the presentation of one stimulus to activate the representation of, or bring to mind, the other stimulus. Many researchers assume that learning about the relationships between events in the environment (the phenomenon) takes place via the formation of links between mental representations of those events (the mechanism). The present paper argues against this position and aims to show that associative learning results, at least in humans, not from the automatic formation of links, but from the operation of controlled reasoning processes. These processes result in beliefs about the world in the form of propositions rather than simply links that allow one representation to activate another. Thus, in the context of the present argument, the term "associative learning" refers to the ability to learn about relationships between events, not to a mechanism by which mental links are formed. In order to distinguish the two main approaches to theorizing about mechanisms of associative learning, we refer descriptively to the automatic link formation mechanism and its alternative, the propositional approach.

A core difference between the two approaches (propositional and link-based) is related to the way in which knowledge is assumed to be represented. As Shanks (2007, p. 294) points out, propositional representations "have internal semantic or propositional structure in the same way that language does. The English sentences 'John chased Mary' and 'Mary chased John' have the same elements but do not mean the same thing as they are internally structured in different ways. The alternative to such propositional or cognitive representations is an association that simply connects the mental images of a pair of events in such a way that activation of one image causes activation (or inhibition) of the other". Dickinson (1980, p. 85) similarly describes "an excitatory link which has no other property than that of transmitting excitation from one event representation to another." These quotes reveal that a proposition differs from a link in that it specifies the way in which events are related. For instance, a proposition can specify that the bell *signals* food. In contrast, a link between representations only allows activation to pass between those representations. The link itself has no representational content; there is nothing stored to indicate the nature of the relationship between the stimuli (Fodor, 2003). This means that a proposition has a truth value (see Strack & Deutsch, 2004), but a link does not. That is, a proposition can be shown to be true or false. In the case above, it can be demonstrated that the bell does or does not signal food. A link cannot be shown to be true or false because it does not represent any particular relationship between the bell and food.

Proponents of the automatic link mechanism do not deny that propositional reasoning processes can generate knowledge of relationships between events in the world. However, they argue that the link formation mechanism is able to produce learning independently and in an automatic manner. This point has already been made by Shanks (2007, p. 297). As he says, “It is important to realise that when arguing for a contribution of associative processes, supporters of this approach have never denied that rational causal thinking takes place … Rather, the question is whether all causal thought is of this form, or whether instead there might be a separate type of thinking (associative) when people make intuitive judgments under conditions of less reflection”. Likewise, McLaren, Green, and Mackintosh (1994) “agree there exist two qualitatively different types of learning” (p. 315) “an associative system which cumulates information about contingencies between events and a cognitive system with beliefs and reasons for those beliefs” (p. 327) “By associative learning, we mean learning that can be characterised in terms of the establishment of links between representations” (p. 316). They assume that the formation of links occurs “automatically, regardless of the subject’s plans or intentions” (p. 321). Thus, the alternative to the propositional approach is a dual-system approach; behavior is determined by both the propositional reasoning system and the automatic link formation mechanism. A critical issue then is whether there is evidence for the second component of the dual-system approach, the automatic link formation mechanism.

It is important to be clear that our aim is not to evaluate individual models of learning or propositional reasoning, of which there are many. Our aim is simply to compare the broad class of dual-system models with the broad class of propositional models. It is for this reason that we use the terms “propositional approach” and “dual-system approach”. These two approaches differ in fundamental and testable ways. To summarize, the propositional approach suggests that controlled reasoning processes are necessary for learning to take place, and learning results in beliefs about the relationship between events. This can be contrasted with the idea that learning is sometimes the consequence of the automatic formation of excitatory and inhibitory links between stimulus nodes or representations.

In this paper, we present a brief and selective survey of the literature on associative learning (for more complete reviews of some specific aspects of the literature see Lovibond and Shanks, 2002; De Houwer, Vandorpe & Beckers, 2005; De Houwer, in press). In this survey, we find clear support for the role of propositional processes in learning. In stark contrast, little unambiguous support is found for an automatic link formation mechanism. We conclude that there is very little to be lost, and much to be gained, by the rejection of the dual-system approach that incorporates an automatic link-formation mechanism. This is true for our understanding of the basic processes of associative learning (at both the psychological and physiological level) and in the application of learning theory to pathological behaviors in the clinic.

2. The dual-system approach to learning

The dual-system approach incorporates all of the reasoning processes of the propositional approach plus an additional automatic link-formation mechanism. Therefore, it is this link-formation mechanism that is the focus of the present section.

2.1 Learning

As outlined above, the usual view is that links between representations can be formed automatically in the sense that they are independent of the goals, processing resources, and causal beliefs of the individual (see Moors & De Houwer, 2006, for an analysis of the concept "automatic"). Thus, as Le Pelley, Oakshott, and McLaren (2005, p. 65) have argued, imposing a cognitive load will "hamper participants' use of cognitive strategies in contingency learning, instead forcing them to rely on 'automatic' associative processes". This implies that these (link-based) associative processes are automatic in the sense that they are efficient (see also Dickinson, 2001, p. 23).

Although the link mechanism is often thought to be efficient and to operate independently of the subject's goals, link formation is not assumed to be completely unconditional. A number of different learning rules have been proposed that can be seen as setting restrictions on the conditions under which the pairing of events leads to the formation of a link between the representations of those events (e.g., Mackintosh, 1975; Pearce, 1987; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Wagner, 1981). For example, it is generally accepted that links will be formed only if the CS is attended (e.g., Mackintosh, 1975; Pearce & Hall, 1980). Similarly, Rescorla and Wagner (1972) proposed that contiguous pairings of a CS and US will not produce an associative link between the two if the representation of the US is already activated (or is unsurprising), for instance because a second pre-trained CS is present on that trial. This is the phenomenon of blocking (Kamin, 1969) – the pre-trained CS will block the formation of a link between the target CS and the US – which is an example of competition between cues to gain "associative strength". Blocking is a very important phenomenon in the study of learning precisely because it shows that contiguous stimulus pairings do not always produce associative learning.

The link-formation mechanism is thought to be responsible not only for blocking but also for many other conditioning phenomena (e.g. conditioned inhibition, overexpectation effects etc.) and is thought to apply equally to all stimuli across different modalities and in a wide range of species. The generality of the phenomena (perhaps most importantly blocking) across these different situations and species is often argued to demonstrate that all species possess a common learning mechanism (e.g., Dickinson, Shanks, & Evenden, 1984). The mechanism must, it is sometimes further argued, be very simple and automatic because surely species such as the humble rat could not possess the complex hypothesis testing abilities of humans.

2.2 Performance

The link model provides a ready explanation for conditioned responses (CRs) such as salivation to a CS that has been paired with food. Once the link is formed, activation can be transmitted from one representation to another just as a piece of copper wire conducts electricity. Thus, when a CS such as a bell is presented on test, it activates the mental representation of that bell. This activation is then transmitted along the link and so the US representation also becomes activated (see Figure 1). Activation of the US representation produces salivation (the CR) because it is functionally equivalent to actual physical presentation of food. Thus, the link mechanism provides a very simple and intuitive account of why, when a CS is presented in the absence of the US on test, behaviors consistent with actual US presentation, such as salivation, are often observed.

Of course, this characterization of operation of the link model is overly simplistic and easily discredited (see Wagner & Brandon, 1989). Within this model, activation of the US representation by the CS (via the link) is equivalent to activation of the US representation by presentation of the US itself. Associative learning theorists are well aware that presentation of the CS and US do not have exactly the same consequences; the CS is not a substitute for the US. Wagner's (1981) influential Sometimes Opponent Processes (SOP) model of associative learning addresses this issue. Wagner distinguishes between a primary and a secondary state of activation, termed A1 and A2 respectively. It is only when a US is physically present that its representation (or some part thereof) will be activated into the (primary) A1 state. Following earlier CS-US pairings, presentation of the CS will associatively activate the US representation into the (secondary) A2 state. Thus, Wagner's model postulates different states of activation to distinguish between perception of the US when it is physically present and anticipation of that US.

There are also other ways in which a US representation can be activated into the A2 state. When a US is presented (and its representation is activated into A1), removal of that US will allow the representation to decay into A2. In this case, A2 activation of the US representation would seem to equate to memory of the US. One thing that is striking about this model is that it does not distinguish between memory for a US in the recent past and anticipation of a US in the future (which have very different behavioural consequences, see Bolles & Fanselow, 1980). That is, both US memory and US anticipation are represented by A2 activation. Further refinement would be needed to accommodate this important distinction, However, what is important is that if one postulates different states of activation, then the idea of simple activation can come to mean different things, and the link model becomes much more flexible.

Anticipatory CRs such as salivation or fear are not the only responses said to be produced by the link mechanism. Learning theorists have also applied this same approach to the analysis of human contingency learning. An example of a contingency learning task is the allergist task (e.g., Larkin, Aitken, & Dickinson, 1998). Participants play the role of an allergist who is asked to determine which food cues produce an allergic reaction outcome in a fictitious Mr. X. In the case of simple conditioning, Mr X eats a food such as carrots on each trial and always suffers an allergic reaction. Participants learn that carrots are associated with the allergic reaction. The automatic link-formation mechanism is thought to operate in this scenario just as it does in Pavlovian conditioning; a carrot-allergic reaction (cue-outcome) link is formed, such that presentation of the cue is able to activate the representation of the outcome. When a food that has been followed by the allergic reaction during training is judged to be allergenic on test, it is argued that this judgment is the consequence of the cue-outcome link that has formed.

In fact, Pearce and Bouton (2001) suggest that the link between cue and outcome can serve to represent a whole range of different associative relationships. This further implies that a causal relationship between the cue and outcome (e.g., drinking alcohol causes headache) is represented in exactly the same way as a predictive relationship (e.g. hearing the platform announcement predicts, but does not cause, the arrival of a train). It also implies that causal and predictive relationship are represented in the same way as purely referential relationships, in which the cue merely refers to the outcome without an expectation that the outcome will actually occur (e.g. the word "sun" uttered at night refers to the sun but does not produce an expectation that the sun will appear in the immediate future), or the relationship between a category (e.g. animals) and an exemplar of that category (e.g. a cat).

However, these relationships are not equal. It is known, for example, that whether the cues and outcomes in an associative learning experiment are presented in a causal or a predictive scenario has a profound effect on the pattern of responding seen on test (Pineno, Denniston, Matute, Beckers & Miller, 2005; Vadillo & Matute, 2007; see also Waldmann, 2000 for a similar argument in the context of causal and diagnostic learning). The simple link mechanism, because it cannot capture the precise nature of the associative relationship between cue and outcome, cannot explain these effects and so cannot explain many aspects of human associative learning. Of course, as was pointed out above, the automatic link formation mechanism has been argued to be only one system in a dual-system approach to learning. It is open to proponents of this approach to argue that the differences observed between causal and predictive cues are a consequence of the second, propositional process not the automatic links (e.g., Vadillo & Matute, 2007). We shall return to this issue below.

In summary, the dual-system approach suggests that, in addition to the reasoning processes that produce conscious propositional knowledge, there exists an automatic, hard-wired mechanism that produces links between CSs and USs (or cues and outcomes). In Pavlovian conditioning, these links allow the presentation of the CS to activate the US representation, and this produces a CR. The link formation mechanism is also thought (under certain circumstances) to be responsible for the learning of other types of relations including predictive, causal and referential relations, and is assumed to operate in all species including humans.

3. The propositional approach to learning

According to the propositional approach, associative learning depends on effortful, attention-demanding reasoning processes. The process of reasoning about the relationship between events produces conscious, declarative, propositional knowledge about those events.

3.1 Learning

When we learn that Mr X has an allergy to carrots, or that a bell will be followed by food, we use the same processes of memory and reasoning that we use to plan our grocery shopping, play chess or behave appropriately at a black tie function. When presented with a bell, we may recall that the last time the same bell rang, we received food. Given a number of assumptions (e.g. that relations are stable over time and that the bell is a potential signal for food), this might lead us to hypothesize that we are about receive food again. We may also recall having previously hypothesised that the bell signals food. When we do indeed receive food, that experience constitutes a test (a confirmation) of our hypothesis. Thus the strength of our belief in the bell-food relationship will increase. The encoding of this hypothesis in memory, and the degree to which we have confidence in it constitutes learning. There is no mental link between the bell and food, but a proposition of the form “when I hear a bell, I will receive food”.

Propositions can be regarded as qualified mental links, that is, links that specify how two events are related. This approach is also adopted in the Bayesian network approach to the analysis of belief acquisition and revision (see Lagnado, Waldmann, Hagmayer & Sloman, 2007, for a very useful overview). In Bayes nets, events are joined by, for example, a causal link – an arrow that has a particular strength and direction. Thus, an arrow that points from “bacterial infection” to “peptic ulcer” might indicate that bacterial infection *causes* peptic ulcers. Because the links in Bayes nets represent propositions about relationships, like all propositions, they have truth value (e.g., it is either true or not true that bacterial infection causes peptic ulcers). Thus, the arrows do not simply indicate that activation can spread from one mental representation to another in that direction. Despite these similarities, the Bayes net framework and the propositional approach are not identical. Most importantly, the Bayesian approach is silent as to whether belief acquisition involves controlled or automatic processes. The propositional approach presented here makes the strong claim that associative learning is never automatic and always requires controlled processes.

Associative learning theorists are often concerned, not simply with whether or not a CR is produced, but with the strength of the CR, which is thought to be a measure of “associative strength”. Within the propositional approach, associative strength relates to two things. The first is the belief about the strength of the CS-US relationship. Thus, a belief may be held that a CS is followed by a US on 50% of occasions. This will, of course produce a weaker CR than a belief that the CS is followed by the US on 100% of occasions. The second is the strength of the belief, which will typically be low at the start of training and high after many CS presentations. Thus, associative strength will be jointly determined by how strong the CS-US relationship is believed to be (the content of the belief) and the strength of that belief (the degree of confidence with which it is held).

The description of learning presented above leaves some important issues unspecified. Firstly, we do not specify the nature of the controlled processes, beyond characterizing them as propositional reasoning. That is, we do not propose a new model of propositional reasoning. There are many ways to model reasoning processes (e.g. Braine, 1978, Johnson-Laird, 1983, Chater, Tenenbaum & Yuille, 2006 ; Evans & Over, 1996) some of which are specifically designed to account for the learning of causal relationships between events (e.g. Cheng, 1997; Kruschke, 2006). We would not argue for the virtues of any particular model of reasoning, only that associative learning requires reasoning, however that is achieved.

Secondly, even though we postulate that associative learning is influenced by memory for prior events, we do not propose a new model of memory. Probably the simplest memory model that would be consistent with our view is an instance model of memory (e.g. Hintzman, 1986). According to this model, separate experiences are stored as separate memory traces that can be retrieved on the basis of similarity with the current stimulus input. Thus, a bell can retrieve memories of past occasions on which the bell was presented, and therefore past bell-food pairings.

Thirdly, we do not rule out a role for automatic processes in learning. Memory retrieval has many features of automaticity, and so some of the processes that result in learning must also be automatic. However, this does not imply that learning itself is automatic. According to the propositional approach, recollections of past bell-food pairings alone cannot produce learning. These recollections only serve as one kind of input into the propositional reasoning processes that are responsible for learning. Other kinds of input will include, for example, the knowledge that there was no other signal for food present when bell-food pairings were experienced, and the belief that bells are, in general, potential signals for food delivery.

It is important to make clear that allowing automatic processes of memory (and, indeed, perception) to play a role in learning, does not imply that the propositional approach is simply another dual-system approach. The way in which automatic and controlled processes interact to produce learning in the propositional approach is quite unlike that of the dual-system approach. In the dual-system approach, two incompatible CS-US (e.g. bell-food) relationships might simultaneously be learned by the two systems (although, it should be noted, it is seldom explained how these systems might interact under such circumstances). For example, a strong bell-food link may form in the absence of any belief that presentation of the bell signals food delivery. In contrast, in the propositional approach this is not possible because the automatic processes of perception and memory serve only as an input to the non-automatic processes of propositional reasoning. These two types of process are simply different parts of the same learning system.

Lastly, it is important to be clear on the way in which the propositional approach deals with the role of consciousness in learning. We do not claim that people are necessarily aware of all of the processes that lead to the formation of propositions about relationships between events, including the reasoning processes. What we do claim is that the propositional beliefs themselves are available to consciousness. Thus, it is not possible to have learned about a relationship between two events in the environment without being, or having been, aware of that relationship (see De Houwer, in press).

3.2 Performance

The consequence of entertaining a belief that the bell CS signals the food US (or, in other cases, that the CS causes the US) is that, when I next hear the bell, I shall (all things being equal) anticipate, or expect the food to be presented. Early cognitive psychologists also viewed conditioned responses to be the consequence of US expectancy. They assumed that the strength of the CR (e.g., skin conductance elevation) in conditioning with a shock US would be a product of the strength of the expectancy of shock and the value (intensity or aversiveness) of that shock (e.g., MacCorquodale & Meehl, 1954). However, expectancy was thought of in terms of a link that allowed the CS to activate the US representation. The propositional approach departs from the views of these early theories in that the knowledge of the associative relationship between CS and US is a belief represented in propositional form. Thus, the expectancy of the US when the CS is presented is a consequence of the belief that the CS causes or signals the US.

One problem that is often raised in the context of expectancy-based models of emotional and physiological conditioned responses is how an expectancy can give rise to such responses. We do not have a solution to this long-standing problem. However, we already know that instructions can produce physiological and emotional responses in the absence of any CS-US link. For instance, the mere instruction that an electric shock is forthcoming leads to an increase in fear and skin conductance (Cook & Harris, 1937). Thus, if it is assumed that instructions produce CRs by generating an expectancy of the US, then there must be a process by which US expectancy can generate physiological CRs, even though this process is not yet well understood.

A related issue is that skin conductance and heart rate CRs seem uncontrollable and, in this sense, therefore, automatic. This seems to imply that an automatic learning system is in operation. However, the idea that conditioned responses can arise automatically can be accounted for within the propositional approach in two ways. First, we do not argue that subjects have control over their responses to expected USs, but rather that learning to expect those USs is a non-automatic process. Once there is an expectancy that the US will occur (the subject has learned that the CS that has been presented predicts the US), this can automatically lead to emotional and physiological responses; if one believes that a painful shock is imminent, it is difficult not to experience fear. Second, once a proposition has been formed that the CS causes or signals the US, it will be stored in memory and may be activated automatically. Hence, the presentation of a CS can automatically lead to the expectation of the US (and thus to conditioned responding) if the previously formed CS-US proposition is retrieved automatically from memory. Whether the CS-US proposition is retrieved automatically from memory will depend on a number of factors, including the number of times that the CS-US proposition has been consciously entertained. In summary, although learning results from controlled processes, performance may be automatic.

With regard to performance in causal or contingency learning, the propositional approach applies in a very straightforward way. Take the example of the food-allergy paradigm. Participants are assumed to form propositions about the relation between foods and allergies (e.g., carrots cause an allergic reaction). When asked to rate the contingencies between different foods and allergies, participants simply need to express their propositional knowledge. That is, the report of contingency knowledge is merely the verbal expression of a belief.

3.3 Predictions of the propositional and dual-system approaches

The propositional and dual–system approaches make a number of different predictions about the conditions under which learning will occur, and the pattern of responding that might be observed when different contingencies are in place. Firstly, whether learning can take place in the absence of awareness of the CS-US (or cue-outcome) contingencies is relevant to this debate. The propositional approach assumes that learning involves testing hypotheses and that it results in conscious propositional beliefs. One would, therefore, expect participants who successfully learn the CS-US contingencies to be aware of, and be able to report, those contingencies. By contrast, if learning is automatic, it may take place in the absence of such awareness. Secondly, the propositional approach suggests that all learning is effortful and so should depend on the availability of sufficient cognitive resources. The link formation mechanism, because it is automatic (in the sense that it is efficient) should be less dependent on cognitive resources. Thirdly, hypotheses about how events are related to each other can be acquired by verbal instruction and will be influenced by abstract rules and deductive reasoning processes. Therefore, the propositional approach predicts that learning will similarly be affected by these factors. The automatic link formation mechanism is non-propositional. It cannot, therefore, be affected directly by verbal instruction, rules or deduction.

In Section 4, we present the findings that lend support to the propositional approach. In Section 5, we outline the evidence that has been argued to provide strong support for the dual-system approach. It will be suggested at the end of Section 5 that the balance of evidence strongly favours the propositional approach.

4. Evidence for the propositional approach

4.1 The role of awareness in associative learning

Because learning is assumed to involve the strategic testing of hypotheses and to result in conscious propositional knowledge about relations between events in the world, a propositional approach predicts that learning should be found only when participants have conscious awareness of the relevant relations. If evidence for unaware conditioning were uncovered, this would, therefore, strongly support the existence of multiple learning mechanisms (Lovibond & Shanks, 2002; see also Boakes, 1989; Brewer, 1974; Dawson & Schell, 1985; Shanks & St. John, 1994).

In Pavlovian conditioning of human autonomic responses, for example, a CS (e.g., a light) is paired with an aversive US such as an electric shock. On test, learning is evidenced by the ability of the CS to increase the participant's skin conductance, a measure of fear. The results consistently show evidence for skin conductance CRs only in participants who are aware of the CS-US contingency (for reviews see Dawson & Schell, 1985; Lovibond & Shanks, 2002). Moreover, CRs occur only after the participants become aware of the CS-US contingency. Such results have led to the conclusion that awareness of the CS-US contingency is a necessary condition for Pavlovian conditioning to occur (Dawson & Shell, 1985). Other studies of conditioning with shock USs suggest that the close link between learning and awareness is due to the fact that consciously available hypotheses determine how the participant will respond. For instance, interindividual differences in human autonomic conditioning are closely related to interindividual differences in the extent to which the US is expected at a particular moment in time (e.g., Epstein & Roupenian, 1970). When participants have incorrect beliefs about the association between events or between a response and an event, their conditioned behavior is most often in line with the incorrect beliefs rather than the objective contingencies (e.g., Parton & DeNike, 1966).

Lovibond and Shanks (2002) concluded that the available evidence, from a whole range of conditioning procedures, is consistent with the idea that conditioning is accompanied by awareness. Although there are many papers arguing for unaware conditioning, close inspection reveals, in almost all cases, that the measure of conditioning was most likely more sensitive than that of awareness. This may have been because, for example, a recall rather than a recognition test of contingency awareness was used, or because contingency awareness was only tested after an extinction phase (see Dawson and Schell, 1985; 1987, for excellent reviews of these issues). These flaws have the potential to lead to an apparent dissociation between conditioning and awareness when, in fact, none exists. Only two possible exceptions were identified by Lovibond and Shanks, evaluative conditioning (e.g. Baeyens, Eelen, & Van den Bergh, 1990) and the Perruchet effect (e.g. Perruchet, 1985). We shall return to these in Section 5.

Before we accept that the absence of evidence for unaware conditioning constitutes evidence against the automatic link mechanism, we should consider the alternatives. For example, perhaps the observed concordance between awareness and CRs does not result from the US expectancy causing the CR (as we have suggested), but the CR causing the US expectancy. Thus, following CS-shock training, presentation of the CS will elicit CRs such as increased anxiety, heart rate and arousal. When the participant experiences these physiological CRs, they may then draw the conclusion that the shock is about to be presented, and so they become aware of the CS-US contingency (Katkin, Wiens & Öhman, 2001; Öhman & Soares, 1993; 1994; 1998). Alternatively, it may be argued that, although the link-formation mechanism is automatic in some respects (e.g., it is efficient and independent of the learner's goals), it is not automatic in the sense that it is unconscious. This would be a second way in which the absence of unaware conditioning might be argued not to be inconsistent with the dual-system approach.

To summarize, a demonstration of unaware conditioning would be highly damaging to the propositional approach, and would provide strong evidence for a second (automatic) learning mechanism. However, a large body of literature shows a clear concordance between conditioning and awareness, and provides, therefore, no unique support for an automatic learning mechanism. So what can be concluded from these data? The observed concordance between conditioning and awareness is strongly predicted by the propositional approach. And, although the absence of unaware conditioning cannot be taken as decisive evidence in the present debate (an absence of evidence rarely is decisive), it is only consistent with the existence of the link-formation mechanism if certain additional assumptions are made. Thus, if anything, the data support the propositional approach. Finally, it should be noted that if we acknowledge that learning depends on awareness, then we remove one of the reasons for postulating a dual-system approach in the first place. If all learning is aware, there is less to be gained from postulating an automatic link formation mechanism in addition to a propositional reasoning mechanism.

4.2 Cognitive load and secondary tasks

According to the propositional approach, learning depends on the involvement of propositional reasoning processes that require attentional/cognitive resources. Thus, secondary tasks that consume cognitive resources, or instructions that divert attention away from the target association, are predicted to impair learning. A small decrease in attention may not be sufficient to reduce learning, but any manipulation that is sufficient to interfere with the formation or deployment of propositional knowledge about the CS-US relation should also reduce CRs to that CS. One way in which processes can be automatic is that they require only limited cognitive resources. Thus, if reduced attention to the target relationship leads to a reduction in learning of that relationship, this would seem to suggest that learning is cognitively demanding and, in this sense, not automatic.

The most thorough investigation of the effect of attentional manipulations on conditioning was conducted by Dawson and colleagues in the 1970s (eg., Dawson, 1970; Dawson & Biferno, 1973). They embedded a differential autonomic conditioning design within an "auditory perception" masking task that required participants to answer several questions at the end of each trial concerning the pitch of a series of six tones. In fact, one tone was paired with shock (CS+) and another tone was never paired with shock (CS-). Propositional knowledge of the differential contingency was assessed by online expectancy ratings and by a post-experimental interview. The results were clear-cut. The addition of the masking task substantially reduced both contingency knowledge and differential electrodermal CRs. Participants who were classified as unaware of the differential contingency failed to show any differential CRs. Furthermore, the expectancy ratings and electrodermal CRs were closely related. When the data for "aware" participants were aligned around the trial on which they first showed expectancy discrimination, the electrodermal measure similarly showed differentiation after but not before that point. Dawson's results are not unusual; the same pattern has been observed repeatedly across different conditioning preparations, and there is no convincing example of a differential impact of reduced attention on verbalisable knowledge and CRs (see Lovibond & Shanks, 2002).

The finding that learning processes are disrupted by the addition of a masking task suggests that learning requires cognitive resources and is, in this sense, not automatic. It is, therefore, evidence against an automatic link formation mechanism. However, it might be argued that no psychological mechanism or process places zero requirements on cognitive resources; there are no automatic processes in this very strict sense. There are degrees of automaticity (Moors & De Houwer, 2006). Thus, the link formation mechanism, although cognitively demanding, may be less demanding than other tasks such as reasoning and problem solving. Alternatively, perhaps cognitive load does not prevent the automatic link-formation mechanism itself from operating, but rather, it reduces the degree to which the stimulus input (the CS and US) is processed. If the participant fails to notice the stimuli, there will be no input to the automatic learning system, and nothing will be learned. Either of these interpretations of the effect of cognitive load would, of course, constitute quite a large concession. If all learning depends on cognitive resources, then one of the reasons for postulating the existence of an automatic link formation mechanism has been removed (as was the case for the role of awareness in conditioning, see Section 4.1 above). Moreover, such a concession weakens the testability of Dickinson (2001) and Le Pelley et al.s (2005)'s claim that when the cognitive system is overloaded, the operation of the link mechanism will be revealed. If the link formation mechanism depends on cognitive resources, then imposing a mental load during a learning task cannot, as has been claimed, reveal the operation of that mechanism in the absence of propositional reasoning.

Furthermore, one recent study seems to suggest that the introduction of a secondary task does not simply reduce stimulus processing. This time the evidence comes from studies of blocking in human contingency learning. In blocking, as described previously, pairing of one cue, A, with the outcome (A+) in a first phase prevents learning about the target cue T on subsequent AT+ trials. De Houwer and Beckers (2003) found that blocking in human contingency learning was less pronounced when participants performed a demanding secondary task during the learning and test phases, than when they performed an easy secondary task. In other words, increasing the demands of the secondary task *increased* the degree to which participants learned a T-outcome relationship. Waldmann and Walker (2005) obtained a similar result, attesting to the reliability of this finding. This is the precise opposite of the outcome predicted by the account outlined above, according to which cognitive load has an effect on learning by reducing the degree of stimulus processing. By that account, the secondary task should have reduced learning about T on AT+ trials. The result is, however, in line with the hypothesis that blocking depends on effortful controlled processes, as predicted by the propositional approach; participants were prevented from reasoning that, because A is a cause of the outcome, T is, therefore, redundant.

4.3 Verbal instructions

Many studies have shown that informing participants verbally about a relationship between stimuli is sufficient to produce evidence of learning. In an example presented earlier, if one informs a participant that a tone will always be followed by a shock, the tone will produce an increase in skin conductance, even though the tone and shock have never actually been presented together (Cook & Harris, 1937). Likewise, if one first presents tone-shock trials and then verbally instructs the participants that the tone will no longer be followed by the shock (instructed extinction), the skin conductance CR will be dramatically reduced (e.g., Colgan, 1970). Thus, verbal instructions can lead to the same effects as the actual experience of a contingency, and can interact with knowledge derived from actual experience.

Recent studies have shown that these conclusions also hold for more complex learning phenomena. Lovibond (2003), using an autonomic conditioning procedure, trained a compound of A and T with shock (AT+) and then presented CS A without the US (A-). The A- training in the second phase increased the CR observed to T on test, a phenomenon known as release from overshadowing. Release from overshadowing could result from reasoning that (a) at least one of the cues A or T must signal the shock on AT+ trials and (b) because A was subsequently found to be safe, T must be the signal. Importantly, Lovibond (2003) also found release from overshadowing also when the AT+ and A- trials were described verbally (Experiment 2) and when the AT+ trials were actually presented, but the subsequent A- contingency was described verbally (Experiment 3). This shows that knowledge acquired verbally and that acquired by direct experience is represented in a similar way. Thus, the implication is that the knowledge acquired by experience is propositional in nature.

It is very difficult to explain effects such as instructed conditioning in terms of an automatic link mechanism. Perhaps the mention of the bell activates the representation of the bell, and the mention of the shock activates a representation of shock. This contiguous activation might foster the formation of a link between these two representations (mediated learning, Holland, 1990). Of course, this theory is easily refuted; verbal instructions that “on none of the following trials will the bell be followed by shock” activate the bell and shock representations in the same way, but will not produce an anticipatory response.

Perhaps knowledge in propositional form creates CS-US links in some way that we have not yet considered. However, even if this translation process were possible, there is a deeper problem with this general idea. Proponents of the dual-system approach would like to argue for a distinction between the acquisition of conscious propositional knowledge on the one hand and automatic learning on the other. Allowing that a single verbal instruction might produce a link between two representations of the same kind as does the experience of multiple training trials seems to blur this distinction. Remember that, in their analysis of causal learning, the dual-system theorists also argue that the links formed by the automatic system can generate propositional knowledge. Taken together, these two ideas suggest that all propositional knowledge is immediately translated into links, and all knowledge in the form of links can be translated into propositional form. One of the two systems is, therefore, redundant. The only coherent solution to this problem is to assume that there is a single system, and the evidence presented here suggests that this system is propositional in nature. The experiments presented in the following section, concerning the effects of abstract rules and deductive reasoning in conditioning, lend further support to this conclusion.

4.4 Abstract rules and deductive reasoning

Shanks and Darby (1998) reported a striking demonstration of the use of rules in associative learning. They presented A+, B+, AB-, C-, D-, CD+ trials together with I+, J+, M-, and N- trials. During a test phase, participants judged that the outcome was more likely to occur after the (previously unseen) compound MN than after the (also previously unseen) IJ compound. In terms of links between representations, this is the reverse of the prediction based on the elements that made up the compounds. Participants appeared to have learned a rule from observing trials on which cues A-D were presented, that the outcome of compounds of two stimuli (i.e., AB-, CD+) is the reverse of the outcome of the individual elements that make up that compound (i.e., A+, B+, C-, D-). They then applied this reversal rule to cues I-N.

Other evidence for the role of propositional reasoning in human associative learning comes mainly from studies on cue competition, in particular, blocking (see De Houwer, et al., 2005, for review). For example, De Houwer, Beckers and Glautier (2002) observed blocking only when it was possible to infer deductively that cue T in the A+/AT+ design was not associated with the outcome. Because T does not add anything to the effect of A alone (i.e., the outcome was as probable and as intense on A+ trials as on AT+ trials), it can be inferred that T is not a cause of the outcome. However, De Houwer et al. argued that this inference is valid only if it is assumed that the effect of two causes is additive (that when two causes are presented in compound, a larger than normal effect will be produced). De Houwer et al. provided one group of participants with an alternative explanation for why T did not add anything to the effect of A. They told these participants that A alone already caused the outcome to a maximal extent. That is, the outcome was at ceiling on A+ trials. In this case, participants can reason that no increase in the effect was seen on AT+ trials, not because T was non causal, but because an increase in the size of the effect was impossible. In line with the idea that blocking is based on propositional reasoning, no blocking effect was found in this condition (causal ratings of T were not reduced as a consequence of prior A+ trials).

Many other studies have confirmed this result. Beckers, De Houwer, Pineno, and Miller (2005; see also Lovibond, Been, Mitchell, Bouton & Frohardt, 2003) raised doubts in their participants minds about the inference underlying blocking by giving pretraining in which the effect of two cues was shown to be subadditive (i.e., G+, H+, GH+, and I++, where + stands for a US of low intensity and ++ for a US of high intensity). Blocking was significantly smaller after this type of pretraining than after pretraining that confirmed the additivity of causes (i.e., G+, H+, GH++, I+). Mitchell and Lovibond (2002), using a similar approach, showed blocking of skin conductance CRs only when blocking was a valid inference. Finally, Vandorpe, De Houwer, and Beckers (2007) obtained the same result in a causal judgment study that involved a very complex design. This is important because dual-system theorists often argue that the link-formation mechanism will be revealed in very complex tasks such as that used by Vandorpe et al. (see the discussion above concerning cognitive load, Section 4.2), and so the propositional system is unable to operate or is offline (e.g., Dickinson, 2001; Le Pelley et al., 2005). Vandorpe et al.'s results showed, however, that propositional reasoning processes can operate even in these complex tasks.

4.5 Conclusions

Many experiments, using a wide range of procedures, have shown a concordance between associative learning and contingency awareness. Furthermore, results of experiments in which a secondary task was imposed are consistent with the operation of a cognitively demanding reasoning process, especially in the case of blocking. Thus, manipulations that prevent reasoning also prevent the learning mechanism from operating. Many more experiments have demonstrated the impact of verbal instructions, rules and deductive reasoning processes on the acquisition of associative knowledge. These data make a very strong case for the idea that associative learning is based on reasoning processes that yield conscious propositional knowledge.

Of course, the dual-system approach cannot be said to be inconsistent with these findings, because it incorporates both the link formation and propositional reasoning systems. However, what is important is that, within the dual-system account of the data outlined above, the link mechanism itself is redundant. We now turn to the evidence that has been argued to provide unique support for the link-formation mechanism.

5 Evidence for the automatic formation of links

Dual-system theorists point to a number of sources of evidence that they believe provide unique support for link formation models. Firstly, although associative learning is generally accompanied by awareness of the CS-US contingency, there are two learning procedures that do seem to provide some evidence of unaware conditioning (see Lovibond & Shanks, 2002). These are evaluative conditioning and Perruchet's (e.g., 1985) findings relating to the effects of trial sequence in partial reinforcement schedules. Secondly, some experiments have demonstrated learning that is not always rational (or normative). The absence of rationality has been argued to support the idea that learning can result from an automatic link mechanism. Lastly, some theorists have made the point that some neuroscientific data suggest the existence of a multiple learning systems. We address these lines of evidence in turn.

5.1 Unaware associative learning

In evaluative conditioning research (see De Houwer, Thomas, & Baeyens, 2001; De Houwer, 2007, for reviews), neutral stimuli (across a range of modalities) have been shown to increase or decrease in rated pleasantness as a consequence of pairings with strongly liked or disliked stimuli. Some researchers have provided evidence for evaluative conditioning in the absence of awareness (Baeyens et al., 1990; Dickinson & Brown, 2007; Fulcher & Hammerl, 2001; Walther & Nagengast, 2006; and see Stevenson, Prescott & Boakes, 1998, for a related finding). However, insensitivity of testing procedures and aggregating awareness scores across both participants and items may have hidden some contingency awareness in these studies (see Lovibond & Shanks, 2002, for a review). An example of this second issue can be seen in Dickinson and Brown (2007). They found that their participants, when analysed as a single group, did not demonstrate reliable contingency awareness but did show evaluative conditioning. However, Wardle, Mitchell, and Lovibond (2007) reanalysed these data and found that when participants were divided into two groups, aware and unaware, it was only the aware group that produced a reliable conditioning effect. Other researchers have suggested an even more fine-grained analysis. They have argued that, although participants might show very little contingency awareness when the cues are aggregated, they are, nevertheless, aware of the outcomes that a subset of cues were paired with. It is possible that it is this subset of cues that are responsible for the evaluative conditioning observed in earlier studies (Pleyers, Corneille, Luminet & Yzerbyt, 2007).

It is very difficult to provide a satisfactory demonstration of unaware conditioning simply by showing conditioning in the absence of awareness. This is because it is very difficult to be sure that the awareness measure and the conditioning measure are equally sensitive. Lovibond and Shanks (2002) identified Baeyens et al.'s (1990) finding as being the most convincing evidence of unaware evaluative conditioning, because flavor-flavor conditioning was seen in the absence of any contingency awareness, but color-flavor conditioning was not seen despite awareness of the color-flavor contingency. The latter finding appears to confirm that the awareness measure used was sensitive (albeit to contingencies involving different stimuli). Thus, participants in the flavor-flavor condition appear to have been unaware of the contingencies they were exposed to. Given the uniqueness of this finding, it is important that Baeyens et al's design is replicated, perhaps with the awareness measure used by Dickinson and Brown (2007), and that the awareness-learning relationship is analysed at the item level. Even more convincing than Baeyens et al's (1990) dissociation would be a demonstration of conditioning in participants unaware of the flavor-flavor contingencies, but not in participants aware of those same contingencies (rather than color-flavor contingencies). This is exactly the reverse association (see Dunn & Kirsner, 1988) sought by Pierre Perruchet in his analysis of eyeblink conditioning and cued reaction time learning. It is to this work that we now turn.

Perruchet (1985) exposed participants to a pseudo random series of tone-air puff and tone alone trials and measured both eyeblink CRs and expectancy that an air puff would be delivered on the following trial (tones appeared on every trial). Participants' self-reported expectancy of an air puff followed the gambler's fallacy. Thus, after a run of 3 tone-air puff trials, participants tended to predict that the tone would not be followed by an air puff on the next trial. Conversely, after a run of 3 tone alone trials, an air puff was strongly predicted to follow the tone on the next trial. The eyeblink CR, however, followed the opposite pattern; eyeblinks to the CS were most likely to be observed on trials following a run of tone-air puff trials and least likely following a run of tone alone trials. Thus, recent CS-US pairings appeared to strengthen the CS-US link and increase the probability of the CR, despite a reduction in US expectancy. Perruchet has more recently observed the same dissociation using a simple cued reaction time task (Perruchet, Cleeremans & Destrebecqz, 2006).

Perruchet's dissociations between US expectancy and the occurrence of the CR in eyeblink conditioning (and the equivalent effect in the cued reaction time task) are certainly intriguing. However, the findings are somewhat peculiar and are open to alternative interpretation. They are peculiar in the sense that the dissociation is not really between contingency awareness and the observation of the response (CR or reaction time). Participants know the contingency from the start of the experiment and the training trials confirm this; the tone will be followed by the US on 50% of trials. The effect observed seems to be much more a performance effect. Furthermore, the recency of CS-US pairings is perfectly confounded with recency of US presentations in this experiment. The observed fluctuation in the CR may, therefore, be due to sensitization produced by US recency alone, and not an associative phenomenon at all. Perruchet's own experiments (see also Weidermann, Tangen, Lovibond, & Mitchell, in press) go some way to ruling out this alternative explanation, but further work remains to be done. Despite these issues, Perruchet's gambler's fallacy effect remains the strongest available evidence for dissociation between a CR and the conscious expectancy of a US.

5.2 Rationality

It is often assumed that rationality is a hallmark of the propositional system. If behavior is rational, then a propositional mechanism was in operation, if it is not rational, an automatic mechanism was in operation (Shanks and Dickinson, 1990; Shanks, 2007). Therefore, if it can be shown that associative learning is non-rational, it must be based on the automatic formation of links. The example of irrational behavior that most readily comes to mind is phobia. For example, spider phobics can be fearful of spiders despite claiming to know that spiders are not harmful. This would appear to undermine the idea that learning is a propositional process – how could such a system produce behavior that contradicts the verbally reported belief?

There are three ways that the irrational behavior of spider phobics can be explained, each of which is consistent with the propositional approach to learning. Firstly, the verbally reported belief that spiders are not harmful may simply be a consequence of social demands; the patient may believe the spider to be harmful but not wish to contradict the clinician's view that the spider is harmless. Secondly, this phenomenon may relate to performance, not to learning. The patient may have a long standing and strong belief that spiders will do him or her harm. He or she may also have acquired more recently a perhaps more fragile appreciation that certain spiders are not harmful. On presentation of a harmless spider, the old belief that spiders are harmful may be retrieved automatically from memory and thus lead to fear (see Section 3.2). Because the retrieval of the old belief occurs automatically, the resulting fear might seem irrational and difficult to control. According to the propositional model, both beliefs (that the spider is harmful and that it is not harmful) will have been acquired through a process of propositional reasoning. Lastly, there is, in fact, little evidence that specific phobias of this kind result from learning at all, and thus they may have a genetic etiology (see Menzies & Clarke, 1995, for review). If fear of spiders has a large genetic component that affects behavior independently of learning, the fact that fear remains even when it is known that spiders are not harmful does not represent a challenge to the propositional approach to associative learning.

Nevertheless, there are examples of what appears to be irrational associative learning. Karazinov and Boakes (2007) trained participants on a causal learning task with a conditioned inhibition design (X+/XT-). Thus, X was followed by the outcome when presented alone (X+) but not when it was presented in compound with the target cue (XT-). This training can give rise to inhibition; presentation of T has the ability to reduce the causal attribution to another exciter, Y, on test. This seems to be a rational inference because T prevented the outcome produced by X in training, and so might prevent the outcome that would otherwise have been produced by Y on test. Karazinov and Boakes (2007) found the reverse effect, however, when participants were given little time to think during training. Thus, participants did not learn that T prevented the outcome, but appeared to learn that it caused the outcome. Karazinov and Boakes concluded that participants did not have time to reason about the relationship between T and the outcome, and so their behavior was the result of the automatic formation of a (second order) link between T and the outcome (or between T and the response of giving a high causal rating).

There are other related findings in the literature. For example, Le Pelley et al. (2005) paired cue A with two outcomes (A-O1O2) in a first phase of training and found blocking following a second phase in which cue T was added (AT-O1O2); pretraining with A reduced the degree to which an association between T and the two outcomes was learned. This blocking was disrupted, however, when one of the outcomes changed in the second phase (AT-O1O3). Not only did participants learn to associate T and O3 (they failed to show blocking with respect to the outcome not predicted by A), but also T and the unchanged outcome, O1. Le Pelley et al. (2005) argued that, because learning an association between T and O1 is not rational (O1 is predicted by A), and was not observed in a much simpler version of the task, the learning of T-O1 association must be due to a non-rational, automatic mechanism.

Shanks (2007) presented the following phenomenon as the most compelling evidence of an irrational link-formation mechanism in the context of contingency learning. In one condition, the probability of the outcome in the presence of the cue (P(O/C)) was 0.75, and the outcome did not occur in the absence of the cue (P(O/~C) = 0). In the other condition, the probability of the outcome both in the presence and in the absence of the cue was 0.75. Thus, although the probability of the outcome following the cue was equivalent in both cases (0.75), the outcome was contingent on the cue in the first condition, but not in the second. It has been found that judgments of the probability that the outcome will follow the cue are greater in the former case than in the latter. Thus, the cue-outcome contingency appears to have an impact on the judgment of outcome probability, despite the fact that this probability is identical in both cases (see De Houwer, Vandorpe & Beckers, 2007; Lagnado & Shanks, 2002; Lopez, Cobos, Cano & Shanks, 1998; Price & Yates, 1993). It is irrational to give a higher rating of probability when the contingency is increased but the probability of the outcome stays the same. Shanks attributed these higher probability ratings to the formation of links between cues and outcomes that have a contingent relationship.

We agree that these are very interesting findings, and each suggests that our reasoning abilities are sometimes not optimal. However, we do not think that these findings provide evidence for an automatic link-formation mechanism. The irrational behavior observed can equally be attributed to sub-optimal operation of the reasoning system[1]. In each case, an explanation for the behavior can be given that is consistent with the propositional approach. For example, when given little time to ponder over the implications of seeing X+ and XT- trials, perhaps Karazinov and Boakes' (2007) participants mistakenly thought that T might somehow signal the presence of X, which itself caused the outcome. Such an inference would lead to the conclusion that T itself might be associated with the outcome to a greater extent than the control cue. Perhaps Le Pelley et al.'s (2005) participants knew that something about the outcomes had changed between A-O1O2 trials and AT-O1O3 trials, but could not remember exactly what had changed. As a consequence, they may have concluded that it was safest to assume that T caused O1 and O3 equally.

Finally, in the studies Shanks (2007) refers to, participants may merely have been confused about the meaning of the term "probability" in the test instructions. It is not at all obvious that participants would readily distinguish between probability and contingency in the way that the experimenters did. Alternatively, participants in the non-contingent condition probably assumed that there existed some other cause of the outcome. Then, on test, they may have thought that the experimenter was asking about the probability of the outcome following the cue, but in the absence of any other potential causes. That is, an assumption may have been made that the cue was presented in a different context on test.

These alternative explanations might be argued to be somewhat far fetched. However, they are presented only to demonstrate that irrational behavior is not inconsistent with the operation of an imperfect propositional reasoning system cooperating with an imperfect memory system. It might also be argued that this position leaves the propositional approach untestable. This is not so. Firstly, one can test propositional explanations of irrational behavior empirically. For instance, if Le Pelley et al.'s (2005) finding is due to confusion as to which outcome changed between the two phases of training, increasing the distinctiveness of the two outcomes should reduce the unblocking effect with respect to O1. If the impact of contingency on probability judgments featured by Shanks (2007) depends on confusion about the instructions given on test, then the effect should be reduced in magnitude if these instructions leave less room for misunderstanding. Also, presenting the test question in terms of frequency ("you will see 10 further trials on which the cue will be present, on how many will the outcome occur?") rather than probability should reduce the size of the effect (see Gigerenzer & Hoffrage, 1995, for an example of frequency formats reducing base rate neglect). If, on the other hand, the participants assumed that the test context was different from the training context, then making it explicit that the cue was presented in the same context on test should eliminate the effect. Secondly, and more importantly, evidence that participants are not always rational when they learn does not undermine the main predictions of the propositional approach; that learning will occur only when participants are aware of the cue-outcome (or CS-US) contingencies, will be disrupted by secondary tasks and will be affected by verbal instructions, rules and deductive reasoning processes.

5.3 Dissociable systems within the brain

One could argue that a dual-system approach is supported by neurological data showing that different brain regions are involved in different types of learning. These different brain regions could be seen as the neurological basis of different learning systems. For example, there is now abundant evidence that the amygdala plays an important role in, for example, fear learning (e.g. Le Doux, 2000; Öhman & Mineka, 2001). A quite different area of the brain, the cerebellum, has been shown to be important in conditioning of the nictitating membrane (Thompson, 2005). Thus, based on such neuroscientific dissociation data, it might be argued that the amygdala is part of a fear learning system that is quite separate from the system responsible for nictitating membrane conditioning.

This conclusion is, however, not necessarily correct (see Henson, 2006, for a detailed discussion of the validity of theoretical inferences based on neuroscientific dissociation data). One alternative interpretation is that neither the amygdala nor the cerebellum is able to produce learned behavior alone, but that they operate as individual components in a coordinated learning system. For instance, these brain regions might be important in processing specific kinds of stimuli or generating specific kinds of responses rather than being responsible for the learning process as such. Thus, the *learning* may take place neither in the amygdala nor cerebellum but in another part of the brain entirely, or, indeed, in many parts of the brain simultaneously. A related argument can also be applied to the idea that the striatum and its dopaminergic afferents are responsible for habitual behaviour (Jog, Kubota, Connolly, Hillegaart & Graybiel, 1999) but prefrontal areas are responsible for higher level cognition. Again, these dissociations seem to imply separate learning systems. However, they may simply reflect a single learning system solving problems of differing complexity or concreteness (see Chater, in press).

Thus, although there can be no doubt that recent advances in the neurosciences have provided a wealth of knowledge about the brain mechanisms necessary for learning, these findings are not inconsistent with the single-system view of learning . Furthermore, the available behavioural evidence concerning human associative learning does not support the view that there are multiple learning systems. The behavioural evidence, therefore, presents a challenge to neuroscientists to discover how a single, integrated, propositional learning system with multiple subcomponents might be implemented in the brain.

5.4 Conclusions

To summarize the data presented in the present section, it would appear that two or three studies provide support for the link formation mechanism. These are demonstrations of the Perruchet effect (Perruchet, 1985; Perruchet et al., 2006) and perhaps one example of flavor-flavor evaluative conditioning (Baeyens et al., 1990). It is important, therefore, that these findings are subject to the closest empirical and conceptual scrutiny in the future. Findings that provide evidence for irrational learning should also be studied further, but they do not provide direct evidence against the propositional approach. Lastly, it is not at all clear that evidence from studies of the brain can inform us as to the existence of distinct learning systems. Overall, therefore, we see no reason to postulate the existence of a link formation system in addition to a propositional reasoning system.

6. Conceptual arguments

There are a variety of reasons why the link mechanism has been so popular as an explanation for associative learning, even in the absence of strong supporting data. In the present section we discuss three of these reasons: (1) the learning models developed within this traditional approach (e.g. Rescorla & Wagner, 1972) seem parsimonious; (2) mental links, and the way they increase and decrease in strength, provide a very intuitive analogy for neural plasticity; (3) researchers are resistant to the idea that non-human animals engage in propositional reasoning. We will evaluate the relative strengths and weaknesses of the propositional and link approaches with regard to these conceptual issues.

6.1 Simple models of learning

The first and perhaps strongest reason for learning theorists' adherence to the idea of a link-formation mechanism is that a range of very tightly specified theories have been developed within this approach. Theories such as those proposed by Mackintosh (1975), Pearce and Hall (1980), Rescorla and Wagner (1972) and Wagner (1981) are formalized, can be simulated on a computer, and can, therefore, make precise and testable predictions. The power of these models comes from the fact that they often make few assumptions but apply to a wide range of phenomena. For this reason, it could be argued that these models are preferable to the propositional approach to learning.

The first thing that needs to be pointed out is that the precision of the predictions of associative models from the link-formation tradition is somewhat overstated. A lot depends on the particular parameter values and the particular model variant from which the predictions are derived. In fact, from experience we have learned that it is difficult to produce a pattern of data that cannot be explained by one or the other variant of these associative models. For example, one can explain blocking (Kamin, 1969) and the opposite phenomenon, augmentation (Batsell, Paschall, Gleeson & Batson, 2001). One can also explain overshadowing (Pavlov, 1927) and the opposite phenomenon, potentiation (Garcia, Brett & Rusiniak, 1989). For each case of competition between cues, the opposite pattern of results can be explained by postulating links ("within-compound associations") between the stimuli that might otherwise be in competition (e.g., Durlach & Rescorla, 1980).

The notion of within compound associations is only one way in which freedom is gained to explain results that are not predicted by the formal versions of the models. Another way is to postulate different levels of generalization between cues. Schmajuk and Larrauri (2008), for instance, added such assumptions to a variant of the Rescorla-Wagner model in order to explain the finding that additivity pretraining can influence blocking (Beckers et al., 2005, see Section 4.4). To recap, blocking is the finding that little is learned about T in a design in which A+ trials precede AT+ trials. According to the Rescorla-Wagner model, blocking occurs because, on AT+ trials, the outcome is already predicted by A. Schmajuk and Larrauri argued that more blocking is seen following additivity pretraining (G+, H+, GH++, I+) than subadditivity pretraining (G+, H+, GH+, I++) because learning about GH during pretraining generalizes to later AT+ trials. In Beckers et al.'s (2005) experiment, the AT compound can be expected to acquire more generalized associative strength from GH following GH++ pretraining (the additive group) than following GH+ pretraining (the subadditive group). This is because the associative strength of GH is higher in the additive group. In other words, participants expect the outcome to a larger extent at the start of AT+ trials in the additive than in the subadditive group. It follows from the Rescorla-Wagner model, therefore, that less can be learned about the T-outcome relation (more blocking will be observed) in the additive group.

There are two problems with this alternative explanation. Firstly, Schmajuk and Larrouri (2008) focus on generalization between compounds (e.g., GH and AT). However, generalization between elements is ignored, as is generalisation from compounds (e.g., GH) to elements of those compounds (e.g., G). Hence, Schmajuk and Larrouri (2008) can explain the results of Beckers et al. (2005) only by choosing very specific and selective parameters of generalization. It is not clear whether the model would still be able to explain the findings of Beckers et al. when more realistic assumptions are made about generalisation between different kinds of cue.

Secondly, as Schmajuk and Larrauri (2008) admit, the explanatory power of their model is limited. There are, for example, other experiments presented by Beckers et al. (2005) that the model is unable to account for, such as the effects of additivity on backward blocking, in which AB+ training is given before A+ training. To explain these data, further assumptions would be required. Elsewhere in the literature there are other similar effects that this model cannot account for. For example, in a similar experiment to that of Beckers et al. (2005), Mitchell, Lovibond and Condoleon (2005) showed that G+, H+, GH- pretraining (subtractivity) can also produce a strong blocking effect. In this case, the compound of two causal cues in pretraining (G+ and H+) was non-causal (GH-). Schmajuk and Larrauri's (2008) model cannot account for blocking in this case; it predicts very little blocking here, because the GH compound acquires no associative strength in pretraining. In contrast, the propositional approach provides a straightforward explanation for the strong blocking seen in both Mitchell et al's subtractivity condition, and Beckers et al.'s (2005) additivity condition. Participants in both of these conditions can reason that T was non-causal because the AT compound did not produce a *different* outcome (either smaller or larger) from that observed when the A cue was presented alone.

The conclusion from the examples above seems clear. While individual models such as the Rescorla-Wagner model are quite parsimonious, the entire class of theories that are assumed to describe the way in which links are formed is not. Although extending models in a posthoc manner is not, in principle, problematic, the evaluation of the extended model against only a single data set (for which that extension was specifically designed) is dangerous. The generalizability of the new model to other data sets must be demonstrated, otherwise there is a risk that a different link-based model will be generated posthoc to account for each observed experimental result.

There is also another issue related to parsimony. In order to account for our manifest ability to, for example, solve problems and play chess, traditional learning theorists must supplement the link-formation system with a system that forms propositions on the basis of reasoning. As we argued above, these theorists are calling for a dual-system approach. No approach that needs two systems can be more parsimonious than an approach that proposes only one of those systems, no matter how parsimonious the second system might be.

Nevertheless, the apparent precision and parsimony of traditional learning models might be an important reason why many researchers are not ready to give up these models. It is important to realize, therefore, that adopting a propositional approach does not imply that one must give up traditional models of learning. The propositional approach is not an alternative to specific learning models such as the Rescorla-Wagner model (or any of its relatives) but to the dual-system approach that postulates an automatic link formation mechanism. We can clarify this argument using Marr's (1982) distinction between functional and algorithmic levels of explanation. Both functional and algorithmic models make predictions about which pattern of input (e.g., learning trials) leads to which pattern of output (e.g., CRs or causal ratings). Only algorithmic models, however, incorporate assumptions about the processes and representations that translate the input into the output. That is, models at the algorithmic level make assumptions about *how* the stimulus input is processed to produce the output. The propositional approach and the automatic link formation mechanism are thus clearly explanations at the algorithmic level because they do incorporate (different) assumptions about how the input is processed to produce the output (i.e., controlled reasoning vs automatic link formation and activation) and about the nature of the representations over which these processes operate (i.e., propositions vs links between stimulus representations).

Many individual models of associative learning, however, can be regarded as functional models. Take the example of the Rescorla-Wagner model. In essence, this is a mathematical formula that allows one to predict whether a CR will be observed given information as to the nature of the learning trials experienced. Hence, it is a functional model. It is not an algorithmic model because Rescorla and Wagner (1972) do not commit to a particular type of underlying process. Their model was developed to account for *what* is learned under certain conditions. This can be contrasted with models at the algorithmic level that give an account of *how* this learning takes place. In fact, Rescorla and Wagner are explicitly agnostic about algorithmic level explanations (that is, how organisms learn, and therefore why they behave according to the Rescorla-Wagner model). They offer two quite different algorithmic level explanations, one in the language of links and another in terms of the constructs of expectancy and surprise. Thus, when the Rescorla-Wagner model is tested against other models such as the Pearce-Hall model, it is the fit of the mathematical formulae to the behavior that is being tested (i.e., predictions at the functional level), not the nature of the underlying processes or representations (e.g., automatic formation of links or propositional reasoning). From this perspective, a functional model such as the Rescorla-Wagner model is not incompatible with the propositional approach because the two can be seen as focussing different levels of explanation.

In fact, from this point of view, the Rescorla-Wagner model can even be thought of as a simple mathematical model of propositional reasoning, not, as is usually assumed, a model of link formation. At the functional level, it captures many of the operating principles of propositional reasoning. To take one simple example, a belief is most likely to change when it is demonstrated to be wrong. That is, when the belief leads to an expectancy that is violated. The Rescorla-Wagner model captures the essence of this idea; according to this model, learning only takes place (or beliefs only change) when the outcome on a learning trial is not predicted (i.e. that outcome is surprising).

Lastly, it is interesting that so many learning models developed since the 1960s include constructs such as limited capacity working memory, selective attention and interference in memory (Bouton, 1993; Mackintosh, 1975; Pearce & Hall, 1980; Wagner, 1981). We would argue that these constructs describe much more naturally the operation of controlled cognitive processes of propositional reasoning operating in cooperation with the memory system than they do the automatic formation of links.

6.2 There are links in the brain

A second reason for the continuing success of the link formation mechanism is that the idea of a link between mental representations that can increase or decrease in strength is a very powerful analogy for links between neurons in the brain. When associative learning theorists think in terms of the mental link between representations, there seems no doubt that this mechanism feels more real by virtue of its similarity to the hardware in which it must be implemented. However, there are two problems with this claim. Firstly, this implicit reductionism loses all of its force when it is considered that the dual-system approach also postulates complex propositional reasoning capacities that cannot be explained (at least at the present time) in terms of links between representations. These more complex capacities must also be implemented in the brain. Within the dual-system approach, therefore, both systems must have strong (and equal) neural plausibility. Secondly, although a link between a CS and US representation might resemble two connected neurons in the brain, mental representations are not identical to neurons and links are not identical to dendrites. Representations and links between representations are unobservable theoretical constructs. They are invented by psychologists in order to help understand behavior at an algorithmic level. In that sense, they are no more neurologically plausible than other theoretical constructs such as propositional representations.

A very similar argument applies to the success of parallel distributed processing (PDP) models as support for the link formation approach. In PDP models, structures with properties very similar to a collection of interconnected neurons are simulated within a computer. The strengthening of links within such PDP models is very similar to the strengthening of dendrites between neurons. Thus, both PDP models and neurological structures are structures (simulated in the computer or present in the brain) in which algorithmic processes can be implemented (see Marr, 1982). The link model described in Figure 1 is quite different from these PDP models, just as it is different from structures in the brain. This is because, in Figure 1, links are formed between nodes that each represent a stimulus in a symbolic manner (i.e., the CS and US). In contrast, a single node in a PDP model does not represent anything, just as a neuron in the brain does not represent anything.

In PDP models, representations are an emergent property of the network, and correspond to particular patterns of activation across a number of nodes. PDP models thus offer a way to implement representations of stimuli and relations in a nonsymbolic, distributed manner. It is certainly true that the link model in Figure 1 is one possible algorithmic-level model that can be implemented in a PDP network. But models of highly complex cognitive abilities, such as propositional reasoning, can, in principle, also be implemented within PDP models, just as they are in the brain.

In summary, the idea of a link can be used in many different ways, and it is important that these different uses are not confused. In this section we have distinguished between links at the implementational level (neurons and PDP networks) and the idea that links form between representations, which is a model at the algorithmic level (see Figure 1). The present paper does not focus on the implementational level. Rather, we aim to distinguish between two algorithmic models of associative learning, one in which links are automatically formed that transmit excitation between representations, the other in which beliefs are formed, as a consequence of controlled processes, about the relationship between the events. We would argue that both the dual-system approach (incorporating the automatic link-formation mechanism) and propositional approaches are equally consistent with a link-based implementation such as a PDP model or, indeed, the brain.

6.3 Propositional reasoning in non-human animals

Although our subject matter here is human learning, we would not want to argue that humans possess a unique cognitive learning system. This stance implies that non-human animal learning is also a process of belief acquisition. Thus, the complex representational system we possess evolved from similar, but simpler, cognitive systems in our ancestors, and many differences observed between human and non-human learning are quantitative, not qualitative.

We have argued that learning is the consequence of an interaction between propositional reasoning and memory for past events. There is also evidence for primitive versions of these abilities in non-human animals. For example, Clayton and Dickinson (1998) have demonstrated episodic-like memory in scrub jays. There is also some evidence to support the idea that rats are able reason about cause and effect (Beckers, Miller, De Houwer & Urushihara, 2006; Blaisdell, Sawa, Leising & Waldmann, 2006). For example, Beckers et al. (2006) followed De Houwer et al. (2002) and Mitchell and Lovibond's (2002) approach to the demonstration of propositional reasoning in blocking, but they used rat subjects. Beckers et al.'s data closely paralleled those found with human participants. This supports the idea that rats engage in propositional reasoning. If propositional reasoning abilities underlie associative learning in humans, and these abilities are shared (perhaps in a primitive form) by other species, then it is not unreasonable to suggest that propositional reasoning may also be responsible for associative learning in non-human animals. Whatever the merits of this view, one should at least be open to the possibility that learning in animals is not always based on an automatic link formation mechanism but could also result from other, reasoning-like processes.

Of course, there must be limits to this line of argument. In the extreme case, surely invertebrates such as aplysia do not have conscious beliefs. Indeed, we would agree that it would not be useful to apply the propositional approach to aplysia. Rather than representing the two events and the relationship between them, such that one event leads to anticipation of the second event, aplysia simply learn to respond to a particular stimulus. That is, a stimulus-response (S-R) relationship is learned by which a certain input leads to a certain response in a reflexive manner and thus without the involvement of mental representations (see Moors, 2007). However, humans, and many other animals, have in the course of evolution been endowed with a more flexible system that allows responding to be more contextually appropriate. For example, the more sophisticated system is, unlike an S-R mechanism, sensitive to changes in the reinforcement value of the outcome (e.g, Adams & Dickinson, 1981). This is because the mental representations of the events and their relationship intervene between the stimulus and the response. In other words, we suggest that humans have cognition and aplysia do not. Between these two extremes lies a continuum of cognitive complexity. Animals with more sophisticated cognitive abilities use these abilities to learn about their environment, so that they can, to the greatest extent possible, adapt their behavior when that environment changes. It would now appear that at least certain non-human animals have cognitive capabilities that go beyond the simple automatic formation of links. It would be surprising if those capabilities were not utilized in the process of learning to adapt to and control the environment.

One last important point is that all human and non-human animals (including aplysia) also display plasticity at the neural level. Within all species, reflexive (and thus cognitively unmediated) learning can be observed at the neural level. This reflexive type of learning, however, falls beyond the scope of both the propositional *and* the link formation approach. As indicated earlier, these approaches operate at the algorithmic level, that is, at the level of psychological processes and representations. Neither approach operates at the implementational (in this case neural) level.

7. Implications for the lab and clinic

In this section we see how the present proposal fits with the way that psychology has changed over the past half a century, both from a theoretical and an applied (clinical) perspective.

7.1 The cognitive revolution

The received view is that behaviorism (more particularly S-R theory) gave way to the cognitive revolution in the mid 1950s at the time when the computer was invented and when a number of findings were published for which no parsimonious S-R account could be provided (see Gardener, 1985 for a review). Within learning research, it became clear that many phenomena, such as sensory preconditioning, blocking, and reinforcer devaluation, could not be explained in S-R terms and were better explained by a model in which an internal representation of the CS was connected to an internal representation of the US (a stimulus-stimulus (S-S) link; see Dickinson, 1980; Mackintosh, 1974). An S-S model is a giant leap towards a fully-fledged symbolic system, because it postulates that associations between stimuli in the environment are represented by links between internal representations of those stimuli.

In 1973, Seligman and Johnson published a cognitive or expectancy-based theory of instrumental learning. On this theory, if a rat presses a lever to obtain food (or avoid shock), it does so because it desires food (or wishes to avoid the shock) and believes that the lever press will produce that outcome (see Dickinson, 1989). However, Seligman and Johnson maintained the view that Pavlovian conditioning results from an automatic mechanism, in which links form between between the CS and US representations. Resistance to the idea that Pavlovian conditioning is the result of the same processes as instrumental conditioning (that is, processes of belief acquisition) continues to the present day.

In this context, the view presented in the present paper should not be seen, as it no doubt is by the majority of psychologists, as an example of extremism. Rather, the belief-based approach to S-S learning is merely a small step in the same direction that we have been heading for the past 50 years; away from S-R learning theory and towards a propositional approach to all learning. Furthermore, as argued above, adoption of the propositional approach does not imply that the important insights gained from research conducted within a behaviorist approach, or within the more recent S-S approach are to be discarded. It is merely that learning theorists have been mapping out the properties, not of the mechanisms that form S-R or S-S links, but of the propositional reasoning processes that result in learning.

7.2 The clinic

The propositional approach is consistent with developments in clinical psychology over the past 20 years. It is now commonly proposed that patients display false or exaggerated beliefs and distortions in reasoning that contribute to their symptoms and maladaptive behavior. For example, anxious patients overestimate the probability and cost of future harm, and patients with anorexia perceive their bodies to be overweight (e.g., Clark, 2004).

Early "cognitive-behavioral" interventions were based on a dual process model of learning (Zinbarg, 1990). They assumed that "behavioral" techniques like reinforcement and extinction worked on unconscious automatic responses whereas "cognitive" (verbal) techniques worked on consciously available beliefs. However, more recent (and more effective) cognitive-behavioral interventions feature a closer integration of experience and language, and hence are more consistent with the propositional approach to learning. For example, exposure to interoceptive sensations (e.g., breathlessness, pounding heart) in panic disorder is used explicitly as a way of testing the patient's catastrophic interpretations (e.g., heart attack) and is linked to verbal information concerning the true causes of those sensations (e.g., hyperventilation, anxiety). Thus, direct experience and language can be seen as two different and potentially synergistic ways of targeting patient's distorted beliefs and thus normalising their behavior (Lovibond, 1993). Further exploration of the ways in which learning experiences impact on propositional knowledge may well facilitate progress in developing effective clinical interventions.

8. Conclusion

Within the propositional approach presented here, learning is not separate from other cognitive processes of attention, memory and reasoning, but is the consequence of the operation of these processes working in concert. There is, therefore, no automatic mechanism that forms links between mental representations. Humans learn the causal structure of their environment as a consequence of reasoning about the events they observe. For example, when a bell is followed by food on a number of occasions, it is inferred that the bell stands in some predictive or causal relationship to the food. Therefore, food will be expected the next time the bell rings. Later ringing of the bell will then generate the belief that food presentation is imminent and so will produce salivation.

The available evidence largely supports the propositional approach to learning. Thus, learning does not take place outside of awareness, it requires cognitive resources and it is affected by verbal instructions, rules and deductive reasoning processes. There are some fragmentary pieces of evidence that seem to indicate a role for a second, automatic mechanism in anticipatory learning, most particularly the dissociation between outcome expectancy and conditioned responding shown by Perruchet (1985). This evidence is, however, far from conclusive. It would seem unwise at this point to base a belief in a dual-system theory of learning on evidence from a very small number of experiments that is yet to be properly evaluated.

If, as the propositional approach suggests, the human cognitive system is a more complex version of a similar system possessed by non-human animals, then animal models of human functioning would no longer be restricted to a narrow range of "associative" phenomena. We may then see animal models of reasoning or attentional control, which may, in turn, lead to the development of drug therapies for deficits in these areas. In the same vein, a single coherent approach could be developed for the treatment of learning-based clinical problems.

There are, therefore, many applied benefits of this new approach. However, fundamentally, what we propose is a change in the way we think about our basic research in learning. The postulation of automatic mechanisms of link formation is pervasive in psychology; the links are used to explain phenomena as disparate as simple conditioned responding and the formation of attitudes to members of an out-group. The propositional approach suggests that these phenomena should be reinterpreted to be the consequence of propositional reasoning leading to the acquisition of new beliefs.

References

Adams, C. D. & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 33B*, 109-121.

Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning. *Cognition & Emotion, 4*, 3-18.

Batsell, WR., Paschall, G.Y., Gleason, D.I. & Batson, J.D. (2001). Taste preconditioning augments odor-aversion learning. *Journal of Experimental Psychology: Animal Behavior Processes, 27,* 30-47.

Beckers, T., De Houwer, J., Pineno, O., & Miller , R.R. (2005). Outcome additivity and outcome maximality influence cue competition in human causal learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31,* 238-249.

Beckers, T., Miller, R. R., De Houwer, J. & 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.

Blaisdell, A., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science, 311*, 1020-1022.

Boakes, R. A. (1989). How one might find evidence for conditioning in adult humans. In T. Archer & L.-G. Nilsson (Eds.), *Aversion, avoidance and anxiety: Perspectives on learning and memory* (pp. 381-402). Hillsdale, NJ: Erlbaum.

Bolles, R.C., & Fanselow, M.S. (1980). A perceptual-recuperative model of fear and pain. *Behavioral and Brain Sciences, 3,* 291-323.

Bouton, M.E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114,* 80-99.

Braine, M.D.S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*, 1-21.

Brewer, W. F. (1974). There is no convincing evidence for operant or classical conditioning in adult humans. In W. B. Weimer & D. S. Palermo (Eds.), *Cognition and the symbolic processe*s (pp. 1-42). Hillsdale, NJ: Erlbaum.

Chater, N. (in press). Rational and mechanistic perspectives on reinforcement learning. *Cognition.*

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Science, 10*, 287-291.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104,* 367-405.

Clark, D.M. (2004). Developing new treatments: On the interplay between theories, experimental science and clinical innovation. *Behavior Research and Therapy, 42*, 1089-1104.

Clayton, N. S. & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature,* 395, 272-274.

Cook, S. W. & Harris, R. E. (1937). The verbal conditioning of the galvanic skin reflex. *Journal of Experimental Psychology, 21*, 202-210.

Colgan, D. M. (1970). Effect of instructions on the skin conductance response. *Journal of Experimental Psychology, 86*, 108-112.

Dawson, M.E. (1970). Cognition and conditioning: Effects of masking the CS-UCS contingency on human GSR classical conditioning. *Journal of Experimental Psychology, 85,* 389-396.

Dawson, M. E., & Biferno, M. A. (1973). Concurrent measurement of awareness and electrodermal classical conditioning. *Journal of Experimental Psychology, 101,* 55-62.

Dawson, M. E. & Schell, A. M. (1985). Information processing and human autonomic classical conditioning. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (Vol., 1, pp. 89-165). Greenwich, CT: JAI Press.

Dawson, M. E. & Schell, A. M. (1987). Human autonomic and skeletal classical conditioning: The role of conscious cognitive factors. In G. Davey (Ed), *Cognitive processes and Pavlovian conditioning in humans* (pp. 27-55). Oxford: John Wiley & Sons.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10*, 230-241.

De Houwer, J. (in press). The propositional approach to associative learning as an alternative for association formation models. *Learning and Behavior.*

De Houwer, J., & Beckers, T. (2003). Secondary task difficulty modulates forward blocking in human contingency learning. *Quarterly Journal of Experimental Psychology, 56B*, 345-357.

De Houwer, J., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology, 55A*, 965-985.

De Houwer, J., Thomas, S. & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853-869.

De Houwer, J., Vandorpe, S., & Beckers, T. (2005). On the role of controlled cognitive processes in human associative learning. In A. J. Wills (Ed.), *New directions in human associative learning* (pp. 41-63). Mahwah, NJ: Erlbaum.

De Houwer, J., Vandorpe, S. & Beckers, T. (2007). Statistical contingency has a different impact on preparation judgements than on causal judgements. *Quarterly Journal of Experimental Psychology, 60,* 418-432.

Dickinson, A. (1980). *Contemporary animal learning theory.* Cambridge: Cambridge University Press.

Dickinson, A. (1989) Expectancy theory in animal conditioning. In S.B. Klein & R.R. Mowrer (Eds), *Contemporary Learning Theories: Pavlovian conditioning and the status of traditional learning theory (*pp. 279-308). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dickinson, A. (2001). Causal learning: An associative analysis. *The Quarterly Journal of Experimental Psychology, 54B*, 3-25.

Dickinson, A., & Brown, K. J. (2007). Flavor evaluative conditioning is unaffected by contingency knowledge during training with color–flavor compounds. *Learning & Behavior, 35,* 36-42.

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgment of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology, 36*A, 29-50.

Dunn, J.C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95,* 91-101.

Durlach, P. D. & Rescorla, R. A. (1980). Potentiation rather than overshadowing in flavor aversion learning: An analysis in terms of within-compound associations. *Journal of Experimental Psychology: Animal Behavior Processes, 6,* 175-187.

Epstein, S. & Roupenian, A. (1970). Heart rate and skin conductance during experimentally induced anxiety: The effect of uncertainty about receiving noxious stimuli. *Journal of Personality and Social Psychology, 16*, 20-28.

Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences, 7,* 454-459.

Evans, J. St. B. T. (2008). Dual-process accounts of reasoning judgment and social cognition, *Annual Review of Psychology, 59,* 6.1-6.24.

Evans, J. St. B. T., & Over, D.E. (1996). *Rationality and Reasoning*. Hove, UK: Psychology Press.

Fodor, J.A. (2003). *Hume variations*. Oxford, UK: Clarendon Press.

Fulcher, E. P., & Hammerl, M. (2001). When all is revealed: A dissociation between evaluative learning and contingency awareness. *Consciousness & Cognition, 10,* 524-549.

Garcia, J., Brett, L. P., & Rusiniak, K. W. (1989). Limits of Darwinian conditioning. In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary learning theories: Instrumental conditioning and the impact of biological constraints on learning* (pp. 181-203). Hillsdale, NJ: Erlbaum.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York, NY, US: Basic Books.

Gigerenzer, G, & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684-704.

Henson, R. (2006). Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in cognitive sciences, 10,* 64-69.

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review, 93*, 411-428.

Holland, P. C. (1990). Event representation in Pavlovian conditioning: Image and action. *Cognition, 37*, 105-131.

Jog, M.S., Kubota, Y., Connolly, C.I., Hillegaart, V., &Graybiel, A.M. (1999). Building neural representations of habits. *Science, 286*, 1745-1749.

Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge: Cambridge University Press.

Kahneman, D., & Frederick, S. (2002). *Representativeness revisited: Attribute substitution in intuitive judgment.* In T.D. Gilovich, D.W. Griffin, & D. Kahneman (Eds), Heuristics and biases (pp49-81). New York: Cambridge University Press.

Kamin, L. J. (1969). Predictability, surprise attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279-296). New York: Appleton-Century -Crofts.

Karazinov, D. M. & Boakes, R. A. (2007). Second order conditioning in human predictive judgments when there is little time to think. *Quarterly Journal of Experimental Psychology, 60*, 448-460.

Katkin, E. S., Wiens, S., & Öhman, A. (2001). Nonconscious fear conditioning, visceral perception, and the development of gut feelings. *Psychological Science, 12*, 366-370.

Kruschke, J.K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review, 113,* 677-699.

Lagnado, D. A. & Shanks, D. R. (2002). Probability judgment in hierarchical learning: A conflict between predictiveness and coherence. *Cognition,* 83, 81-112.

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154-172). Oxford: Oxford University Press.

Larkin, M. J., Aitken, M. R., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1331-1352.

Le Doux, J.E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23,* 155-184.

Le Pelley, M., Oakeshott, S., & McLaren, I. (2005). Blocking and Unblocking in Human Causal Learning. *Journal of Experimental Psychology: Animal Behavior Processes, 31,* 56-70.

López, F. J., Cobos, P. L., Caño, A., & Shanks, D. R. (1998). The rational analysis of human causal and probability judgment. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 314-352). Oxford, UK: Oxford University Press.

Lovibond, P.F. (1993). Conditioning and cognitive-behavior therapy. *Behavior Change, 10,* 119-130.

Lovibond, P. F. (2003). Causal beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 97-106.

Lovibond, P.F., Been, S.L., Mitchell, C.J., Bouton, M.E., & Frohardt, R. (2003). Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. *Memory and Cognition, 31,* 133-142.

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. (1974). The Psychology of Animal Learning. Academic Press.

Mackintosh, N. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*(4), 276-298.

MacCorquodale, K., & Meehl, P.E. (1954). Edward C. Tolman. In W. K. Estes et al. (Eds), *Modern learning theory* (pp. 177-266). New York: Appleton-Century-Crofts.

Marr, D. (Ed.). (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.

McLaren, I. P. L., Green, R. E. A., & Mackintosh, N. J. (1994). Animal learning and the implicit/explicit distinction. In N. C. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 313-332). London: Academic Press.

Menzies, R.G., & Clarke, J.C. (1995). The etiology of phobias: A nonassociative account. *Clinical Psychology Review, 15*, 23-48.

Mitchell, C. J., & Lovibond, P. F. (2002). Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity. *The Quarterly Journal of Experimental Psychology 55B,* 311-329.

Mitchell, C.J., Lovibond, P.F., Condoleon, M. (2005). Evidence for flexible rule use in blocking of causal judgments. *Learning and Motivation, 36,* 77-87.

Moors, A. (2007). Can cognitive methods be used to study the unique aspect of emotion: An appraisal theorist's answer. *Cognition & Emotion, 21*, 1238-1269.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological bulletin, 132,* 297-326.

Öhman, A., & Mineka, S. (2001). Fears, phobias and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*, 483-522.

Öhman, A., & Soares, J. J. F. (1993). On the automatic nature of phobic fear: Conditioned electrodermal responses to masked fear-relevant stimuli. *Journal of Abnormal Psychology, 102*, 121-132.

Öhman, A., & Soares, J.J.F. (1994). “Unconscious anxiety”: Phobic responses to masked stimuli. *Journal of Abnormal Psychology, 103*, 231–240.

Öhman, A., & Soares, J.J.F. (1998). Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following non-recognized fear-relevant stimuli. *Journal of Experimental Psychology: General*, 127, 69–82.

Parton, D. A. & DeNike, L. D. (1966). Performance hypotheses of children and response to social reinforcement. *Journal of Personality and Social Psychology, 4*, 444-447.

Pavlov, I. (1927). *Conditioned Reflexes*: (1927) New York, NY, US: Oxford University Press.

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

Pearce, J.M., & Bouton, M.E. (2001). Theories of associative learning in animals. *Annual Review of Psychology, 52,* 111-139.

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

Perruchet, P. (1985). A pitfall for the expectancy theory of human eyelid conditioning. *Pavlovian Journal of Biological Sciences, 20*, 163-170.

Perruchet, P., Cleeremans, A., & Destrebecqz (2006). Dissociating the effects of automatic activation and explicit expectancy on reaction times in a simple associative learning task. *Journal of Experimental Psychology: Learning Memory and Cognition, 32,* 955-965.

Pineño, O., Denniston, J. C., Beckers, T., Matute, H. & Miller, R. R. (2005). Contrasting predictive and causal values of predictors and of causes. *Learning & Behavior, 33,* 184-196.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33*, 130-144.

Price, P. C. & Yates, J. F. (1993). Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Memory & Cognition, 21*, 561-572.

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*. New York: Appleton-Century-Crofts.

Schacter, D.L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory & Cognition, 13,* 501-518.

Schmajuk, N., & Larrauri, J. (2008). Associative models can describe both causal learning and conditioning. *Behavioral Processes, 77,* 443-445.

Seligman, M. & Johnston, J. (1973). A cognitive theory of avoidance learning. In F.J. McGuigan & D.B. Lumsden (Eds.), *Contemporary approaches to conditioning and learning* (pp. 69-110). Washington: Winston.

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology, 60,* 291-309.

Shanks, D. R., & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 405-415.

Shanks, D.R., & Dickinson, A. (1990). Contingency awareness in evaluative conditioning: A comment on Baeyens, Eelen, and Van den Bergh. *Cognition & Emotion, 4,* 19-30.

Shanks, D. R. & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral & Brain Sciences, 17*, 367-447.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Stevenson, R. J., Prescott, J. & Boakes, R. A. (1998). Changes in odor sweetness resulting from implicit learning of a simultaneous odor-sweetness association: An example of learned synesthesia. *Learning and Motivation, 29*, 113-132.

Strack, F. & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247.

Thompson, R.F. (2005). In search of memory traces. *Annual Review of Psychology, 56*, 1-23.

Vadillo, M. A. & Matute, H. (2007). Predictions and causal estimations are not supported by the same associative structure. *Quarterly Journal of Experimental Psychology,* 60, 433-447.

Vandorpe, S., De Houwer, J. & Beckers, T. (2007). Outcome maximality and additivity training also influence cue competition in causal learning when learning involves many cues and events. *Quarterly Journal of Experimental Psychology, 60*, 356-368.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 223-266). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wagner, A.R., & Brandon, S.E. (1989). Evolution of a structured connectionist model of Pavlovian conditioning (AESOP). In S.B. Klein & R.R. Mowrer (Eds), *Contemporary Learning Theories: Pavlovian conditioning and the status of traditional learning theory (*pp. 149-189). Hillsdale, NJ: Lawrence Erlbaum Associates.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 53-76.

Waldmann, M. R., & Walker, J.M. (2005). Competence and performance in causal learning. Learning and Behavior, 33, 211-229.

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology: Animal Behavior Processes, 32*, 454-459.

Wardle, S.G., Mitchell, C.J., & Lovibond, P.F. (2007). Flavor evaluative conditioning and contingency awareness. *Learning & Behavior, 35,* 233-241.

Weidermann, G., Tangen, J., Lovibond, P.F., & Mitchell, C.J. (in press). Is Perruchet's dissociation between eyeblink conditioned responding and outcome expectancy evidence for two learning systems? *Journal of Experimental Psychology: Animal Behavior Processes.*

Wilson, T.D., Lindsey, S., & Schooler, T.Y. (2000). A model of dual attitudes. *Psychological Review, 107,* 101-126.

Zinbarg, R.E. (1990). Animal research and behavior therapy Part I: Behavior therapy is not what you think it is. *The Behavior Therapist, 13,* 171-175.

Footnote

[1] The recent reasoning literature often attributes non-normative performance on reasoning tasks to an automatic process, which is part of a dual-process or dual-system view of reasoning (e.g. Sloman 1996; Stanovich, 1999). Quite confusingly, this automatic process is sometimes labeled "associative". However, no link-formation mechanism is imputed here. "Associative" in this context refers to a heuristic whereby responding is determined by the overall similarity of the test stimulus to stored prototypes. Therefore, the automatic component of this particular dual-system model operates at the level of performance, not learning – it is quite different from the link mechanism that is the focus of the present paper.