BEHAVIORAL AND BRAIN SCIENCES (2009) **32**, 183–246
Printed in the United States of America
doi:10.1017/S0140525X09000855

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: association; associative link; automatic; awareness; conditioning; controlled; dual-system; human associative learning; propositional

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 et al. 2000). In each case, one system is generally characterized as conscious, cold, and calculating; the other, as unconscious, affective, and intuitive. In this target article, 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 Fig. 1) and, therefore, produce salivation in much the same way as would actual presentation of the US itself.

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 associative learning is silent as to the psychological mechanism responsible for learning. The second meaning of 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

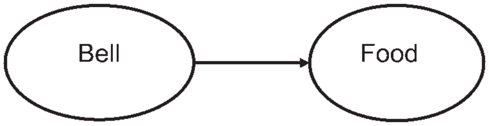

Figure 1. Elipses indicate mental representations (of the bell and the food). The arrow between the two elipses indicates the mental link formed as a consequence of bell-food pairings. The bell ringing produces salivation because it activates the mental representation of food, which, in turn, produces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

contiguous (with some restrictions) pairings of those physical stimuli. These mental links then allow the presentation of one stimulus to activate the representation of – that is, 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). Our target article argues against this position and aims to show that associative learning results, 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. Hence, 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.

Chris Mitchell is senior lecturer at the University of New South Wales. From 1991–1997 he investigated associative learning in rats at University College London. He then worked at Unilever Research, Port Sunlight, as a consumer psychologist, before returning to academia in 2000 to pursue his interests in attention, memory, and associative learning. He has published in the *Journal of Experimental Psychology: Animal Behavior Processes*, the *Journal of Experimental Psychology: Learning, Memory, and Cognition*, and the *Journal of Experimental Psychology: General*. He is an Associate Editor of the *Quarterly Journal of Experimental Psychology*.

Jan De Houwer is Professor of Psychology at Ghent University. He is the author of over 100 publications in the field of experimental psychology, including publications in *Psychological Bulletin* and the *Journal of Experimental Psychology: General*. His main research interest is the manner in which spontaneous (automatic) preferences are learned and can be measured. He is Editor of *Cognition and Emotion* and was Laureate of the Royal Flemish Academy of Belgium for Science and the Arts.

Peter Lovibond is Professor of Psychology at the University of New South Wales. He has qualifications in experimental and clinical psychology, and has carried out research on animal learning and motivation, human learning, cognition, psychophysiology, and anxiety and depression. He is a Consulting Editor for *Learning and Behavior*, the *Journal of Experimental Psychology: Animal Behavior Processes*, and the *Journal of Abnormal Psychology*. He is a Fellow of the Australian Psychological Society and the Association for Psychological Science.

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). 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. (Shanks 2007, p. 297)

Likewise, McLaren et al. (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.

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

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 target article, 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 De Houwer 2009; De Houwer et al. 2005; Lovibond & Shanks 2002). 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 section 2.

2.1. Learning

As outlined in section 1, 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 et al. (2005a, 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 et al. 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 Fig. 1). Salivation (the CR) is observed because activation of the US representation 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

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

earlier CS-US pairings (conditioning), presentation of the CS will associatively activate the US representation into the (secondary) A2 state. Thus, Wagner's model postulates two different states of activation to distinguish between perception of the US when it is physically present (the A1 state) and anticipation of that US (the A2 state).

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 behavioral consequences; see Bolles & Fanselow 1980). That is, both US memory and US anticipation are represented by A2 activation of the US representation. 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 et al. 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 a 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 relationships 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 to 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 (Pineño et al. 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 in section 1, 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 further on.

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, to play chess, or to 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 when we hear that bell, we are about to receive food again. We may also recall having previously hypothesized 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 et al. 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, they, like all propositions, have truth value (e.g., it is either true or not true that bacterial infection causes peptic ulcers). Therefore, 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. Therefore, 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. First, 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; Chater et al. 2006; Evans & Over 1996; Johnson-Laird 1983), 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, in whichever way this is achieved.

Second, 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, of past bell-food pairings.

Third, 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 2009).

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 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). Hence, 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 about the pattern of responding that might be observed when different contingencies are in place. First, 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. Second, 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. Third, 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 favors 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 with 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.

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

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 & 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 et al. 1990a) 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 rather from 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 participants experience 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 et al. 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. Therefore, 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. Hence, 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 (e.g., 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 verbalizable 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 (2005a) 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

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

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 (see sect. 3.2), 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 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 the knowledge acquired verbally and that acquired by direct experience are 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 these instructions 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 $A+$, $B+$, $AB-$, $C-$, $D-$, and $CD+$ 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 et al. (2002) observed blocking only when it was possible to

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

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. (2002) 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. (2002) 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 et al. (2005; see also Lovibond et al. 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 et al. (2007a) 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 in section 4.2 concerning cognitive load), and so the propositional system is unable to operate or is off-line (e.g., Dickinson 2001; Le Pelley et al. 2005a). Vandorpe et al.'s (2007a) 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. First, 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. Second, 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, it has been suggested that some neuroscientific data indicate the existence of a multiple learning system. We address these lines of evidence in turn.

5.1. Unaware associative learning

In evaluative conditioning research (see De Houwer et al. 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. 1990a; Dickinson & Brown 2007; Fulcher & Hammerl 2001; Walther & Nagengast 2006; and see Stevenson et al. 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 analyzed as a single group, did not demonstrate reliable contingency awareness but did show evaluative conditioning. However, Wardle et al (2007) reanalyzed 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 with which a subset of cues were paired. It is possible that it is this subset of cues that are responsible for the evaluative conditioning observed in earlier studies (Pleyers et al. 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 (1990a)

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

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 analyzed at the item level. Even more convincing than Baeyens et al's (1990a) 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 pseudorandom 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. Hence, after a run of three 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 three 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 et al. 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 Weidemann et al., 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 2007; Shanks & Dickinson 1990). 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, arachnophobes 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 arachnophobes can be explained which are consistent with the propositional approach to learning: (1) 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. (2) 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 sect. 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. (3) There is, in fact, little evidence that specific phobias of this kind result from learning at all, and therefore 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 they appeared to

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

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. (2005a) 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. (2005a) 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 a result of 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 ($P(O/{\sim}C) = 0$). 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 et al. 2007; Lagnado & Shanks 2002; López et al. 1998a; 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 (2007) 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 (2005a) participants knew that something about the outcomes had changed between *A-O1O2* trials and *AT-O1O3* trials, but they 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.

First, one can test propositional explanations of irrational behavior empirically. For instance, if Le Pelley et al.'s (2005a) 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 ten 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. Second, 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 instance, 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). Therefore, 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, however, is 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 behavior (Jog et al. 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).

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 behavioral evidence concerning human associative learning does not support the view that there are multiple learning systems. The behavioral 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. 1990a). 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; and (3) researchers are resistant to the idea that nonhuman animals engage in propositional reasoning. We will evaluate the relative strengths and weaknesses of the propositional and link-based 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 et al. 2001). One can also explain overshadowing (Pavlov 1927) and the opposite phenomenon, potentiation (Garcia et al. 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 (2008) 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

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

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 Larrauri (2008) focus on generalization between compounds (e.g., *GH* and *AT*). However, generalization between elements is ignored, as is generalization from compounds (e.g., *GH*) to elements of those compounds (e.g., *G*). Hence, Schmajuk and Larrauri (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 generalization between different kinds of cue.

Secondly, as Schmajuk and Larrauri (2008) admit, the explanatory power of this 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 explain. For example, in a similar experiment to that of Beckers et al. (2005), Mitchell et al. (2005) showed that *G*+, *H*+, and *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*−). The variant of the Rescorla-Wagner model proposed by Schmajuk and Larrauri (2008) 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 (2005) 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 post hoc 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 post hoc 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 it is an alternative 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 (1972) 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. Hence, 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 focusing 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

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

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.

First, 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. Second, 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 Fig. 1). This target article 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 nonhuman 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 nonhuman animal learning is also a process of belief acquisition. Therefore, the complex representational system we possess evolved from similar, but simpler, cognitive systems in our ancestors; and many differences observed between human and nonhuman 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 nonhuman 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 to reason about cause and effect (Beckers et al. 2006; Blaisdell et al. 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 rats as subjects. Beckers et al.'s (2006) 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

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

be responsible for associative learning in nonhuman 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 nonhuman 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 nonhuman animals (including *Aplysia*) also display plasticity at the neural level. Within all species, reflexive (and therefore, 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 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 Gardner 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 an S-S model postulates that associations between stimuli in the environment are represented by links between internal representations of those stimuli.

In 1973, Seligman and Johnston 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 Johnston maintained the view that Pavlovian conditioning results from an automatic mechanism, in which links form 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 target article 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

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

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 patients' distorted beliefs and thereby normalizing 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 are 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 nonhuman 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.

NOTE

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 target article.

Open Peer Commentary

Associative learning requires associations, not propositions

doi:10.1017/S0140525X09000867

Frank Baeyens, Debora Vansteenwegen, and Dirk Hermans
Department of Psychology, University of Leuven, 3000 Leuven, Belgium.
frank.baeyens@psy.kuleuven.be
deb.vansteenwegen@psy.kuleuven.be
dirk.hermans@psy.kuleuven.be
http://ppw.kuleuven.be/ppw/index.htm
http://ppw.kuleuven.be/ppw/english/
http://ppw.kuleuven.be/leerpsy/

Abstract: We discuss findings on evaluative conditioning (EC) that are problematic for the "conscious reasoning/propositional knowledge" account of learning, namely, dissociations between conscious beliefs and acquired (dis)liking. We next argue that, both for EC and for Pavlovian learning in general, conditioned responding cannot rationally be inferred from propositional knowledge type "CS refers to/signals US," and that, therefore, performance cannot be explained.

There is much in this target article that we can fully endorse, but unfortunately, not the general conclusion that Pavlovian conditioning depends on high-level cognitive processes type "conscious propositional reasoning," the core outcome of which is propositional knowledge about stimulus relations. We believe there is something fundamentally wrong with this proposal *in general*, but we start with the more modest claim that, both empirically and conceptually, Mitchell et al.'s approach in the target article does not capture the essence of *evaluative conditioning* (EC) (Baeyens et al. 2001b).

There are many demonstrations of EC effects going counter to participants' conscious propositional beliefs, including situations (1) in which unconditioned stimulus (US) occurrence is attributed to an irrelevant, non-correlated characteristic of the conditioned stimuli (CSs), but acquired (dis)liking follows the objective CS-US contingency (Baeyens et al. 1990b; 1996b; 2001a); (2) in which US occurrence is correctly believed to be conditional (modulated), but acquired (dis)liking is unconditional (Baeyens et al. 1996a; 1998); or (3) in which US occurrence is no longer expected, but conditioned (dis)liking still persists (Vansteenwegen et al. 2006). These dissociations suggest that EC falls beyond the scope of Mitchell et al.'s model.

Moreover, Mitchell et al.'s account of EC faces more serious, conceptual problems: (i) How does one infer CS (dis)liking from the propositional-declarative knowledge type "CS refers to the (dis)liked US"? From a logical/rational/inferential point of view, there is no causal connection *at all* between entertaining the propositional belief, "CS refers to (dis)liked US," and (dis)liking a CS. "Disliking" a flavor-CS does not follow any more from entertaining the belief that "This flavor refers to the bad tasting US Tween20," than that it would follow from it that the CS should be "liked," or should evoke feelings of pride, envy, or plain misery. Related to this: (ii) How does one explain that a (correct) propositional belief about the CS-US relationship is not a *sufficient* condition for EC effects (Baeyens et al. 1990b)? Finally, (iii) if EC would be based upon a proposition with truth-value, how does one deal with the observation that EC is resistant to extinction; that is, that acquired (dis)liking is not affected by the subsequent experience (and resulting propositional belief) that the CS is no longer accompanied by the US? The alternative is to accept that EC does not depend on propositional attitudes (thinking/believing *that* "CS refers to US"), but reflects a less-than-propositional state (thinking *of* US) resulting from the operation of an associative mechanism: CS presentation activates a representation of a (dis)liked US, and the activation of this representation causes or instantiates an approach/avoidance response-tendency that phenomenologically equals a feeling of "(dis)liking," such that "(dis)liking" the CS causally follows from mere presentation of the CS. Such an association is not true or false, but is simply established or not established; and once it has been formed, it remains there forever.

One could argue that EC is just a special case, and that Mitchell et al.'s approach is valid for all instances of Pavlovian learning *except for EC*. We are not sympathetic to this possibility either. The problem of linking conditioned responding ("performance") to specific "belief" states (resulting from entertaining a particular propositional mental content) is not restricted to EC, but spreads to the whole domain of Pavlovian conditioning. Mitchell et al. indeed admit that a detailed causal/mechanistic account of the translation of propositional beliefs (and concomitant/resultant expectancies) to specific physiological/behavioral responses is not the forte of their model. It is a long way from the propositional attitude, "I believe that the tone predicts a shock," to increases in heart rate, muscle tension, breath regulation, hormone release, or the activation of escape/avoidance responses. It is not just a long way, but also a way that is not specified *at all* by the (content of the) propositional attitude. In principle, Pavlovian conditioned responses are not rationally connected to propositional knowledge about the CS-US relationship and cannot (logically) be inferred from it; nor do they (necessarily) result from the interaction with other propositional beliefs (i.e., from reasoning) (Shanks & Dickinson 1990). A model that can explain all but conditioned behavior, is lacking something quintessential. Again, the alternative is to accept that conditioned responding causally results from the associative activation of the US-representation – a theory that may indeed require (much) refinement, but at least offers an account that works *in principle*.

Why is it that Mitchell et al. have arrived at this problematic account of learning? First, in some arguments purportedly favoring the "conscious reasoning/propositional knowledge" idea, two issues are erroneously mixed up. The data on the role of awareness, and the influence of cognitive load/secondary tasks, favor an account of Pavlovian learning that acknowledges the importance of *"controlled" processing* indeed; but these arguments do not bear at all upon the issue whether the acquired knowledge should take the form of *structured mental representations*, or of non-propositional associative links between representations. Second, many of the observations that *do* favor an account in terms of representations with combinatorial syntax and semantics (and structure sensitivity of processing), are derived from experiments that invoke processes/faculties that simply go beyond the scope of what associative learning theory reasonably could be expected to explain. The findings on the influence of verbal instructions, abstract rules, and deductive reasoning show that people indeed can use language and reason, and can transform complex, structured mental representations in situations that require more than simple registration of stimulus co-occurrence. But how does this favor a "conscious reasoning/propositional knowledge" account of *associative learning* any more than does a demonstration that humans can play chess or understand poems?

Mitchell et al.'s account of learning tries to get rid of a dual-system approach, and while doing so, throws away the very notion of associations. According to our analysis, there is not much ground to justify this radical stance. Moreover, Mitchell et al. still adhere to a dual-system approach of mental processes *in general*. Even though the very act of learning is supposed to take place in a conscious reasoning system, this system is said to get inputs from a non-conscious, automatic perceptual/memory system; and the only thing Mitchell et al. ultimately propose is to resect plasticity from one part of the dual system. Parsimony, where art thou?

Propositional learning is a useful research heuristic but it is not a theoretical algorithm

doi:10.1017/S0140525X09000879

A. G. Baker,[a] Irina Baetu,[a] and Robin A. Murphy[b]

[a]*Department of Psychology, McGill University, Montréal, Québec, H3A 1B1, Canada;* [b]*Cognitive Perceptual and Brain Sciences Unit, Division of Psychology and Language Sciences, University College London, London, WC1E 6BT, United Kingdom.*

andy.baker@mcgill.ca
http://www.psych.mcgill.ca/faculty/abaker.html
irina.baetu@mail.mcgill.ca
robin.murphy@ucl.ac.uk

Abstract: Mitchell et al.'s claim, that their propositional theory is a single-process theory, is illusory because they relegate some learning to a secondary memory process. This renders the single-process theory untestable. The propositional account is not a process theory of learning, but rather, a heuristic that has led to interesting research.

In a conditioning experiment, a light might signal food. There are at least two interesting questions about this conditional relationship. The first is a psychophysical question: Is the animal sensitive both to the events and to the conditional rule that describes the relationship between the events? The second question is: From what internal mechanism, or algorithm, does this conditional rule emerge? Mitchell et al. argue that learning emerges from an internal representation of the conditional rule as a proposition. This theory is contrasted with the view that the propositional representation emerges from internal associative links between the events that do not involve direct propositional mechanisms.

Cognitive penetrability and the absence of automaticity? Because it rejects the automatic link processes in learning, a propositional mechanism is more parsimonious than "dual theories" that postulate two mechanisms This parsimony is illusory because the target article authors themselves describe a second process involving memory retrieval. "Learning" is propositional, but memory is sometimes automatic. An association formed as a proposition may, in memory, become automatic. This distinction is important for Mitchell et al.'s argument since the propositional view relies on the cognitive penetrability (Brewer 1974) of conditioning experiments. That is to say, cognitive manipulations, such as instructions, can directly modulate conditioning without having direct experience with the conditioning events. Automatic memories would not be easily susceptible to cognitive manipulations. So, if a learned rule is penetrable, it is propositional; but, if the rule is not penetrable, it has become automatic and thus reflects memory, not learning. Cognitive penetrability as used here is therefore circular. If people are aware of the contingency, then instructions can influence behaviour; so the mechanism is propositional. If not, the measurement tool is not sensitive enough, or the behaviour has become automatic in memory.

Cognitive load manipulations designed to show that increasing load disrupts propositional reasoning provide empirical evidence for the presence of automatic links. Imposing a load, while participants are exposed to a redundant-cue blocking paradigm ($A+$ $AX+$), eliminates blocking because the load prevents the propositional process (De Houwer & Beckers 2003; Waldmann & Walker 2005). However, even though the propositional process is challenged, participants still respond to both *A* and *X*. If the propositional process is blocked, then this must mean the learning here is "automatic." Mitchell et al. imply that simple pairings are automatic memory processes and that the important issues in learning are phenomena like blocking and inhibition. Even associationists have argued that such phenomena are more susceptible to modulation than excitatory pairings (Bouton 2004; Swartzentruber & Rescorla 1994). Mitchell et al. seem to be restricting their analyses to those learning phenomena that are least stable and most susceptible to modulation.

The autonomic conditioning preparations might not be ideal for demonstrating strong automatic learning. Most animal conditioning preparations generate learning in nearly 100% of animals. An important finding of many of the cognitive penetrability experiments is that some people learn, and some do not, but only those who "know" the contingency show conditioned responses. Hence, in many experiments, a significant proportion of participants do not learn. This hardly makes us confident that these are biologically prepared (Seligman 1970) learning preparations suitable for evaluating the formation of strong "automatic" associations. Penetrability may reflect weak conditioning, or perhaps no conditioning at all, and therefore begs the question as to what cognitive penetrability means for human conditioning. Furthermore, unlike the rats in a conditioning experiment involving shocks, participants in these autonomic learning tasks are not naïve. They have learned to expect motivational events following verbal instructions. They have learned to relax when told the teacher will not check their homework today. Is it surprising that they will feel anxious when told that a tone will be followed by a shock, or will relax when subsequently told it will no longer be followed by the shock?

We believe there are learned pairings that are not cognitively penetrable. These include evaluative conditioning (Baeyens et al. 1990a) and probably strong aversions and flavour preferences. Cultural dishes evoke strong positive reactions and these are learned. Few of us have a strong positive reaction to eating live grubs, but in some cultures some people do. In a classroom demonstration, we have difficulty getting our students to salivate at the thought of crunching a grub between their teeth, as we might with students within the appropriate culture. These reactions are learned but not penetrable. This conclusion would be again open to the circular criticism that these processes were overlearned and became automatic in memory.

Propositional learning as a valuable heuristic. We prefer lower-level psychological theories to explain rule learning and propositional reasoning, rather than have them as fundamental primitives of learning. If propositional learning is impossible to disconfirm and has not eliminated automatic learning, then is it valuable? We would argue that it is. High-level rule-based and symbolic cognitive views can generate important research. These cognitive views can either be a theory about internal mechanisms, or they can represent a sometimes normative description of the physical and statistical mechanisms in the world. For example, Tolman (1948) challenged the S-R (stimulus-response) psychologists' automatic theories with his notion of the cognitive map, and this notion generated interesting research and challenged the automatic or associative approach. Now, however, animals' abilities to navigate in the world to goals are beginning to be explained by lower-level theories (Diez-Chamizo et al. 1985).

Elsewhere, arguing against the notion that internal mental images of three-dimensional drawings may be rotated in much the same way they are in the physical world (a notion which generated a great deal of interesting research), Pylyshyn (1973) pointed out that this is not sustainable as a cognitive theory. Mental images are computations but not direct mental representations. Our own research on contingency learning in humans and rats was originally motivated by theories (e.g., Cheng 1997) that humans and animals internalized the notion of computing contingencies (Baker & Mackintosh 1979; Baker et al. 2003). We have subsequently argued that for both animals and humans these computations of correlations, and even computations of the dependencies in multiple-event chains, emerge from a connectionist network (Baetu & Baker, in press; Murphy & Baker 2004; Wasserman et al. 1993).

Around the time of the "cognitive revolution," a number of rule-based and symbolic computational models emerged. Interest in these models has diminished because connectionist models have accounted for much of the data they initially explained, as well as for some they did not (Shultz 2003). More recent symbolic models even incorporated sub-symbolic (automatic) modules (e.g., Anderson & Lebiere 1998). Moreover, it has long been known that connectionist models can generate truth tables, logical operations, and many other linear and nonlinear rules (McCulloch & Pitts 1943). "Higher-level" cognitive processes and developmental stages emerge from connectionist or automatic architectures (e.g., Shultz et al. 1994).

This brings us to propositional learning mechanisms. These propositions map the normative relationship between events in the world. Although we take issue with some of the data, we see Mitchell et al.'s work as a psychophysics of "propositional" relationships. They show the rules and mechanisms in the world people and animals can represent and pose a challenge for connectionist or other lower-level theories. We are confident that the principles of propositional learning will emerge from connectionist principles that, contrary to the authors' claims, provide an algorithm and not a computational description (Marr 1982).

The truth and value of theories of associative learning

doi:10.1017/S0140525X09000880

Tom Beckers and Bram Vervliet

Department of Psychology, University of Amsterdam, 1018 WB Amsterdam, The Netherlands.

tom.beckers@uva.nl

http://home.medewerker.uva.nl/t.r.j.beckers/

b.vervliet@uva.nl

Abstract: In this commentary, we assess the propositional approach to associative learning not only in terms of veridicality and falsifiability, but also in heuristic value. We remark that it has furthered our knowledge and understanding of human, as well as animal, associative learning. At the same time, we maintain that models developed from the association formation tradition continue to bear great heuristic value as well.

In their target article, Mitchell et al. present a detailed and very thoughtful evaluation of the potential evidence in favour of automatic association formation as a source of associative learning effects. They convincingly argue that there is presently very limited, if any, evidence for the existence of a separate, non-propositional association formation module. They rightfully point out that, given the obvious need to postulate the existence of a propositional module in order to explain many aspects of human associative learning, an association formation module that does not add explanatory power to the propositional module is entirely redundant.

Still, the claim that any instance of human associative learning (i.e., any change in performance that is a result of the presence of regularity in the relation between events in the world) must by necessity be due to the operation of controlled, propositional reasoning processes, is a strong one. Prima facie, Mitchell et al.'s claim seems ill-fitted to the existence of a phenomenon such as evaluative conditioning: If, for example, I have developed a liking for white wine because of spending many pleasant holidays in France, it would seem that this liking for white wine does not need to reflect any knowledge about the relation between white wine and anything else, beyond the fact that white wine makes me think of France. Essentially, the fact that white wine makes me think of France is non-propositional (I cannot be right or wrong for being reminded of France upon smelling white wine). As such, evaluative conditioning effects very strongly appear to result from the mind being carried from one idea or representation to another, without any intermediate processing, much like what is the presumed mode of operation of an association (Fodor 2003).

It is then perhaps not surprising that some of the best evidence for automatic association formation comes from an evaluative conditioning study (Baeyens et al. 1990a). Still, the description of the propositional approach as offered by Mitchell et al. leaves open the possibility that even evaluative conditioning effects, although perhaps resulting from automatic, non-propositional memory retrieval processes (an object "automatically" making you think of something pleasant), do necessitate the conscious, falsifiable establishment, in propositional form, of a link between events (if only of the form "event *A* co-occurred with event *B*") at some earlier point in time. That is, the fact that evaluative conditioning effects, at performance, are almost by nature non-propositional (the fact that *A* makes you think of *B* is not something that you can subsequently evaluate as correct or wrong), does not preclude that they perhaps only occur if people at some point have consciously noticed some sort of real-world relationship between *A* and *B* (such as "*A* has repeatedly co-occurred with *B*," a statement which you can obviously evaluate to be true or false).

Does this render the propositional account unfalsifiable? Surely, the fact that performance may reflect automatic memory retrieval of propositional knowledge stored earlier and may moreover reflect propositional knowledge indirectly (such as when stored propositional knowledge about the co-occurrence of two events influences your subsequent evaluation of one of both), does make falsification of the general framework difficult, but not impossible. It would suffice to convincingly demonstrate associative learning about entirely subliminally presented CSs to rule out a role for propositional reasoning altogether. The debate about whether such evidence already exists still seems to be open (see Wiens & Öhman 2002 vs. Shanks & Lovibond 2002).

However, the most important contribution of the propositional approach to associative learning is not to be situated in proving the association formation approach wrong. As Mitchell et al. point out, what is perhaps most important, is that it has provided a new perspective on conditioning, not only in humans (where at least a contribution of reasoning processes to learning has long been acknowledged), but also in animals. This perspective has not only enabled us to unveil the importance of rule learning in animal Pavlovian fear conditioning (Beckers et al. 2006), but also to highlight the parallels between extinction learning and rule learning in terms of context sensitivity and generalisation (Wheeler et al. 2008). As such, the propositional approach has opened up a whole new framework for the understanding and the prediction of human and animal conditioning phenomena, the impact of which is bound to further increase over the coming years.

And perhaps this is where a caveat about the propositional approach to associative learning, in turn, is warranted. Notwithstanding the impressive amount of evidence that the propositional approach is more veridical than the association formation approach, it seems beyond argument that models developed within the association formation tradition have continuing heuristic value as well. As an example, just recently Leung and Westbrook (2008), in a series of extremely elegant experiments, demonstrated that the degree of additional extinction accrued by a cue exhibiting spontaneous recovery is governed by both individual prediction error of the cue and common prediction error of all cues present during an extinction trial. Does such a finding invalidate the propositional nature of associative learning? Not necessarily (probably not, one might even argue). Still, it is obvious that experiments like these would never have been designed, and these findings never revealed, on the basis of our current understanding of propositional reasoning. As such, it may simply be too early for one truth to govern our inquiries into human and animal associative learning. Keeping our antennas open to discover empirical phenomena in the realm of associative learning and conditioning will probably necessitate a willingness to entertain a variety of models and approaches for some time to come.

What's reason got to do with it? Affect as the foundation of learning

doi:10.1017/S0140525X09000892

Eliza Bliss-Moreau[a] and Lisa Feldman Barrett[b,c]

[a]*Department of Psychiatry and Behavioral Sciences, California National Primate Research Center, University of California–Davis, Davis, CA 95616;* [b]*Department of Psychology, Boston College, Chestnut Hill, MA 02467; and* [c]*Massachusetts General Hospital/Harvard Medical School, Charlestown, MA 02129.*

eblissmoreau@ucdavis.edu barretli@bc.edu
www2.bc/edu/~barretli

Abstract: We propose that learning has a top-down component, but not in the propositional terms described by Mitchell et al. Specifically, we propose that a host of learning processes, including associative learning, serve to imbue the representation of the conditioned stimulus (CS) with affective meaning.

In the target article, Mitchell et al. characterize associative learning phenomena according to the *relationship* established between two previously unrelated stimuli (i.e., a conditioned stimulus [CS] and an unconditioned stimulus [US]). Associative learning, they suggest, occurs when the CS becomes *propositionally related* to the US using effortful, controlled, and rational processing. We believe this view does not account for important questions about how the representations in question (the CS and US) are modified by experience. Furthermore, this view makes assumptions about how stored representations are activated. We suggest that stimulus representations are realized by multimodal states reflecting both exteroceptive and interoceptive information brought online by a combination of top-down (e.g., propositional) and bottom-up (e.g., stimulus-driven) processes. In this view, learning occurs when the multimodal representation of a stimulus acquires an affective component. Propositional change is not necessary for learning.

A growing body of evidence suggests that the human brain captures statistical regularities in sensory-motor patterns and stores them as representations. These representations are used to continuously organize incoming sensations during the process of predicting what those sensations stand for the in world (Bar 2003; 2007; Kveraga et al. 2007). External sensations always occur in a context of internal sensations from the body. As a result, the sensory-motor pattern that is stored for future use will always include a representation of the interoceptive state of the body. The brain predicts what sensations refer to in the world in part based on prior experiences of how those external sensations have influenced, or changed, internal sensations from the body on prior encounters (cf. Barrett & Bar, in press). These bodily changes are referred to as "affective." Affective states can be described hedonic (pleasure or displeasure) with some degree of arousal (for a recent review, see Barrett & Bliss-Moreau, in press). These ideas are consistent with a growing body of research demonstrating that knowledge about the world is "embodied," or grounded, by a network of broadly distributed, diverse, multimodal states which are encoded during the experience of a given stimulus (see Barsalou 2008). What you know about an object is therefore based, in part, on its affective impact in the past.

When it comes to learning, changes in a CS's meaning can be thought of as the process by which the multimodal representation of the stimulus is changed by any experience. The most fundamental change occurs because the representation of the CS is experienced in a context of affective arousal that is derived from the representation of the US. Any number of relationships between the CS and US could serve to alter the representation of the CS. The CS and US could be paired in time or space, associated semantically, or even explicitly coupled via rule-based learning. In our view, the need to differentiate types of learning in terms of how the relationship between CS and US is established (as exemplified by Mitchell et al.'s model) is eliminated. All learning can be subsumed under the same general, basic mechanism that exists in all organisms that possess the capacity to generate affective responses to stimuli in the environment. Thus, to some extent, any change in the representation of a CS is *affective learning* (Bliss-Moreau et al. 2008).

In typical classical conditioning paradigms, examples of USs include shocks (e.g., Vervliet et al. 2005), very loud noises (e.g., Neumann & Waters 2006), and even sexual arousal (e.g., Hoffmann et al. 2004). These USs act on the nervous system directly to generate a robust affective response in a bottom-up or stimulus-driven way that is automatic and unconscious. Other USs, such as negative words or pictures, have a less robust bottom-up effect on the learner's nervous system. Instead, such USs have top-down effects because they have propositional meaning. The difference in the bottom-up potency of different USs leads some theorists to believe that different models are required to account for learning phenomena. According to the affective learning perspective, this is not so – changes in affect can and do occur via both bottom-up and top-down processing and therefore with both types of USs. For example, evidence from instructed learning paradigms demonstrates that the representation of a CS can be changed by telling a person that a US will be presented after the CS, even if the US is never presented (e.g., Olsson & Phelps 2004). According to the affective learning perspective, the set of instructions that indicates when the (promised, but never presented) shock will occur sufficiently alters the learner's affective state so that the interocpetive representation of this affective change is integrated into the representation of the CS. We have demonstrated that people can learn the affective value of other people when presented with propositional information about those people (e.g., seeing the phrase "hit a small child" presented with a picture of Sally) (Bliss-Moreau et al. 2008). In this example, the representation of "hit a small child" has an affective component which is integrated into the representation of Sally.

It is possible that some USs are exclusively experienced either via automatic, effortless associative processing *or* via effortful, controlled propositional processing (but not both), as Mitchell et al. and most dual-process theories suggest (for an extensive review, see Evans 2008). A more likely scenario, however, is that the two types of processing are often active in parallel and serve to constrain each other to make meaning of a given stimulus in a given context. For example, the sound of gunfire is aversive and may have an automatic effect on the nervous system. But, for a person who has never experienced war, that automatic processing may be constrained by propositional information about the "shoot-'em-up" Western movies he or she remembers from childhood, resulting in a relatively neutral experience. For a war vet, the automatic processing may be constrained with propositional information gained in the experience of fighting and killing, resulting in a highly aversive experience. Propositional learning, even for a stimulus that has semantic meaning, is not required.

By focusing on how stimulus representations are changed as a result of internal experience, a whole host of learning phenomena can be united under one principle. Our hope is that by approaching learning from this perspective, the field will generate new hypotheses about the way that people learn about the world.

Learning without thinking

doi:10.1017/S0140525X09000909

R. A. Boakes
School of Psychology (A18), University of Sydney, NSW 2006, Australia.
bobb@psych.usyd.edu.au

Abstract: The main conclusion to draw from Mitchell et al's article is that it is difficult to disentangle cognitive and learning processes in contingency and causal experiments. More compelling evidence for human associate learning comes from research where, because of the type of events involved, participants are unable or unlikely to think about the relationships between the events.

The conclusion Brewer (1974) drew from his review of conditioning research using human participants came as a great shock. For decades it had been very widely accepted – and not just by the many behaviorists of those times – that the processes described by S-R (stimulus-response) or reinforcement theorists based on animal evidence also provided a basis for at least simple aspects of human behavior in a manner that was independent of belief or awareness. Brewer's conclusion was that there was no convincing evidence to support this assumption. His alternative account of what goes in a conditioning experiment is captured by the following quote: "The college sophomore does not leave his higher mental processes outside the door when he walks into the experimental room, but he uses them to try to understand what is going on and what he should do about it" (p. 2). Despite many subsequent attempts to show his conclusion to be wrong, Brewer (1974) clearly was correct about the overwhelming influence of "higher mental processes" in determining a participant's behavior in the kind of conditioning experiment – mainly "conditioning" of autonomic responses or of small movements – that he reviewed.

The article by Mitchell et al. can be seen as a successor to Brewer (1974), in which a similar argument is directed mainly at experiments from the past two decades that have used causal or predictive scenarios in experiments to test principles of associative learning. One similarity between past and present research is the overwhelming use of college students as participants, a population that has been selected on the basis of thoughtfulness and then encouraged to be curious. An odd aspect of too many causal judgment experiments is that, although the researchers want their participants' higher mental processes to operate in order to understand the sometimes complex instructions, interpretation of the results assumes the absence of any such influence following a participant's first response. In this respect, many points made by Mitchell et al are salutary, including the important one that associations are not expectancies. When applied to animal data, the absence from associative learning theories of explanations as to how links provide the basis for expectancies is rarely problematic; but when applied to human experiments, this is a major lacuna.

Mitchell et al.'s arguments are less compelling when one looks beyond the kind of study on which they focus. Odors provide an example of stimuli that people find hard to describe or identify, and therefore difficult to think about. Stevenson and I have suggested that this is the principal reason that conditioned changes in an odor's perceptual properties are independent of the poor explicit memory participants have for the stimulus contingencies they were given during training (Stevenson & Boakes 2004). For a similar reason, changes in human flavor preferences produced, for example, by caffeine-based conditioning, also appear independent of belief or awareness (e.g., Chambers et al. 2007). Moving beyond the laboratory, the strong aversions developed by cancer patients undergoing chemotherapy appear to develop quite independently of patients' valid beliefs about the cause of their distress; and, in a way, that shows very strong parallels with conditioned taste aversions in rats and many other animals (e.g., Bernstein 1985).

Particularly interesting, and far less widely known, examples come from the study of placebo effects, an area where a similar debate has continued as to the contributions of conscious expectancies and conditioning without awareness (Stewart-Williams & Podd 2004). Here it has turned out that, whereas some placebo effects are strongly determined by patients' beliefs about their treatment and are sensitive to the information that is provided, others depend on past treatment history in a way that is independent of belief and insensitive to information. Thus, in one major study, verbal suggestions accompanying medication had a large effect on reactions to pain and on the motor responses of Parkinson's patients, but no detectable effect on hormonal and cortisol levels, whereas the latter could, however, be altered by a placebo treatment following a conditioning procedure (Benedetti et al. 2003).

What such examples suggest is that the conclusion to draw from Mitchell et al. is that the kind of research they review is not likely to reveal much about human learning. Preventing student participants from "reflection" so that their responses in, say, a causal judgment experiment are "intuitive" (Shanks 2007) turns out to be very difficult. Mitchell et al. refer to three examples that appear to have achieved this, including our own study in which we placed participants in a causal judgment experiment under strong time pressure, so giving them "little time to think" (Karazinov & Boakes 2007). Mitchell et al. attempt to explain away the result; but in doing so, they appear to accept that the non-rational response given by the average participant under these conditions must be based on a within-compound association and not on any kind of logical inference. Nevertheless, even though a few studies of this kind appear to have been successful in reducing the influence of logical inferential thinking, this is not enough to justify a confident return to this way of studying human associative learning.

Rats and infants as propositional reasoners: A plausible possibility?

doi:10.1017/S0140525X09000910

Leyre Castro and Edward A. Wasserman

Department of Psychology, The University of Iowa, Iowa City, IA 52242.
leyre-castroruiz@uiowa.edu **ed-wasserman@uiowa.edu**
http://www.psychology.uiowa.edu/

Abstract: Mitchell et al. contemplate the possibility of rats being capable of propositional reasoning. We suggest that this is an unlikely and unsubstantiated possibility. Nonhuman animals and human infants do learn about the contingencies in the world; however, such learning seems not to be based on propositional reasoning, but on more elementary associative processes.

Whether advanced cognitive competencies can be grounded in more elementary perceptual or associative processes is a matter of considerable current interest. Goldstone and Barsalou (1998) suggested that the distinction between low-level perception and high-level cognition is inapt; there is no *chasm*, but rather a *continuum*, from perception to conception. Similarly, Leech et al. (2008) have proposed that relational priming is a basic building block for analogical reasoning. And, in the realm of language development, sequential learning and domain-general mechanisms may pave the way for language (Christiansen et al. 2002). These examples illustrate how complex mental abilities could emerge from simpler behavioral mechanisms, thereby opening the door to understanding the ontogeny and phylogeny of higher-order cognition.

Mitchell et al. proceed in the opposite direction. They hypothesize that "associative learning depends on high-level cognitive processes that give rise to propositional knowledge" (target article, Abstract). For Mitchell et al., a basic and general ability – learning the relations between environmental events – is not due to the formation of links between representations of the events (associative account), but to the formation of mental propositions about how the events are related (propositional account).

On this view, propositions predicate some property of a subject, are generally held to be true or false, and are combined by laws of logical inference (Braine 1978). Propositional reasoning thus entails processing and storing verbal premises and assertions. For example, "blocking" results from applying the following rule (De Houwer & Beckers 2003): If cue *A* alone causes the outcome with a particular intensity and probability, and if cues *A* and *B* together cause the outcome with the same intensity and probability, then cue *B* does not cause the outcome.

How organisms such as infants and nonhuman animals, who do not have language, deploy such propositions to infer relations about events is unclear; but Mitchell et al. believe that possibility and cite two studies (Beckers et al. 2006; Blaisdell et al. 2006) which hint at rats' engaging in inferential propositional-based reasoning.

Beckers et al. (2006) found that, after presenting rats with $A+$ trials in Phase 1 and $AX+$ trials in Phase 2, blocking (e.g., low responding to X) did not take place if the rats had earlier experienced $C+$, $D+$, and $CD+$ trials. According to Beckers et al., rats possess (as do humans) prior knowledge that, when two potential causes are presented together, a larger effect should occur than when only one cause is presented – additivity. Mitchell et al. agree, but they never say if this prior knowledge is innate or acquired; and if it is acquired, then how all organisms have come to the same understanding.

Nevertheless, armed with the additivity assumption, rats infer that X is not the cause of the outcome, when $AX+$ trials are presented after $A+$ trials and the outcome remains the same. But, if rats are pretrained with $C+$, $D+$, and $CD+$ trials, then they can tell that the additivity assumption is now false. The rats consequently reassess their beliefs; when they are later presented with $A+$ trials followed by $AX+$ trials, they deduce that X, as well as A, is a cause of the outcome.

But, perhaps something simpler is happening here. Haselgrove (under review) has shown that Rescorla and Wagner's (1972) associative model can readily explain the results of Beckers et al. (2006). Haselgrove noted that five out of the six experimental cues in Beckers et al. were from the same modality (audition); all of them were of the same duration; and all of them were trained in the same context. Under those conditions, it is conceivable that the cues used for pretraining and the cues used for blocking entailed a common element. The Rescorla and Wagner model predicts that the conditioned properties acquired by the pretraining cues can transfer to the blocking cues via this common element. Thus, generalization of the pretraining contingencies to the experimental contingencies can explain Beckers et al.'s findings – reasonably and elegantly.

Mitchell et al. also consider the Blaisdell et al. (2006) project as support for propositional reasoning in animals. Here, rats could distinguish between a situation in which the outcome had been observed and a situation in which the outcome had been produced by the rats. This distinction actually reflects the dichotomy between classical and instrumental conditioning. In classical conditioning organisms learn that events in the environment are related to one another, whereas in instrumental conditioning organisms learn that their own actions change environmental events. The fact that animals can distinguish between observation (classical conditioning) and intervention (instrumental conditioning) is not new. Killeen (1981) showed that pigeons can discriminate whether their own behavior or "something else" caused changes in a light. How associative models can accommodate the specific data of Blaisdell et al. is a challenge, but it is hardly compelling evidence that rats are propositional reasoners.

Mitchell et al. are curiously silent about the emergence of associative learning. Must we also assume that human infants engage in complex propositional reasoning? If so, then this proposal is difficult to reconcile with studies of human development. Clancy et al. (1976) found that use of the *if*-clause (the meaning, without containing the connective) emerges at 2 or 3 years of age. And, "If p, then q" propositions (the type involved in blocking) are not understood until children are 6 years old (Braine & Rumain 1983). Importantly, there seems to be consensus that children's early-developing inferences are likely to be acquired as part of learning their language (Braine & Rumain 1983; Falmagne 1975). Does that mean that human infants who have not yet acquired language cannot learn about regularities between environmental events? Doubtful. Sobel and Kirkham (2006) found backward blocking in 8-month-old children; so, infants exhibit the same associative learning phenomena as do animals and human adults.

The great advantage of associative accounts is that their mechanisms seem to be available to all species across all developmental stages. Without explaining its origin and its developmental trajectory, we cannot fully comprehend any psychological process, particularly one as essential as associative learning.

Rational models of conditioning

doi:10.1017/S0140525X09000922

Nick Chater

Division of Psychology and Language Sciences, University College London, London WC1E 6BT, United Kingdom.

n.chater@ucl.ac.uk

http://www.psychol.ucl.ac.uk/people/profiles/chater_nick.htm

Abstract: Mitchell et al. argue that conditioning phenomena may be better explained by high-level, rational processes, rather than by non-cognitive associative mechanisms. This commentary argues that this viewpoint is compatible with neuroscientific data, may extend to nonhuman animals, and casts computational models of reinforcement learning in a new light.

Mitchell et al. provide an important critical challenge to the presuppositions underlying current theories of human conditioning, both in psychology and the neurosciences. They suggest that human contingency learning results from reasoning processes over propositional knowledge, rather than from an elementary process of forming associations. This commentary focuses on three questions raised by this analysis, and concludes with a perspective on the origin of contradictory forces in the control of behavior which does not invoke a clash between a cognitive and associative system.

Multiple neural systems for decision making? Mitchell et al. argue that behavioral evidence makes a case against a distinct associative learning system. Yet the idea that there are multiple, competing, neural systems underpinning decision making is very widespread within neuroscience. One line of evidence for multiple systems comes from double dissociations in human neuropsychology, and, perhaps most strikingly, from animal lesion studies (see, e.g., Coutureau & Killcross 2003; Killcross & Coutureau 2003). Yet such studies provide only tentative evidence for functionally distinct systems, rather than differential engagement of a single system (Chater 2003; Shallice 1988). Consider an analogy with allergies: Some people cannot eat prawns, but can eat pine nuts; other people can eat pine nuts, but not prawns. But we cannot, of course, conclude that there are two distinct digestive systems that process these different foods. Instead, a *single* digestive system deals almost uniformly with all foods, but exhibits two biochemical "quirks" leading to the selective allergies. Thus, a *single* processing system can in principle yield striking double dissociations of function (Chater, in press). Hence, double dissociations in humans, and animal lesion studies yielding double dissociations, are weak evidence for distinct processing systems. The same caveats apply to studies in which reinforcement learning is selectively impaired not by a lesion, but by a pharmacological intervention (e.g., a dopamine agonist, Pizzagalli et al. 2008). Similar issues arise, too, with neuroimaging studies. Such studies reveal differential neural activity under different task conditions. But such differential activity may nonetheless be entirely compatible with the existence of a single, unitary, decision-making system.

Is animal conditioning associative? Mitchell et al.'s account may be correct with regard to people. But perhaps rats really do use dedicated associative learning mechanisms. Indeed, this latter assumption is widespread in the comparative literature (e.g., Mackintosh 1983). Nonetheless, there are at least three reasons to doubt this. (1) Many aspects of animal cognition are highly sophisticated and seem to go far beyond the scope of purely associative mechanisms (e.g., Wasserman & Zentall 2006). (2) Associative theories of learning typically assume gradual modifications; yet actual behavior is roughly all-or-none (Gallistel et al. 2004), just as though the animal is adopting or rejecting a hypothesis about possible environmental contingencies. The familiar smooth learning curves arise only from data averaging. (3) Putative conditioning phenomena in animals appear to be highly sensitive to rational factors (Courville et al. 2006). So, for example, blocking (Kamin 1969) can be rationally understood in terms of "explaining away" (Pearl 1988); the slower rate of extinction from partially reinforced contingencies has a natural statistical explanation; and so on.

The role of computational models of reinforcement learning. There have been remarkable recent developments in computational models of reinforcement learning (Dayan & Abbott 2001) – often implicitly or explicitly viewed as capturing the computational principles of a distinct, striatal, non-cognitive, learning system (Jog et al. 1999). If Mitchell et al. are right, then such computational models should perhaps be interpreted differently: as providing an account of rational inferences that can be drawn from data concerning actions and rewards, given minimal background knowledge. But where background knowledge *is* available (e.g., about likely causal connections between actions, events, and rewards), we should expect that such knowledge will be incorporated appropriately (Gopnik & Schulz 2007). According to this perspective, computational models of reinforcement learning apply to a narrow class of situations, in which background causal knowledge is restricted, rather than describing the operation of a particular neural system that drives behavior.

Clash of reasons, not clash of mechanisms. One intuitive appeal of the idea of a split between associative and cognitive systems, competing for the control of behavior, is a potential explanation for many paradoxical aspects of human behavior, both in laboratory studies of, for example, time-discounting and weakness-of-will and in real-world phenomena of addiction, depression, or phobias (Epstein 1994; McClure et al. 2004).

If, following Mitchell et al., we reject evidence for a distinct associative system, how are we to explain the origin of internal cognitive conflict? One straightforward approach (Chater, in press) is to propose that internal conflict arises from a "clash of reasons" rather than a clash of systems. In almost all nontrivial reasoning problems, different lines of argument appear to favour different conclusions. One source of reasons, among many, may be past experience (including the "reinforcement history"). Moreover, reasons are often not equally persuasive; nor are they equally easy to evaluate. When paying close attention and given sufficient time, it may become evident one reason is valid, whereas another reason is weak. But when attention is reduced, the weaker reason may nonetheless prevail. Therefore, to choose a classic example from probabilistic reasoning, the reasoner may decide that, given information about, say, Linda's intellectual and political background, it is more likely that Linda is a feminist bank teller, than that she is a feminist (Tversky & Kahneman's [1983] conjunction fallacy), because there is a better overall match with the former description (for which at least the first part matches), than the second description (which seems entirely incongruous). Considered reflection on probability may, or may not, lead the reasoner to draw the opposite conclusion.

More generally, it seems entirely possible that there will be systematic differences between responses when time and attention are limited and responses when time and attention are plentiful (see Cunningham & Zelazo [2007] for a similar perspective on apparent dissociations between two putative routes underpinning social cognition, as exemplified by, e.g., Bargh & Chartrand 1999); and concomitant differences in the degree to which brain areas are activated in the contemplation of different reasons. In summary, observing battles for control of the behavioral "steering wheel," and evidence for different behavioral and neural bases for the competitors, need not be interpreted as indicating a clash between distinct mechanisms (e.g., associative vs. cognitive), but might equally arise from a clash of reasons within a unified cognitive system.

Is propositional learning necessary for human autonomic classical conditioning?

doi:10.1017/S0140525X09000934

Michael E. Dawson[a] and Anne M. Schell[b]

[a]*Department of Psychology, University Southern California, Los Angeles, CA 90089-1061;* [b]*Psychology Department, Occidental College, Los Angeles, CA 90041.*

mdawson@usc.edu

schell@oxy.edu

http://college.usc.edu/labs/dawson/home/index.cfm

http://departments.oxy.edu/psych/index.html

Abstract: Additional support is presented for the necessity of awareness of the CS-US relation in human autonomic conditioning. However, possible limitations and exceptions regarding this general rule are discussed. Limitations include the lack of relationship between conditioned response (CR) strength and degree of awareness, and an important exception may be the finding of conditioning with backwardly masked CSs of a biologically prepared nature.

Mitchell et al., in their interesting and provocative treatment of human associative learning, link propositional learning with awareness. Specifically, they conclude that "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" (sect. 8, para. 2). Our research strongly supports this conclusion regarding human autonomic classical conditioning, as the authors noted. Therefore, our comments focus on the role of awareness of the CS-US (conditioned stimulus-unconditioned stimulus) contingency in human autonomic classical conditioning.

Confirming findings. Research from our laboratories is even more supportive than Mitchell et al. indicate of the position that human autonomic classical conditioning is propositional and requires significant cognitive resources. As they point out, much of our research has embedded conditioning within a distracting cognitive masking task with the result that subjects who become aware of the CS-US contingency successfully condition, and subjects who remain unaware do not. They then raise the "devil's advocate" possibility that an automatic associative-link mechanism might exist, but the cognitive load imposed by a masking task may act to prevent conditioning by reducing the degree to which the CS and US are processed, and hence reduce the input to the link mechanism. However, in the cognitive masking task that we have used, the CS is specifically the focus of the subject's attention (it must be judged on some dimension and remembered), and expectancy of the US is constantly reported by the subject.

For instance, subjects were presented with a series of tones on each trial, one of which was the CS, and were required to determine which tone matched a preceding tone in pitch. Subjects indicated expectancy or non-expectancy of the US by pressing a series of buttons continuously during the tones (Dawson & Biferno 1973). Hence, the failure to condition without awareness cannot be attributed to the failure to attend and process the CS and US. We have found this necessity of awareness not only with typically used CSs, such as tones or colored lights, but also with odor CSs, stimuli often thought to be capable of eliciting conditioned emotional responses without a supporting conscious memory (Marinkovic et al. 1989). Another line of evidence in support of the importance of cognitive resources not mentioned by Mitchell et al. is that performance on a secondary reaction time task performed during the conditioning session shows deterioration during the CS exactly when the conditioned responses are elicited (Dawson et al. 1982).

Perplexing exceptions. Although we are strong advocates of the position that human classical conditioning cannot occur without awareness of the CS-US relationship, we find ourselves in the unusual position of noting that there may be limitations and exceptions to this general proposition. First, Dawson and Furedy (1976) reviewed evidence in support of a "necessary-gate" hypothesis that included the following propositions: (1) contingency awareness is necessary, but not sufficient, for human autonomic classical conditioning (e.g., researchers in this field often observe participants who are aware of the CS-US relationship, give strong URs [unconditioned responses], but do not show conditioning); (2) the degree of contingency awareness has a gate, but not analog, relation to the strength of the conditioned response (i.e., once a critical minimum level of awareness has developed, there is little or no relationship between the strength of the conditioned autonomic response and the degree of accuracy or certainty in the learned proposition); and (3) contingency awareness is not necessary for performance of a response that has been previously conditioned. This second proposition is contrary to the position of Mitchell et al. in section 3.1 that the strength of the CR *is* related to the "strength of belief" in the CS-US contingency.

Second, conditioning using biologically prepared CS-US relations (e.g., angry faces associated with aversive events) may be possible without awareness of the presence of the CS, as demonstrated by Öhman and his colleagues (see Esteves et al. 1994). Esteves et al. (1994) used as CSs in a discrimination conditioning paradigm pictures of angry or happy faces that were backwardly masked to prevent conscious awareness of their presence. These CSs were paired with an electric shock US. Conditioned skin conductance responses were observed following the angry face but not the happy face CS (see also Öhman & Soares 1998; Morris et al. 2001). These results indicate that humans can be conditioned to a stimulus they do not consciously perceive, if that stimulus is evolutionarily prepared to be associated with the US. However, when biologically prepared CS-US relations were embedded in a distracting cognitive masking task, which ensured conscious perception of the individual CS, autonomic conditioning only occurred among aware subjects and only after they became aware (Dawson et al. 1986).

Thus, there is conflicting evidence of whether autonomic conditioning can occur without awareness when biologically prepared CS-US relations are involved. When biologically prepared CSs are backwardly masked and subjects are presumably unaware of the CSs' presence, there is evidence of unaware conditioning. When the CSs are part of a distracting masking task, one which ensures that the CSs are perceived and discriminated, there is no evidence of unaware conditioning.

A possible explanation of these conflicting results is that when higher cortical processes become involved, as when the CSs are perceived during a distracting cognitive task, propositional learning is the dominant force in controlling autonomic conditioning. Propositional learning will be dominant even if an incorrect proposition has been tacitly learned – that the CS has no particularly predictive value. However, when these higher cortical processes concerning the CSs are prevented from occurring, as when the CSs are effectively backwardly masked, and when they are biologically prepared, then the automatic associative learning processes become the dominant force. Therefore, under conditions where the CSs are perceived, propositional learning is necessary for human autonomic classical conditioning.

Straw-men and selective citation are needed to argue that associative-link formation makes no contribution to human learning

doi:10.1017/S0140525X09000946

Dominic M. Dwyer,[a] Michael E. Le Pelley,[a] David N. George,[a] Mark Haselgrove,[b] and Robert C. Honey[a]

[a]*School of Psychology, Cardiff University, Cardiff CF10 3AT, United Kingdom;* [b]*School of Psychology, University of Nottingham, Nottingham NG7 2RD, United Kingdom.*

dwyerdm@Cardiff.ac.uk lepelleyme@Cardiff.ac.uk
georged@Cardiff.ac.uk mark.haselgrove@Nottingham.ac.uk
honey@Cardiff.ac.uk
http://www.cardiff.ac.uk/psych/contactsandpeople/lecturing/dwyer-dominic-m-overview_new.html
http://www.cardiff.ac.uk/psych/contactsandpeople/lecturing/le-pelley-mike-overview_new.html
http://www.cardiff.ac.uk/psych/contactsandpeople/researchfellows/george-d-n-dr-overview_new.html
http://www.psychology.nottingham.ac.uk/staff/mxh/
http://www.cardiff.ac.uk/psych/contactsandpeople/lecturing/honey-rob-overview_new.html

Abstract: Mitchell et al. contend that there is no need to posit a contribution based on the formation of associative links to human learning. In order to sustain this argument, they have ignored evidence which is difficult to explain with propositional accounts; and they have mischaracterised the evidence they do cite by neglecting features of these experiments that contradict a propositional account.

In their target article Mitchell et al. contend that associative learning is best explained as the result of effortful cognitive processes based on propositions, and that there is no need to posit a contribution based on the formation of associative links. The target article greatly overstates the case for rejecting the contribution of associative links, however, because of its remarkably selective citation and interpretation of prior work.

Although there are situations where propositional accounts provide good explanations of human learning, there are also many examples where the pattern of results is more easily explained by associative-link accounts. For example, if beliefs are based on post hoc reasoning about events stored in memory, changing the temporal order in which these events are stored should not affect judgments; and yet several studies have demonstrated trial-order effects in learning (Collins & Shanks 2002; Dickinson & Burke 1996; López et al. 1998b). Similarly, the evidence relating to differences in diagnostic and predictive learning is overstated: many studies have found no impact of this manipulation (e.g., Arcediano et al. 2005; Cobos et al. 2002; López et al. 2005; Tangen et al. 2005). Nor is mention made of cancer patients who form food aversions despite being fully aware that chemotherapy treatment, rather than any food consumed, produced their illness (e.g., Bernstein & Webster 1980). But the most surprising omission is of research on blocking by two of the target article's authors themselves: Mitchell et al. (2006) argued that for participants to reason that a blocked cue *T* is non-causal requires knowledge of the outcome with which *T* was paired. Their results, however, revealed blocking of *T* without any memory of the *T*–outcome relationship, which led them to conclude that "associative cue-competition would appear to be a strong candidate mechanism for noninferential forward blocking in humans" (p. 842).

In the target article, Mitchell et al. also neglect dissociations in patterns of learning under different circumstances that are challenging for single-process propositional accounts, but which follow naturally from dual-process approaches to learning. For example, López et al. (2005) demonstrated that participants were generally insensitive to the difference between predictive and diagnostic tasks but did show differences when instructions made clear the importance of this distinction in causal order. Likewise, although Mitchell et al. cite Shanks and Darby (1998) as providing evidence for propositional learning, they ignore the fact that results consistent with this account were only observed in participants who had learnt the task well; those who learnt less well showed responses consistent with associative-link theories (similar responses to novel compounds and the elements that comprise those compounds). Le Pelley et al. (2005a) demonstrated that participants show an unblocking effect when information is presented on a trial-by-trial basis, but not when presented in a questionnaire format that would facilitate the use of propositional learning. Moreover, Mitchell et al.'s analysis of this unblocking effect does not stand up to scrutiny. Participants could indeed remember what had changed between *A–O1O2* and *AT–O1O3* trials: Their accuracy in predicting O1 on *AT–O1O3* training trials was significantly higher than that for their predictions of O3.

It is unfortunate that Mitchell et al. address only a straw-man version of associative-link accounts. For example, they assert that associative theories see learning as proceeding without awareness. Few associative theorists would agree with this characterisation, however: Why should people *necessarily* remain unaware of links that are formed? Although the issue of awareness is orthogonal to associative accounts (Shanks 2007), the target article explicitly states that any example of learning without awareness would be highly damaging to the idea that all learning is propositional. Yet, Mitchell et al. cite two examples of dissociations between propositional knowledge and conditioned reactions in humans (flavour conditioning and the "Perruchet effect"), and the essential features of both studies have been replicated (see Lovibond & Shanks 2002). In addition, the fact that associative learning can occur in anaesthetised animals (see Lovibond & Shanks 2002) indicates that the idea that propositional mechanisms can explain all animal learning (other than that produced by S-R mechanisms) must be, beyond reasonable doubt, false (see also Iselin-Chaves et al. [2005] for learning under anaesthesia in humans). It warrants repeating that any dissociation between proposition knowledge and learning is fatal to the current account, and such dissociations do exist (albeit that unambiguous evidence is not widespread).

The insistence that associative accounts rely on nodes that represent whole stimuli in a symbolic manner is also a mischaracterisation. Foreshadowed by Estes's (1950) stimulus sampling theory, associative models explicitly acknowledge that any stimulus comprises multiple features that might each be shared with other stimuli (e.g., Blough 1975; Brandon et al. 2000). This undermines attempts in the target article to characterise the concept of generalisation as an unjustified assumption by which "freedom is gained to explain results" (sect. 6.1, para. 3). In fact, this is an integral and fully specified feature of almost all current associative learning models, and flows directly from the idea that whole stimuli should be considered as collections of potentially overlapping features. Within-compound associations are also treated as "get-out clauses" despite following naturally from, and being explicitly predicted by, standard associative principles. Furthermore, there is evidence for their existence (Rescorla & Durlach 1981) and influence upon cue-competition (e.g., Batsell et al. 2001; Durlach & Rescorla 1980). Although there are examples of particular associative-link models being modified in light of an inability to account for particular results, this does not undermine the fact that principles of generalisation and within-compound associations are instantiated within associative-link models as a class.

Finally, Mitchell et al. criticise associative theory for lacking parsimony because it must predicate two sources for human learning (associative-link and propositional mechanisms). However, associative-link theories are very parsimonious in other ways. Most notably, they can explain aspects of human learning (e.g., sensitization, habituation, perceptual learning) which lie beyond propositional mechanisms. Although associative models inherently require dual-process accounts of human learning, propositional accounts are inherently multiple-process with respect to other phenomena. Hence, proposition-only accounts of human learning are no more parsimonious than dual-process accounts when considered in a broader context.

Operating principles versus operating conditions in the distinction between associative and propositional processes

doi:10.1017/S0140525X09000958

Bertram Gawronski[a] and Galen V. Bodenhausen[b]

[a]*Department of Psychology, The University of Western Ontario, Social Science Centre, London, Ontario N6A 5C2, Canada;* [b]*Department of Psychology, Northwestern University, Evanston, IL 60208-2710.*

bgawrons@uwo.ca **http://publish.uwo.ca/~bgawrons/**

galen@northwestern.edu

http://www.psych.northwestern.edu/psych/people/faculty/bodenhausen/

Abstract: Drawing on our Associative-Propositional Evaluation (APE) Model, we argue for the usefulness of distinguishing between basic operating principles of learning processes (associative linking vs. propositional reasoning) and secondary features pertaining to the conditions of their operation (automatic vs. controlled). We review empirical evidence that supports the joint operation of associative and propositional processes in the formation of new associations.

In contrast to a common assumption of dual-process models, Mitchell et al. argue that the formation of new associations in human memory is an exclusive product of controlled, propositional inferences, and that there is no empirical evidence for automatic processes of associative linking. In response to Mitchell et al.'s conclusion, we argue that their analysis conflates the distinction between the basic operating principles of a given process (i.e., associative linking vs. propositional reasoning) and secondary features pertaining to the conditions of its operation (i.e., automatic vs. controlled). If the conceptual independence of these dimensions is taken into account, the reviewed evidence regarding features of automaticity will be diagnostic about the operation of a particular type of process only to the degree that there is perfect overlap between the two dimensions (automatic = associative; controlled = propositional) – which seems debatable on both conceptual and empirical grounds.

Based on our own Associative-Propositional Evaluation (APE) Model (Gawronski & Bodenhausen 2006; 2007), we argue that the formation of a new association in memory should be understood as an *effect* that could be the result of two conceptually distinct *mechanisms*, associative linking and propositional reasoning. In our APE Model, we define *associative linking* as the creation of a new association between two concepts based on the mere co-occurrence of objects or events independent of the perceived validity of their relation. *Propositional learning* is defined as the creation of a new association as a result of syllogistic inferences about the validity of a given relation. The primary difference between the two processes is their dependency on subjective validity, in that only propositional learning, but not associative linking, takes the perceived validity of relations into account (see also Strack & Deutsch 2004). As such, the two mechanisms should lead to the same outcome when the co-occurrence of two objects or events is interpreted as reflecting a valid relation. However, the two mechanisms may lead to different outcomes when the co-occurrence between two objects or events is regarded as non-diagnostic or invalid. This conceptualization incorporates Mitchell et al.'s emphasis of truth values as a core feature of propositional reasoning. However, it differs from Mitchell et al.'s approach, in that assumptions about automatic features represent empirical claims about the boundary conditions of the operation of the two processes, rather than defining characteristics that could be conversely used to identify their operation in a particular case.

To empirically distinguish between the two processes, we suggest that the actual operation of associative and propositional processes should be identified by means of their interactive effects on associations and beliefs. In the APE Model, we define *associations* as mental links between concepts independent of their subjective truth or falsity; *beliefs* are defined as the endorsed relations that are implied by validated or invalidated associations. This distinction has proven its usefulness in the social-cognitive literature, showing that activated associations can produce behaviors that are congruent with these associations, even when the relations implied by these associations are regarded as invalid (for a review, see Strack & Deutsch 2004).

More importantly, there is suggestive evidence that such dissociations can sometimes be due to antagonistic effects of associative linking and propositional reasoning during the encoding of new information (e.g., Gawronski et al. 2008; Rydell et al. 2006), supporting the usefulness of the proposed distinction in the formation of new associations. The basic notion of these studies is that the mere co-occurrence between two objects can create a mental association between these objects, even though the validity of the implied relation is rejected at the propositional level. Empirically, these differences are often reflected in dissociations between implicit and explicit measures (Fazio & Olson 2003), such that implicit measures (e.g., sequential priming tasks) reflect the mere co-occurrence between the two objects, whereas explicit measures (i.e., self-reported judgments) reflect the perceived validity of the implied relation.

Other evidence that is consistent with the notion of associative linking comes from research on spontaneous trait transference (e.g., Skowronski et al.1998), in which communicators have been shown to become associated with the traits they ascribe to others. In most cases, there is no logical basis to infer that a communicator has a particular trait (e.g., tidy) simply because he or she describes that trait in another person. Hence, it seems reasonable to assume that any such associations are the product of associative linking rather than propositional reasoning (Carlston & Skowronski 2005). To be sure, such associative linking processes may still depend on perceivers' attention, processing goals, or awareness of the co-occurrence. However, this by itself does not make the underlying learning process propositional, as defined in the proposed conceptualization.

Another important issue in this context is Mitchell et al.'s concern that proposing mutual interactions between associative and propositional processes would make the distinction between the two processes obsolete. Such interactions are a core assumption of our APE Model, which assumes that mutual interactions between the two processes are reflected in different mediation patterns of experimentally induced effects on activated associations and endorsed beliefs (Gawronski & Bodenhausen 2006). Specifically, we argue that associative linking will often produce parallel effects on associations and beliefs, such that newly created associations provide the basis for explicitly endorsed beliefs. Conversely, newly created associations may be the product of propositional inferences, such that new beliefs generated in the course of validating currently accessible information may be stored in associative memory. Drawing on the abovementioned distinction between implicit and explicit measures, the first case is assumed to produce parallel effects on both kinds of measures, with effects on the explicit measure being fully mediated by the implicit measure. In contrast, the second case should produce parallel effects on both kinds of measures, with effects on the implicit measure being fully mediated by the explicit measure.

An illustrative demonstration of these diverging mediation patterns is a recent study by Whitfield and Jordan (submitted), who combined an implicit evaluative conditioning (EC) procedure (Olson & Fazio 2001) with a propositional impression formation task that used descriptive information about the conditioned stimulus. Their results showed that both the EC procedure and the impression formation task produced parallel effects on both explicit and implicit measures. However, in line with the predictions of the APE Model, EC effects on the explicit measure were fully mediated by the implicit measure, whereas impression formation effects on the implicit measure were fully mediated by the explicit measure (for related findings, see Gawronski & LeBel 2008; Gawronski & Strack 2004; Gawronski & Walther 2008).

Taken together, these results suggest that a conceptual distinction between associative and propositional processes in terms of their operating principles (rather than automatic vs. controlled features) has testable and empirically supported implications. More importantly, our analysis implies that the formation of new associations in memory can be the product of either associative or propositional processes, and that Mitchell et al.'s insightful review may speak only to the automatic versus controlled nature of these processes rather than to the general irrelevance of associative processes in human learning.

Rational constructivism: A new way to bridge rationalism and empiricism

doi:10.1017/S0140525X0900096X

Alison Gopnik

Department of Psychology, University of California at Berkeley, Berkeley, CA 94704.

Gopnik@berkeley.edu

Abstract: Recent work in rational probabilistic modeling suggests that a kind of propositional reasoning is ubiquitous in cognition and especially in cognitive development. However, there is no reason to believe that this type of computation is necessarily conscious or resource-intensive.

There is a paradox at the heart of cognitive science. Human beings (and some animals) seem to have abstract, hierarchical, structured representations of the world. These representations allow us to make a wide range of novel predictions and produce a wide range of novel behaviors. And these representations seem to be accurate – they capture the structure of the world, and they improve as we learn more about the world. But the information provided by our senses, our one direct source of evidence about the world, is very different from these representations. It is a noisy, probabilistic, and chaotic set of contingencies among specific concrete inputs, apparently far removed from the true structure of the world itself.

In the past 2000 years of western philosophy, and the past 50 years of cognitive science, there have been two very different approaches to resolving this paradox. One tradition (nativist, rationalist, propositional, "East Coast") argues that cognition does indeed involve abstract, hierarchical, structured representations. It only appears, however, that we infer these representations from the evidence of our senses. In fact, these representations must be there innately, and are only slightly modified by learning. Small details may be filled in by experience, or alternative parameters may be triggered by different experiences. But the fundamental structure of the representations is there from the start. The alternative tradition (empiricist, associationist, connectionist, "West Coast") argues that it only appears that we have abstract, hierarchical, structured representations. In fact, our novel predictions and behaviors are based on the complex contingency patterns among individual sensory inputs, patterns that we extract through associative mechanisms.

There have sometimes been arguments for a kind of dismissive co-existence between these two approaches. The rationalists say that most cognition is the result of innate abstract representations, but mere associationist processes may play a role in very automatic, low-level kinds of behavior. The empiricists say that associations are responsible for most cognition, but there may be explicit, conscious, and sophisticated propositional reasoning layered on top. These two-process views both suggest that there is some relationship between the sophistication, power, and likely domain of the representations and their computational character – associations are "low-level" and propositions are "high-level." They just disagree on whether most cognition falls on one side or the other.

The target article is in this general tradition, though it endorses the idea that propositional representations can account for even classical associationist phenomena, such as conditioning. But Mitchell et al. also argue that the propositional representations they endorse are resource-intensive, subject to conscious reflection, and can be understood as beliefs – they are "high-level."

In cognitive development, going back to Piaget, there has been a long tradition of trying to elude the rationalist/empiricist dichotomy with "constructivist" theories. A constructivist account should allow us to actually infer highly structured representations accurately from patterns of contingency in the data. The most recent constructivist project has been the "theory theory" – the idea that children develop intuitive theories from evidence in the way that scientists do. But the theory theory, like earlier constructivist theories, has suffered from a lack of computational precision and specific learning mechanisms.

However, in the last 10 years or so there has been increasing excitement about a new theoretical view that provides a computationally rigorous basis for the constructivist project. This approach might be called "rational probabilistic modeling." This view, unlike classical empiricist views, proposes structured, abstract, hierarchical representations. But unlike classical rationalist views, it sees those representations as probabilistic and learned. Moreover, the kind of learning that is involved is not the simple method of association, but is a form or rational probabilistic induction, often involving Bayesian methods. This general theoretical approach has been applied to a very wide range of kinds of cognition, including "low-level" automatic cognition such as vision, motor control, and syntax, as well as "high-level" conscious cognition such as category and word learning and causal learning. (For some recent examples and reviews of this work see Chater & Manning 2006; Chater et al. 2006; Gopnik & Schulz 2004; Gopnik & Tenenbaum 2007; Gopnik et al. 2004; 2007; Regier & Gahl 2004, Tenenbaum et al 2006; Xu & Tenenbaum 2007; Yuille & Kersten 2006).

Causal knowledge and learning, one of the foci of the target article, has been a particularly fruitful venue for these new theories. In our own work, we have shown that even very young children represent the causal structure of the world and reason about those representations in a rational way, in accordance with the principles of causal Bayes nets (Gopnik et al. 2004). As Mitchell et al. mention, this "causal Bayes net" approach seems to be very convergent with the approach that is presented here. But there is one important difference. The computations that are involved in rational probabilistic models have no necessary link to issues of high- versus low-level, animal versus human, conscious versus unconscious, or resource-dependent versus automatic.

Indeed, many of these models have their conceptual roots in vision science. At least since Irv Rock and arguably since Helmholz, vision scientists have seen vision as a process of hypothesis generation and testing. The visual system inferentially reconstructs an accurate representation of the visual world; it solves "the inverse problem." We know that these inferential processes are much more constrained and complex than simple associations, and they have been well-modeled as a kind of Bayesian inference. But they are unconscious, automatic, and low-level (see Yuille & Kersten 2006).

Thinking about the issue developmentally also makes this point vivid. We know that even very young babies are capable of sophisticated kinds of statistical and inductive reasoning – reasoning capacities that go far beyond simple associative mechanisms. In fact, arguably infants have more powerful learning capacities than adults. It seems unlikely, however, that infant resource allocation or consciousness parallels that of adults, though undoubtedly infants are conscious. From a computational point of view, propositional reasoning does indeed go all the way down, as Mitchell et al. argue. The same kinds of rational computations play an essential role in cognition from vision to causation, and from infancy to adult science. But this is orthogonal to the question of how those computations are related to resource management or phenomenology.

Cognition, consciousness, and the cognitive revolution

doi:10.1017/S0140525X09000971

John D. Greenwood

CUNY Graduate Center and Ph.D. Program in Philosophy, City University of New York, New York, NY 10016.

jgreenwood@gc.cuny.edu

http://web.gc.cuny.edu/philosophy/people/greenwood.html

Abstract: It is argued that the cognitive revolution provided general support for the view that associative learning requires cognitive processing, but only limited support for the view that it requires conscious processing. The point is illustrated by two studies of associative learning that played an important role in the development of the cognitive revolution, but which are surprisingly neglected by Mitchell et al. in the target article.

I would like to make some historical remarks about the late twentieth century development of conditioning theory during the time of the "cognitive revolution" in psychology, that I hope will have some epistemic bearing on the theory of the propositional nature of human associative learning advanced by Mitchell et al.

It is true that the cognitive revolution in psychology was marked by a general move from theories based upon stimulus-response (S-R) or response-reinforcement connections, to theories based upon the cognitive processing of representations, including cognitive theories of classical and operant conditioning (Greenwood 2001; 2008). This progression may be said to provide general support for Mitchell et al.'s view that human associative learning is propositional in nature. However, it is not true that the cognitive revolution provided general support for the view that such cognitive processes are consciously mediated. On the contrary, a common and justified response to neobehaviorist critiques of the cognitive revolution, to the effect that it marked a return to the bad old days of "introspective psychology" (Amsel 1989; Skinner 1985), was that contemporary cognitive psychology had demonstrated that human subjects have very limited access to cognitive processes (Nisbett & Wilson 1977; Nisbett & Ross 1980). And, in general, Mitchell et al. need to distinguish evidence for the view that associative learning is based upon cognitive processing – much of which is drawn from the study of animal learning – from evidence for the view that associative learning is based upon conscious representation of association – much of which is drawn from the study of human learning.

One puzzling feature of Mitchell et al.'s defense of a propositional theory of associative learning is that they embrace a feature of the associationist tradition that goes back to Hume; namely, the view that the strength of a learned association is a function of the number of times ideas or stimuli, responses, and reinforcement are observed to occur together. For if one assumes this feature, then it is natural to think of learning as an automatic function of frequency of repetition: that the connection between even cognitive or propositional representations is "stamped in," as Edward B. Thorndike (1898) put it. Yet this feature does not appear to play any critical role in a propositional theory of associative learning.

Hence, I was surprised that the authors did not mention Jose Garcia's (Garcia & Koelling 1966) studies of conditioned taste aversion, perhaps the most powerful challenge to the traditional view. It was an axiom of the behaviorist learning tradition that the optimal temporal interval for conditioned learning of a connection between unconditioned and conditioned stimuli (in classical conditioning) or response and reinforcement (in operant conditioning) is a fraction of a second, and that conditioned learning requires the repeated association of stimuli, responses, and reinforcement. Yet Garcia's studies demonstrated that rats could learn to avoid saccharin water after a single trial and with up to a 12-hour interval between their drinking saccharin water and the artificial inducement of radiation sickness. So antithetical was this result to the behaviorist learning tradition that one of Garcia's professors at UC-Berkeley told him that such an outcome was frankly impossible (Bolles 1993), and Garcia had great difficulty in getting his studies published in mainstream psychology journals (Lubek & Apfelbaum 1987).

These studies provided powerful evidence for cognitive processing in conditioned learning. Garcia's rats were able to identify potential environmental causes of sickness via a cognitive process analogous to Mill's method of difference. As Mackintosh (1978) put it, these and later studies (Kamin 1969; Revuski 1971) demonstrated that:

> Simple associative learning is simple in name only. Animals do not automatically associate all events that happen to occur together. If they did, they would be at the mercy of every chance conjunction of events. In fact, they behave in an altogether more rational manner. By conditioning selectively to good predictors of reinforcement at the expense of poor predictors, and by taking their past experience

into account, they succeed in attributing reinforcers to their most probable causes. (Mackintosh 1978, p. 54)

However, these studies did nothing to demonstrate that such forms of conditioned learning depend on consciousness of the connection between novel stimuli and their consequences. So one needs to carefully distinguish the evidence for cognitive processing in associative learning from evidence that consciousness of a connection is necessary for associative learning.

Which is not to say that there is no evidence for this view, or that none was forthcoming during the course of the cognitive revolution. So I was also surprised that Mitchell et al. did not mention Dulany's (1968) studies of verbal conditioning. Earlier studies had indicated that human subjects' employment of linguistic items – such as use of plural nouns – could be manipulated by social reinforcement without their awareness, a form of conditioning commonly known as the "Greenspoon effect" (Greenspoon 1955). Dulany's work suggests that in many of these studies, not only were subjects conscious of the relevant response-reinforcement connection, but that consciousness was a condition of associative learning.

My own favorite example (albeit anecdotal) is the following episode described in Skinner's (1987) book *On Further Reflection*, in which he reminisced about his attempt to instrumentally condition his daughter's foot movements – when she was 3 years old – by rubbing her back:

> I waited until she lifted her foot slightly and then rubbed briefly. Almost immediately she lifted her foot again, and again I rubbed. Then she laughed. "What are you laughing at?" I said. "Every time you raise my foot you rub my back!" (Skinner 1987, p. 179)

Learning in simple systems

doi:10.1017/S0140525X09000983

Geoffrey Hall
Department of Psychology, University of York, York YO10 5DD, United Kingdom.
gh1@york.ac.uk
http://www.york.ac.uk/depts/psych/www/people/biogs/gh1.html

Abstract: Studies of conditioning in simple systems are best interpreted in terms of the formation of excitatory links. The mechanisms responsible for such conditioning contribute to the associative learning effects shown by more complex systems. If a dual-system approach is to be avoided, the best hope lies in developing standard associative theory to deal with phenomena said to show propositional learning.

After experiencing an electric shock following a squirt of water to its siphon, *Aplysia* will show a change in its behavior. The water squirt will come to evoke a dramatic gill withdrawal (a response readily evoked by the shock, but, previously, only weakly by the squirt itself). This is a classic example of associative learning as a *phenomenon*. What is the mechanism responsible for this phenomenon?

Throughout most of the target article, Mitchell et al. are unambiguous in their answer to this sort of question. The associative learning effect depends, they say, on an effortful, attention-demanding, reasoning process that produces conscious, propositional knowledge about the relationship between events. This is the only mechanism normally allowed. Why do they waver (as they do) when it comes to the case of *Aplysia*? (Although they focus on human learning, the authors assert that the human system is not likely to be unique and that a process of belief acquisition also underlies animal learning.) The obvious answer – that it seems implausible to attribute such processes to a mollusc equipped with just a few hundred neurons – means that they must find some other explanation for learning in *Aplysia*. What they offer (a reflexive, stimulus-response [S-R] mechanism) is entirely in accord with what we know from neurophysiological research on this animal (Carew et al. 1983) and will be widely accepted.

But Mitchell et al.'s acceptance of this analysis has major implications for their central thesis, implications that they scarcely acknowledge. If classical conditioning procedures can produce S-R learning in *Aplysia*, might they not do so elsewhere? And the evidence currently available (see Hall 2002, for a recent review) supports the conclusion that S-R association formation plays a role in generating the conditioned responses shown by higher vertebrates even in the more complex training procedures used for these animals. In allowing the existence of this mechanism, Mitchell et al. have let in, by the back door, a version of the dual-system approach that they profess to reject entirely.

It should be acknowledged, however, that modern studies of classical conditioning in animals, conducted within the associative tradition, have been concerned to show that the effect goes beyond simple S-R learning. As Mitchell et al. themselves point out, it is difficult to explain, by way of the S-R mechanism, the observation that procedures designed to change the value of an outcome (such as sating the animal for a given food) will reduce the vigor of a conditioned response evoked by a stimulus that has previously been paired with that food. This observation has been taken to indicate that the animal has learned something about the relationship between the stimulus and its outcome. Perhaps we need to turn our attention to a more modest version of Mitchell et al.'s thesis, considering its application to just this form of conditioning. Perhaps this form of conditioning, at least, is solely to be explained in terms of propositional reasoning.

Standard associative learning theory offers an alternative interpretation. It suggests that the conditioning procedure establishes an excitatory link between the central representations of the signal and its consequence (i.e., it envisages an S-S [stimulus-stimulus], as opposed to an S-R association). As the signal evokes a response by way of its excitatory effect on the representation of the outcome, the effects of outcome devaluation are readily explained. How are we to choose between this account and one that allows the animal to reason that the signal results in the occurrence of the outcome? Again, we may turn to the behavior shown by (relatively) simple systems to provide an answer. Perhaps the most thoroughly worked-out S-S theory of conditioning is that developed by Wagner (e.g., 1981), and he has applied it in detail to the case of eyeblink conditioning in the rabbit. The neurophysiology of this phenomenon has been investigated extensively, and Wagner's theoretical mechanisms map on very well to the systems identified in the cerebellum and brain stem nuclei (Wagner & Donegan 1989). It seems that S-S learning need not involve higher brain structures (decorticate rabbits maintain conditioned responding, Mauk & Thompson 1987). The argument is again one of plausibility, but it surely seems more reasonable to endow the cerebellum with the ability to form S-S associations than with the ability to reason about the relationship between events.

If we accept the foregoing arguments (and thus the reality of S-R and S-S excitatory links), then we must reject the central proposal of the target article: that propositional reasoning is the sole source of the associative learning phenomenon. What remains is the far less radical proposal that excitatory link mechanisms play little or no part in generating associative learning effects in human subjects. This is a question that has been worked over repeatedly and has not been resolved. Regrettably, it seems to come down to a matter of personal preference – are we more impressed by the (to me, still surprising) finding that the development of opinions about the allergenic properties of foodstuffs often seems to follow associative principles of the sort embodied in the Rescorla-Wagner model (Rescorla & Wagner 1972) or by the fact that this form of learning shows properties (e.g., the role of awareness; sensitivity to verbal instructions) that lie outside the scope of models of this sort?

Whatever their answers to this question, all are likely to accept the argument of the target article – that it would be a good thing if we could come up with a single theoretical analysis capable of accommodating all the data. Mitchell et al. argue for the propositional analysis. But if the arguments presented above are accepted, we must acknowledge the role of excitatory (S-R and S-S) links in some instances of associative learning. And having done so, parsimony seems to dictate that the next step should be to attempt to extend this sort of account to deal with those features that seem to call for a propositional theory. Success in this enterprise would put paid to the dual-system approach, although not in the way envisaged by Mitchell et al.

A causal framework for integrating learning and reasoning

doi:10.1017/S0140525X09000995

David A. Lagnado

Department of Cognitive, Perceptual, and Brain Sciences, University College London, London WC1E 6BT, United Kingdom.

d.lagnado@ucl.ac.uk

http://www.psychol.ucl.ac.uk/people/profiles/lagnado_david.htm

Abstract: Can the phenomena of associative learning be replaced wholesale by a propositional reasoning system? Mitchell et al. make a strong case against an automatic, unconscious, and encapsulated associative system. However, their propositional account fails to distinguish inferences based on actions from those based on observation. Causal Bayes networks remedy this shortcoming, and also provide an overarching framework for both learning and reasoning. On this account, causal representations are primary, but associative learning processes are not excluded a priori.

The task of providing a unified framework for learning is fraught with difficulties. It must cover a wide diversity of empirical findings, mesh with theories of memory, attention, and reasoning, and be plausible from both a neural and evolutionary perspective. And all this should be achieved with a minimum of postulates and parameters. It is little wonder that numerous contenders have fallen by the wayside.

Mitchell et al. launch a bold challenge to associative theories of learning. They argue that the phenomena of associative learning can be explained in terms of a propositional reasoning system, and that there is scant evidence or need for a separate link-formation system. Their thesis has many positives. One is the attempt to integrate both learning and reasoning in a unified system. This is a good thing – for too long these have been studied in relative isolation from one another, separated by different concepts, paradigms, and terminologies. This division ignores the rich interplay between learning and reasoning, and the possibility that a common framework subserves both. Another positive is the rejection of associative link-formation as automatic, unconscious, and encapsulated from higher-level cognition. There is extensive evidence against this view (De Houwer 2009; Lovibond & Shanks 2002), and it unnecessarily cuts associative theories off from other reasoning processes.

Despite these positives, there are several problems with Mitchell et al.'s account, in particular their desire to replace associative theories wholesale with propositional reasoning. First, Mitchell et al. give few details about this propositional reasoning system, but the details matter a great deal. For example, none of the current models of human reasoning, whether mental models, logic, or probability-based theories, can handle causal inference (Glymour 2007; Sloman & Lagnado 2005). This is because the current models lack the formal machinery to distinguish inferences based on actions from those based on observation. This is crucial if a representational system is to provide a guide for predicting the effects of potential actions.

Causal Bayes networks (CBN) formalize the distinction between intervention and observation (Pearl 2000; Spirtes et al. 1993), and provide an overarching normative framework for both reasoning and learning. A directed link from X to Y represents a causal relation, such that potential manipulations of X can lead to changes in Y. This contrasts with associative, probabilistic, or logical connections between X and Y, which cannot capture the causal direction.

Formalizing the distinction is also critical to causal learning. Associative or probabilistic information by itself is insufficient to distinguish between causal models (e.g., an association between bell and food can be generated by various causal structures, including a model where the experimenter is the common cause of both). Interventions allow the learner to discriminate between covariationally equivalent models and to identify a unique causal structure (e.g., if interventions on the bell do not produce food [nor vice versa], but interventions on the experimenter produce both bell ringing and food, then the experimenter is the common cause of both).

A recent wave of psychological research suggests that people conform to the basic prescripts of CBN (Gopnik et al. 2004; Lagnado & Sloman 2004; Sloman & Lagnado 2005; Steyvers et al. 2003, Waldmann & Hagmayer 2005), and work is ongoing to identify the psychological processes that underpin this behavior. Although the accent is on causal representation, the involvement of associative mechanisms is not thereby excluded (e.g., they might be used to parameterize strengths of hypothesized causal links; Griffiths & Tenenbaum 2005). Moreover, sometimes associative connections are the most that can be established, and suffice as crude guides to prediction (e.g., when interventions are impractical, or at the early stages of inquiry).

Nevertheless, in such contexts associative mechanisms will be unlike the traditional conception that Mitchell et al. rightly criticize. Thus, contingency information is not processed automatically, irrespective of prior beliefs, instructions, or other information. Rather, various sources of evidence are integrated to infer causal structure, including covariation, interventions, temporal order, and prior knowledge (Lagnado et al. 2007). Contingency information is not privileged here; in fact, the interpretation of contingency data will be modulated by other information such as temporal order (Burns & McCormack, under review; Lagnado & Sloman 2004; 2006).

A further problem is that Mitchell et al., along with many learning theorists, assume that propositional and link-formation systems offer competing accounts of human learning. However, these systems need not be incompatible, as each system has distinct representational aims. Thus, a propositional causal model aims to represent how things relate in the external world, whereas an associative link models the reasoning process itself (Pearl & Russell 2001). For instance, the bell → food link (see Fig. 1 in the target article) represents the inference from bell to food, but not how these variables relate in the world (a plausible causal model is: bell ← experimenter → food). These two approaches are not exclusive; it is conceivable that people have causal representations of the world but use associative-like processes for prediction and parameter learning. Mitchell et al. risk setting up a false dichotomy – either propositions or links – without acknowledging that these concepts serve different representational aims.

Another concern is Mitchell et al.'s argument from parsimony. They maintain that a dual system with two components can never be simpler than a single system made up from just one of these components. But this moves too fast, and depends heavily on how simplicity is quantified. Extending a propositional reasoning system to accommodate all learning phenomena might introduce additional complexity, such that a dual system turns out simpler overall. Despite this lacuna, the evidence that Mitchell et al. cite against a dual-system approach is strong. Rather than reject associative mechanisms tout court, however, a third way remains open. Why not endorse a unified framework that takes the interaction between learning and reasoning seriously, but allows for variation in the complexity of representations and inferential processes? For example, modes of representation might range from causal models to associative networks, and computation might range from fully Bayesian to heuristic methods. These variations will be determined by task demands, as well as environmental and cognitive constraints (e.g., information availability; memory, and processing limitations).

The key point is that a unified framework does not require that the same representations and computations are used for every learning problem; multiple processes are available, and are selected or integrated as required. In short, the flexibility of our cognition system is likely to permit various representational and inferential solutions, including both propositional and associative processes.

Trace conditioning, awareness, and the propositional nature of associative learning

doi:10.1017/S0140525X09001009

Nanxin Li

Department of Psychology, Yale University, New Haven, CT 06520.
nanxin.li@yale.edu **http://pantheon.yale.edu/~nl238**

Abstract: The propositional nature of human associative learning is strongly supported by studies of trace eyeblink and fear conditioning, in which awareness of the contingency of a conditioned stimulus upon an unconditioned stimulus is a prerequisite for successful learning. Studies of animal lesion and human imaging suggest that the hippocampus is critical for establishing functional connections between awareness and trace conditioning.

In the target article, Mitchell et al. argue that human associative learning requires participants to be consciously aware of contingencies between a conditioned stimulus (CS) and an unconditioned stimulus (US). This argument is strongly supported by extensive research on trace conditioning. *Trace conditioning* is a type of classical conditioning in which there is a temporal gap between the offset of the CS and the onset of the US. Trace conditioning is a well-established model of declarative learning that can be tested in both humans and nonhuman animals (Woodruff-Pak & Disterhoft 2008).

Awareness of the CS-US contingency has been suggested to be a prerequisite for successful learning of trace conditioning. This idea is supported by numerous studies on differential trace eyeblink conditioning, in which one of the stimuli (CS+) was always followed by an air puff to the eye (US), whereas another stimulus (CS−) was explicitly unpaired with the US. Clark and Squire (1998) found that only the subjects who were aware of the temporal relationships between the stimuli displayed differential responses to the CS+ and CS−. Their findings were replicated in a later study reporting that acquisition of trace eyeblink conditioning was significantly correlated with the awareness of stimulus contingencies (Knuttinen et al. 2001). Clark and Squire (1999) further showed that preventing the awareness of the contingency during conditioning disrupted differential trace conditioning, while providing knowledge about the relationship facilitated learning. Notably, the acquisition of the CS-US contingency and trace conditioning developed in a roughly parallel pattern (Manns et al. 2000).

The necessity of awareness in trace conditioning has also been evident in fear conditioning studies (Carter et al. 2003; Knight et al. 2004; Weike et al. 2007). Using a differential trace fear conditioning paradigm, Weike et al. (2007) reported that the differential fear-potentiated startle and skin conductance responses (SCR) to the CS+ and CS− were observed during conditioning only when subjects were aware of the contingency between the CS+ and the US. In the study by Carter et al. (2003), significant correlations between awareness and differential SCR responses to a CS+ and a CS− were present during extinction. Altogether, these findings indicate that awareness is required for acquisition and expression of fear in trace conditioning.

The importance of awareness in trace conditioning is reflected in its underlying neural mechanisms. Growing evidence suggests that trace conditioning is dependent on the hippocampus, a medial temporal lobe region which is widely believed to be critical for declarative memory (Clark et al. 2002; Shors 2004; Woodruff-Pak & Disterhoft 2008). Previous studies reported that amnesic patients who suffered from hippocampal atrophy failed to learn the CS-US contingency and, therefore, were impaired in trace eyeblink conditioning (Clark & Squire 1998; McGlinchey-Berroth et al. 1997). The performance of the amnesic patients became worse as the trace interval between the CS and the US increased. Lesion studies in animals also show that aspirative or electrolytic lesions to the hippocampus disrupted trace eyeblink conditioning (Beylin et al. 2001; Weiss et al. 1999; Solomon et al. 1986). Notably, a recent single-unit study has shown that hippocampal CA1 neurons display highly accurate timed firing to the trace period. During the CS-alone testing session, CA1 neurons maximally fired in synchronization with trace interval used in the conditioning trials (McEchron et al. 2003). This timed firing was closely associated with behavioral responses to the CS. The findings suggest that hippocampal CA1 neurons are critically involved in encoding the trace interval, which is essential for the CS-US associations in trace conditioning.

With the advance of human brain imaging techniques, recent studies began to look directly into functional connections among awareness, hippocampus, and trace conditioning. A few fMRI studies have reported strong activations of the hippocampus during trace learning (Buchel et al. 1999; Cheng et al. 2008; Knight et al. 2004). The magnitude of hippocampal activation was closely associated with the accuracy of US prediction, which was a direct measurement of CS-US contingency awareness (Knight et al. 2004). Besides the hippocampus, the activations of other brain regions, including middle frontal gyrus, that support attention and working memory were also associated with trace interval. One idea is that the hippocampus contributes to awareness by interacting with the neocortex (McIntosh et al. 2003).

To sum up, trace conditioning requires an active process to form an internal representation of the contingency between relevant stimuli across a temporal gap. In this respect, the requirement of awareness for trace fear learning strongly supports the single-process propositional view of associative learning.

ACKNOWLEDGMENT
I thank my colleague Sun Jung Bang for reading the manuscript.

Is there room for simple links in a propositional mind?

doi:10.1017/S0140525X09001010

Evan J. Livesey and Justin A. Harris

School of Psychology, University of Sydney, Sydney, NSW, 2006, Australia.
evanl@psych.usyd.edu.au
http://www.psych.usyd.edu.au/staff/evanl
justinh@psych.usyd.edu.au
http://www.psych.usyd.edu.au/staff/justinh

Abstract: Against Mitchell et al.'s assertions, we argue that (1) the concordance between learning and awareness does not support any particular learning theory, (2) their propositional approach is at odds with examples of learned behaviours that contradict beliefs about causation, and (3) the relative virtues of the two approaches in terms of parsimony is more ambiguous than Mitchell et al. suggest.

Mitchell et al. state that a demonstration of conditioning without awareness would be a major problem for their propositional approach. They claim the literature is consistent with this approach because convincing evidence for learning without awareness is scarce, and because there is a "clear concordance between conditioning and awareness" (sect. 4.1, para. 6). Lovibond and Shanks (2002) established clear principles for the appropriate assessment of awareness, and found most demonstrations of unconscious learning used awareness tests that were insufficiently sensitive or rigorous. Over time, their evaluation of the literature has fuelled the conclusion that learning *cannot* occur without awareness. It is worth examining that conclusion again. Aside from the observation that the dubious nature of the evidence for unconscious learning is not evidence of absence, one can make a strong case that the concordance between learning and awareness does not inform the current debate at all. This is because a realistic dual-process account should still predict that learning will be harder to observe and less likely to occur in unaware participants.

First, it is wrong to assume that link formation should be observable in all situations and in all participants. Dual-process theories usually assume that conscious reasoning can have a large impact on behaviour, large enough in some circumstances to obscure other more subtle behavioural influences. This is even true of participants who reason incorrectly or erratically during an experiment. We should not be surprised if participants who were aware of an associative relationship showed behavioural evidence of learning. But those who were unaware may fail to show learning because they are deliberating on a spurious hypothesis or concentrating on something unrelated to the task. This may disrupt performance, regardless of whether learning has taken place.

It is also wrong to assume that, just because a learning process is automatic, it will always occur (and always to the same degree). If link formation is affected by selective attention or memory load, then the evidence showing a close correspondence between awareness and learning is highly predictable by any account. To suppose that link formation is affected by cognitive processes is not contrary to its conception as an automatic process. Changing the input to a link-based network inevitably leads to changes in what is (or what is not) learned by that system. A model of automatic link formation would be impervious to manipulations of selective attention and working memory if, and only if, its input was restricted to the earliest levels of sensory processing.

Mitchell et al. argue that if link formation is dependent on cognitive resources, then "one of the reasons for postulating the existence of an automatic link-formation mechanism has been removed" (sect. 4.2, para. 3). But this argument only holds if capacity limitations affect link formation and propositional learning in the same way. Instead, they might affect what the cognitive system *does* in the sense of the inferences that are drawn about the relevant events, whereas those same limitations might affect what the automatic system *receives* in the way of input. With this in mind, it would not be at all surprising if consistently attentive participants were more likely to consciously identify the relevant contingencies and also more likely to learn, whereas participants who attended erratically were less likely to be aware and less likely to learn.

Although Mitchell et al. provide examples where learning is sensitive to rules and instructions, these examples only confirm what *both* approaches already assume – that conscious inferences and beliefs can influence decisions and behaviour. More importantly, some learned behaviours do not show this sensitivity. While Mitchell et al. discuss Perruchet's (1985) dissociation, other prominent examples include conditioned taste aversions and anticipatory nausea and vomiting elicited by cues associated with chemotherapy (Bernstein & Webster 1980; Carey & Burish 1988). Both appear to be very clear cases of uncontrollable, automatically learned responses that have nothing to do with the beliefs of the sufferer.

Mitchell et al. wish to extend their approach to describe learning in other animals. But the utility of the approach is questionable given the obvious obstacles in establishing whether other animals even have awareness or beliefs, let alone in measuring either. And if we assume they do, what can we deduce from an animal's behaviour about the content of its beliefs? As one example, when a pigeon pecks or drinks a key-light during an autoshaping procedure (e.g., Jenkins & Moore 1973), does it believe that the key-light *is* food or water? What reasoning process could give rise to such a proposition? Or should we conclude that the pigeon's behaviour does not reflect the content of its belief? If the latter, then we must concede that the learning process cannot be adequately investigated by examining animal behaviour.

At a superficial level, positing a single mechanism for learning seems more parsimonious than assuming two mechanisms, and Mitchell et al. encourage us to abandon link-based learning on this basis. However, conscious propositional reasoning is computationally more expensive and much less well-specified than the operations underlying the strengthening of associative links. The propositional learner is one who retrieves information from episodes and consciously rationalises which events signal other events within the same episode. This comes with its own set of assumptions that imply considerable computation. For a start, in order to calculate relationships symbolically, the propositional system needs to decide what each event *is*. It also needs to store events in an episodic fashion that includes temporal information. This contrasts with the simple mechanisms by which associative links are strengthened or weakened according to statistical regularities in the environment.

The comparison begs the question: Is it really parsimonious to conclude that all instances of learning are the consequence of an elaborate and cumbersome set of cognitive operations just because we know these operations affect human behaviour in certain circumstances? We conclude that it is more parsimonious to assume learning is the product of a very simple link mechanism, but to evaluate exceptions to this rule where the evidence necessitates. Of course, the issue of parsimony is at the very heart of the single-versus-multiple process debate.

Saliences, propositions, and amalgams: Emergent learning in nonhumans

doi:10.1017/S0140525X09001022

Heidi Lyn[a] and Duane M. Rumbaugh[b,c]

[a]*Department of Psychology, Agnes Scott College, Decatur, GA 30030;* [b]*Great Ape Trust of Iowa, Des Moines, IA 50320; and* [c]*Language Research Center, Georgia State University, Decatur, GA 30034.*

heidilynphd@aol.com

http://userwww.service.emory.edu/~whopkin/Heidipage.html

drumbaugh@greatapetrust.org

http://www.greatapetrust.org/research/drumbaugh/drumbaugh.php

Abstract: We comment on the similarities and differences of Mitchell et al.'s framework for understanding classical and operant conditioning and the theoretical framework put forth by Rumbaugh et al. (2007). We propose that all nonhuman and human learning may be based on *amalgams* created by co-occurring stimuli that share their response-eliciting properties and that these amalgams may be propositional in nature.

Rumbaugh et al. (2007) have argued that stimulus-response learning and reinforcement as constructs, as put forth in classic behaviorism over the course of the past century, are in need of redefinition. The essence of the redefinition entails contiguous stimuli sharing their response-eliciting properties as a function of their relative attributes and strengths (e.g., their saliences), forming *amalgams*. These amalgams are not simply linkages between stimuli; they are new entities based on the properties of the stimuli, their response-eliciting properties, and any other salient features of the environment. The ease with which amalgams form, is a function of the natural history of the species and the constructive bias of its neural system.

The neural system integrates the amalgams into templates that metaphorically define a knowledge base, which includes information about the individual and its assessed capabilities in relation to the resources and risks of its ecological niche. From this knowledge base, emergent behaviors and capacities, with no history of specific training, might take form as overarching principles and rules to service adaptation rationally and creatively in both familiar and novel challenges. This theoretical stance has much in common with that of Mitchell et al. Neither theory supports the idea of reinforcement as a direct line to the CR (conditioned response); nor does either theory support the notion of simple S-S (stimulus-stimulus) links as part of a dual-learning system.

This salience theory is silent on the issue of consciousness; however, its provision for emergent behaviors and capabilities allows for consciousness to emerge as a product of the complexity of the nervous system and its knowledge base. Consciousness, once functioning, might selectively interact with the systems of templates that are always subject to modification across time and experience. Thus, consciousness and symbol-based logic (and propositional knowledge) might either monitor or modify systems of templates and the amalgams they organize. Mitchell et al. make a good case against unconscious link-based learning systems, but they do admit the existence of some evidence for a learning system in which subjects react contrarily to their stated beliefs. Mitchell et al. correctly suggest these findings could be the result of an "imperfect" propositional system, where the subjects have created beliefs about certain contingencies, but those contingencies are flawed in some way. However, there is no reason to assume that all learning, even if propositional in nature, must be fully conscious and accessible.

To illustrate the viability of the salience theory in these situations, we turn to some of Mitchell et al.'s examples. Mitchell et al. discuss the concept of blocking (sect. 4.2), where a pre-trained CS ($A+$) blocks the association of a new CS ($T+$) and the US ($AT+$). This is explainable under the salience theory, since the pre-trained CS (A) would be highly salient and would overshadow any new competitive amalgam formation unless the new CS (T) were equally or more salient. Additionally, the finding that embedded tasks (cognitively demanding tasks presented simultaneously with the presentation of the $AT+$ condition) result in subjects' better learning the T condition (sect. 4.2), is supported by salience theory. In this case, the embedded/distracter task reduces salience of A during the presentation of $AT+$, increasing the salience of T and resulting in some learning of the $T+$ contingency. Finally, in the case when $AT+$ was trained, how did the presentation of $A-$ serve to increase the response to T? Perhaps A and T had become equally salient constituents within new amalgams, but subsequent $A-$ training reduced the saliency of A, in effect, increasing the relative saliency of T (sect. 4.3).

Both Mitchell et al.'s and Rumbaugh's theories are equally supported by findings from a long history of nonhuman cognitive studies, some of which were mentioned in the target article. One of the most reliable indications of learning abilities in nonhuman primates not mentioned in the target article is the transfer index (Rumbaugh & Washburn 2003). This index shows the ability of individuals trained in a simple one-choice test (choose between A and B) to switch a preference from a trained positive stimulus ($A+$) to a new positive stimulus ($B+$), given a certain level of ability on the first trained stimulus (67% or 84%). In other words, when your performance on the first task (always choose A) is increased from 67% to 84%, how is your performance on the second task (switch to always chose B) affected? These studies show that smaller-brained primates are adversely affected by augmented learning; that is, the more they learn in the first task, the worse they do on the second. In larger-brained primates, this finding is exactly reversed (the correlation between transfer index performance and various measures of brain size range from .79 to .82). Additionally, transfer index performance is strongly affected by early environmental conditions (see Rumbaugh & Washburn 2003, for a review). These results are difficult to explain with a straight link-based learning system and suggest that learning processes are altered by elaboration of the brain and by rearing.

Further evidence for a propositional/salience-based learning system comes from language learning in chimpanzees and bonobos. These studies include findings in rapid symbolic associations (association of symbols to new referents on a one-trial basis; Lyn & Savage-Rumbaugh 2000) and "representational play" – treating a toy or toy stand-in as if it were something else (Lyn et al. 2006). Moreover, complex mental representations of symbols have been documented by detailing the errors in a vocabulary test (Lyn 2007). These errors include choosing a visual symbol (e.g., a lexigram that represents "key") based on a photograph of a referent that has an auditory similarity ("TV") – indicating that all levels of representation are activated in the choice task. Additionally, emergent behavior (such as the initial language learning in a bonobo; Savage-Rumbaugh et al. 1986) frequently appears during cognitive training in nonhumans, indicating that nonhumans are able to respond not just based on their learned contingencies, but rather by construction of an emergent, unlearned (self-generated) contingency, a feat difficult to explain through standard S-R (stimulus-response) link-based learning.

ACKNOWLEDGMENTS

The first author, Heidi Lyn, wishes to acknowledge support from NIH grant number HD 56232. Duane M. Rumbaugh's work on this commentary was supported by NICHD 38051 to Georgia State University.

Propositional encodings are a subset of organization theory

doi:10.1017/S0140525X09001034

George Mandler

Department of Psychology, University of California, San Diego, La Jolla, California, 92093-0109.

gmandler@ucsd.edu

http://psy.ucsd.edu/~gmandler/

Abstract: The notion that human associative learning is a usually conscious, higher-order process is one of the tenets of organization theory, developed over the past century. Propositional/sequential encoding is one of the possible types of organizational structures, but learning may also involve other structures.

The main argument of the target article is to show that associative learning is not – as reputedly generally assumed – an "automatic process that is divorced from higher-order cognition." (sect. 1, para. 1) Instead, it is proposed that human associative learning is based mainly on the acquisition of propositions involving the main two terms of the to-be-learned conjunction, from conditioning or quasi-conditioning paradigms of learning, primarily those that have been frequently characterized as formed by an associative "link." Assuming that "learning" refers to the acquisition of new knowledge, in this commentary I show (necessarily briefly) that the notion that human associative learning is neither automatic, nor necessarily unconscious, has a venerable, nearly century-old history, missing from the target article. Furthermore, propositional structures constitute just one part of organization theories (see Mandler 2007, for a more extended history).

The opposition to unconscious, automatic associative processes started in modern times with the work of G. E. Müller (e.g., Müller 1911) and proceeded rapidly with the development of Gestalt theory. Following the work of Wertheimer (1921), Duncker (1926), and Katona (1940), the time was ripe for a full-scale assault on the mechanisms of associative memory. The initial arguments were primarily presented by Asch (1962; 1969) and Asch and Ebenholtz (1962), and generated specific demonstrations of human associative learning by Bower (1970), Bower and Bryant (1991), Mandler and Mandler (1964), Mandler (1968; 1979a; 1979b), and Murdock (1966). In Mandler (1979b), I suggested three possible structures accounting for human associative phenomena: coordination (holistic, unitary organizations), subordination (hierarchical organizations), and pro-ordination (sequential organization). The last is most like the propositional structure proposed in the target article – *A* followed by *B*.

Relevant to the target argument, my colleagues and I have tested human associative learning and demonstrated that holistic structures characterize the storage of verbal associations. Mandler et al. (1981) showed that in verbal human associative learning (sometimes known as paired-associates), "associations" are stored not as "links," but by combining the two terms in a single holistic unit. Tests of free recall, cued-recall, and recognition supported that conclusion.

Propositions about and tests of organizational theory describe the structure of human semantics – the mental organization of meaningful knowledge and experience. Organization defines the structure of memory. It is obvious that propositional structures depend on retrievals from memory, and, albeit without any detailed discussion of memory, Mitchell et al. too assert the centrality of memorial functions, when in section 3.1 (para. 1) they state that the encoding of an associative hypothesis in memory constitutes learning. Organization theory has generally avoided any distinction between learning and memory. The history of the organizational approach discussed the organization of mental contents, which can be seen as "learned" when established and retrieved once the organizational structure is established. Consistent with such an approach, Mitchell et al. also note that subsequent to a bell-food pairing, a bell can retrieve memories of previous pairings. More generally, it may not be initially obvious which of the possible structures applies to a particular learning experiment or paradigm. At present it is not obvious which experimental or experiential situations give rise to one organization or another. The target article seems to claim that all encodings are propositional; in contrast, we have shown that some are holistic and unitary. Specific experimental procedures and probing and testing procedures need to be developed in order to determine which particular structures eventuate from a specific "learning" situation.

Finally, it does not seem obvious that "we have been heading . . . towards a propositional approach to all learning" (sect. 7.1, para. 3). The holistic encoding of word pairs or the hierarchical organization of some lists argues against a single model of underlying structures. A general organizational approach has asserted for some time that learning is indeed "not separate from other cognitive processes" (sect. 8, para. 1). Organization theory has made it possible to see the connectedness of these various functions and processes.

The Proust effect and the evolution of a dual learning system

doi:10.1017/S0140525X09001046

Helena Matute and Miguel A. Vadillo
Department of Psychology, Deusto University, 48080 Bilbao, Spain.
matute@fice.deusto.es
http://paginaspersonales.deusto.es/matute
mvadillo@fice.deusto.es
http://paginaspersonales.deusto.es/mvadillo

Abstract: Proust's madeleine illustrates the automatic nature of associative learning. Although we agree with Mitchell et al. that no compelling scientific proof for this effect has yet been reported in humans, evolutionary constraints suggest that it should not be discarded: There is no reason by which natural selection should favor individuals who lose a fast and automatic survival tool.

> And soon, mechanically, weary after a dull day with the prospect of a depressing morrow, I raised to my lips a spoonful of the tea in which I had soaked a morsel of the cake. No sooner had the warm liquid, and the crumbs with it, touched my palate than a shudder ran through my whole body, and I stopped, intent upon the extraordinary changes that were taking place . . . at once the vicissitudes of life had become indifferent to me, its disasters innocuous, its brevity illusory.
> — Marcel Proust (1913/1922), *Remembrance of Things Past*

The episode of the madeleine in the Proust work cited above is one of the most famous passages of universal literature of all times. Not only is it beautifully written, but the passage also describes an experience that is so personal and so ubiquitous in human nature that any psychologist, from Freud to Pavlov, would love to explain it. We will refer to it as the "Proust effect." To our knowledge, it is the best possible description of associative learning.

The beauty of the target article by Mitchell et al. is that it tries to understand the Proust effect in its entirety, not just a part of it. As such, the article is ambitious, important, and timely. It makes us rethink all the established assumptions about learning. Contrary to all intuitions, Mitchell et al. (almost) convince us that (a) there must be only one learning process, and (b) this unique process must be propositional in nature.

The standard explanation for associative learning is the link approach. Because the narrator in Proust's novel had associated the madeleines with all the happiness of childhood (even though he was not aware of this fact), then tasting one of those cakes now, after so many years, brought back the enormous happiness and all the good feelings from childhood. Thus, the Proust effect reflects a simple, automatic link that was created during childhood and is now expressed, also without effort or knowledge of the contingencies, in the form of a conditioned response (CR). According to the link proponents, there was no propositional learning here, no consciousness of the contingencies while the association was acquired; not even now that it is expressed. Indeed, it will still take the narrator many pages and a considerable amount of thinking and elaborated reasoning to discover why the madeleine was producing the CR.

But the link approach is not as simple as it seems, and Mitchell et al. are correct in highlighting this point: The link approach presupposes a dual (and complex) system. Automatic links need to be complemented with some more-elaborated, rational, and time-consuming forms of learning. This complex learning is at work, for instance, after the CR has occurred and the narrator begins to consciously think about it and tries to identify its cause. Even the most enthusiastic proponents of low-level mechanisms have to admit that people are obviously capable of other forms of learning and reasoning.

What Mitchell et al. suggest is that, if we all agree that propositional learning is needed, why should we maintain a belief in automatic links? Couldn't we assume just a propositional learning process that could account for both the automatic-like and the more complex processes? Are there any experiments that can only be explained by the link mechanism? That there are data to support that propositional learning exists is unquestionable, and the authors make an excellent case of it. That many of the results that have traditionally been explained using the link approach can also be explained by the propositional account is also clear in their target article. Moreover, it is well established today that there are very few experiments that can be explained solely by the link approach (Lovibond & Shanks 2002; Shanks & St. John 1994). What Mitchell et al. are showing is that both the dual and the propositional account can explain the majority of the available evidence. Scientific parsimony becomes then the central argument: If a single process can explain it all, why should science maintain two?

But the argument of scientific parsimony should be confronted against that of natural selection. A simple, low-level process is vital for survival because, by definition, it can do all those things the complex process cannot do: it responds quickly, automatically, and without consciousness or effort to the demands of the environment. Even under high pressure, it provides a fast tool for survival. Its loss would be too costly.

As Mitchell et al. note, natural selection has produced a continuum of complexity in the different species. At one end of this continuum, we find very simple species which have just the link system and no cognition. At the other end, we find the human species, which, according to Mitchell et al., has only the propositional system. If so, Mitchell et al. need to explain why humans (and other evolved animals) should have lost their primitive link system while developing the propositional one. There is no clear evolutionary advantage in losing a fast and automatic tool.

Indeed, there is a growing body of evidence suggesting that learning is actually caused by a multiplicity of different mechanisms and that the insistence of traditional learning theory in a unique, general-purpose learning system was simply a mistake (Gallistel 2000; Tooby & Cosmides 1992; 2005). If natural selection has encouraged flexibility and adaptability, having many different forms of learning must have been favored through the course of evolution.

In sum, Mitchell et al. need to explain not only why consciousness becomes so difficult in the Proust effect, but also what survival advantages a species that extinguishes the link system should have. If all the evidence for the automatic mechanism would come from novels and intuitions, Mitchell et al. would be right that science should ignore it. But we have shown good reasons to believe that the automatic mechanism must still be present in humans. Perhaps the problem is that the Proust effect has always been taken for granted and proofs have not been searched in the right places.

ACKNOWLEDGMENT

The authors of this commentary were supported by MICINN (Spain) Grant SEJ2007-63691.

Both rules and associations are required to predict human behaviour

doi:10.1017/S0140525X09001058

I. P. L. McLaren

School of Psychology, University of Exeter, Exeter, EX4 4QG, United Kingdom.
i.p.l.mclaren@exeter.ac.uk

Abstract: I argue that the dual-process account of human learning rejected by Mitchell et al. in the target article is informative and predictive with respect to human behaviour in a way that the authors' purely propositional account is not. Experiments that reveal different patterns of results under conditions that favour either associative or rule-based performance are the way forward.

In this target article, Mitchell et al. argue for a propositional account of human learning, rather than a dual-process model that allows for propositional and associative (what they call the "link model") processes to operate concurrently. The issue at hand, then, is whether we need to postulate associative processes in addition to propositional ones; the converse argument, whether we need to postulate propositional processes in addition to associative processes, can be left for another time. But let me be quite clear: I am of the view that we need to appeal to both if we are to understand learning in humans.

The approach taken in this commentary is to point out differences in learning and performance under conditions that should favour either propositional or associative learning. Mitchell et al. consider a number of these cases, but perhaps do not do them justice. I take as my first example their review of the Le Pelley et al. (2005a) demonstration of unblocking in humans. In these experiments, it was demonstrated that a design such as $A- > O_1$ followed by $O_2|AB-> O_1$ followed by O_3 revealed that learning to B was greater for outcome 1 (O_1) than in a conventional blocking design where the second phase had the compound followed by O_1 then O_2. This finding was predicted on the basis of Mackintosh's (1975) associative theory of learning, which has received experimental support in animals other than human. To dismiss it by saying that it is possible that in a complex design the human participants had forgotten earlier trials and knew something had changed but were not sure whether it was O_1 or O_2, ignores this background. As an explanation of the phenomenon, it is terrifically weak. We are expected to allow that propositional learning and an automatic memory (that is definitely not associative?) are both imperfect, and so people make mistakes, which just happen to be the ones that associative theories predict.

This does sound rather implausible, and it is, even though the authors reassure us that it can be tested. Their proposal is to make the outcomes more distinctive, thus reducing any confusion between them, and so the effect (unblocking) should go away. In fact, if the outcomes were made that distinct from one another, the same associative theory that predicted the original result would now predict that the effect would go away as well, as the alpha change that leads to unblocking is to some extent reinforcer-specific in this model. This result has also been found in humans in another experiment by Le Pelley and colleagues (Le Pelley et al. 2005b), in which changing outcomes from those that are generally "nice" to those that are generally "nasty" (and vice versa) prevented alpha effects that were generated by manipulating the predictiveness of certain cues during training. So we are left with a "test" of their account that fails to distinguish between it and the very associative theory that motivated the experiment in the first place. Not much of a test!

Mitchell et al. also fail to take into consideration a number of other studies that demonstrate a different pattern of results when learning is dominated by either rule-based (hence propositional) or associative processes. People show a peak shift, like pigeons, when they are tested on a dimension after relatively little experience with it, and when they are unable to verbalise any rule that captures the discrimination (Jones & McLaren 1999; and see Livesey & McLaren, forthcoming). This pattern of responding changes (to a monotonic function across the dimension) after extensive experience with the stimuli and when people can verbalise the correct rule. In the spirit of the target article, I would expect the response to be that this does not demonstrate associative learning, but instead, incorrect rule induction or imperfect application of a rule in some way. If this characterisation of Mitchell et al.'s position is right, then it is impossible to defend against. There will always, with sufficient ingenuity, be some incorrect or imperfect rule that can be appealed to that fits the behaviour. But it simply ignores the fact that these studies were based on predictions made before the fact by a dual-process model, not after the data had been collected. Any sufficiently complete learning system can explain any pattern of results once the pattern is known; the trick is to predict them in advance. The "imperfect rule" approach will never do that.

My final example, which reinforces the point just made, concerns work on sequence learning done by myself and Rainer Spiegel (Spiegel & McLaren 2006). In this series of experiments, we show that the predictions made by the Simple Recurrent Network (SRN), even though at times quite counter-intuitive at first sight, are nevertheless borne out by the experimental results obtained with humans. For example, training on sequences of the form *ABC* ... *CBA* and *ABBC* ... *CBBA* where *C* ... *C* can be one, three, or five *C* terms, leads to the ability to respond faster and more accurately to the term after the first *B* following the *C* terms, as it is predicted by the rule "The number of *B*s after the *C*s is the same as the number experienced before the *C*s." However, this rule would not predict that when tested after acquisition involving sequences with an odd number of C elements on a sequence, such as *ABCCCCBA*, the result would be that no learning was displayed, in that responding to the *A* after the *C*s was not facilitated. This was predicted by the SRN, and was the case in our experiments. Many other counter-intuitive effects are reported in Spiegel and McLaren (2006) that all closely follow the predictions made by this associative model. I am not sure whether Mitchell et al. will opt for the "wrong or imperfect rule" approach here or simply rule these data out of court on the grounds that the SRN is not a standard model of associative learning, but either way the purely propositional account seems implausible to the point of incredulity when confronted by these data.

The target article is an enjoyable attempt to make the case for an exclusively propositional account of human learning, but I fully expect that the learning theories of tomorrow will make use of both rules and associations, and that the attempt to restrict theorising to one or the other will quickly pass.

Associative learning without reason or belief

doi:10.1017/S0140525X0900106X

James D. Miles, Robert W. Proctor, and E. J. Capaldi

Psychological Sciences, Purdue University, West Lafayette, IN 47907-2081.
jmiles@purdue.edu **http://www.jimmiles.org**
proctor@psych.purdue.edu
http://www2.psych.purdue.edu/~proctor/
julie@psych.purdue.edu

Abstract: We discuss the necessity of conscious thinking in the single-system propositional model of learning. Research from honeybees to humans suggests that associative learning can take place without the need for controlled reasoning or the development of beliefs of relationships between objects or events. We conclude that a single learning system is possible, but not if it depends on complex thinking.

Mitchell et al. contribute their view to a long-standing controversy in psychology: Namely, can the range of human and animal behavior be sufficiently attributed to a single processing system rather than multiple competing systems? The authors propose that a single propositional learning system drives behavior. In this learning system, "controlled reasoning processes are necessary for learning to take place, and learning results in beliefs about the relationship between events" (sect. 1, para. 8). The major aspect of this position with which we disagree is: "are necessary for learning." In this commentary, we point out several phenomena that are difficult to explain within a single propositional learning system as it is described.

Mitchell et al. contend that associative learning in the Pavlovian situation involves propositional knowledge and not the automatic formation of links between events. The authors indicate that, in their view, *Aplysia* do not possess propositional knowledge. This view would presumably extend to insects, such as honeybees. Rats and people show three complex conditioning effects: backward blocking, forward blocking, and relative validity. Each of these effects involve multiple conditioned stimuli (CS) occurring concurrently or in isolation along with an unconditioned stimulus (US). Mitchell et al. argue that the resulting patterns of response behaviors in these blocking designs are difficult to explain within a strictly automatic associative-link framework. Therefore, they suggest that these phenomena involve propositional knowledge in humans and possibly in rats. Because, presumably, bees do not possess propositional knowledge but display all three effects, it follows that these effects can occur in the absence of propositional knowledge (see e.g., Blaser et al. 2004; Guez & Miller 2008).

So, the authors are driven to the conclusion that highly similar and complex effects in different species can occur on the basis of highly different learning mechanisms. However, the phylogenic "cut-off" point between these mechanisms is unclear. An alternative possibility is that the same learning mechanism produces the three effects in honeybees, rats, and people. According to that view, the learning mechanism must be some form of simple links rather than propositional. It is much easier to concede that humans can learn like bees than bees can learn via propositional reasoning like humans. Whether or not the same mechanism is responsible for learning in both cases, it must be assumed that links between events occur in order to accommodate the bee data.

Mitchell et al. mention that, at least in more complex animals, it is possible that their single propositional learning system may rely on the multiple connections between subcortical and cortical brain regions (sect. 5.3). As such, it is difficult to imagine an intact propositional system at work in decorticate animals in which most higher brain functions, such as those associated with reasoning skills, have been removed. However, many studies, the earliest of which include Culler and Mettler's (1934) research, have found that conditioned learning still occurs following the general or localized removal of cortical regions. For example, rats with as much as 99% of their neocortex removed showed learning in a T-maze that was equivalent to that of fully intact controls (Thompson 1959). At the furthest extreme, the spinal cord alone is sufficient for associative learning to occur (Patterson et al. 1973). These findings do not preclude the existence of propositional learning, but they show that it is not necessary for associative learning to take place.

Non-propositional learning also appears to occur within humans. The classic argument for learning without awareness is Claparède's (1907) description of a woman suffering from Korsakoff's syndrome (Nicolas 1996). Claparède pricked his patient's hand with a pin hidden in his own, and, although she did not display any declarative knowledge of the experience, she later would withdraw from Claparède when he gestured toward her with his hand. Subsequent research with anterograde amnesiacs provides further examples of similar learning in the absence of conscious, declarative knowledge. For example, Gabrieli et al. (1995) found that eyeblink conditioning remained intact in amnesiacs and suggested that declarative information was not needed for CS-US associations to occur (for more examples, see Cohen & Squire 1980).

It may be argued that aware, conscious reasoning does occur during the learning process in amnesiacs, and the patients are simply no longer aware of this learning experience during the expression of this information at a later time. Even if this were the case, there are two reasons that a propositional-system only account remains problematic based on evidence from amnesiacs. First, it is unclear how learned propositional information can be used later without conscious awareness that the information must be applied. As the target article authors describe their learning system, learned propositional information is expressed through beliefs about the relationships between events, and these beliefs are very unlikely to occur at all in anterograde amnesiacs. Additionally, there is some evidence that learning processes in intact individuals may occur without the need for any conscious awareness of the associations between events in the environment. In tasks such as artificial grammar learning, implicit categorization, and implicit sequence learning, the learning and expression of associations between items or events in the environment occur without the need for awareness. For example, in four-choice tasks that contain long repeating item sequences, participants show evidence of learning the sequence often without awareness or knowledge and concurrently with or in isolation from explicit learning (Song et al. 2007). This further implies that associative learning can take place without reasoning and conscious beliefs about associations.

To conclude, the authors claim that "associative learning is never automatic and always requires controlled processes" (sect. 3.1, para. 2). However, the aforementioned examples provide converging evidence that associative learning can take place without the need for reason or belief. If a single learning system is to account for all forms of learning, the system must accommodate these cases.

Undermining the foundations: Questioning the basic notions of associationism and mental representation

doi:10.1017/S0140525X09001071

Ezequiel Morsella,[a,b] Travis A. Riddle,[a] and John A. Bargh[c]
[a]*Department of Psychology, San Francisco State University, San Francisco, CA 94132-4168;* [b]*Department of Neurology; University of California, San Francisco, San Francisco, CA 94143;* [c]*Department of Psychology, Yale University, New Haven, CT 06520-8205.*
morsella@sfsu.edu **http://bss.sfsu.edu/emorsella/**
triddle@sfsu.edu
john.bargh@yale.edu **http://pantheon.yale.edu/~jab257/home.html**

Abstract: Perhaps the time has come to re-examine the basic notions of cognitive science. Together with previous challenges against associationism, the target article should be viewed as a call to arms to re-evaluate the empirical basis for contemporary conceptualizations of human learning and the notion of "mental representation," a concept that has become too imprecise for describing the elements of cognition.

Has the time arrived to rebuild the foundations of cognitive science? If Mitchell et al. had been the first to challenge the basic notion of associationism, then perhaps the field could survive this seeming coup d'état by simply pointing to the mounting evidence in favor of associationism – the classic research on the changes at the synaptic level that are responsible for classical conditioning (e.g., Kandel 2000; LeDoux 2000; Thompson 2005); and investigations on the implicit/automatic processes mediating the learning of sequences, motor skills, attitudes (Stadler & Frensch 1998), and the execution of various kinds of "input-output" processes (Hallett 2007; Pessiglione et al. 2008). Afterwards, we could all continue to promulgate that the brain learns as it does because "cells that fire together wire together" – a concept that, like the reflex, exemplifies our switchboard intuition regarding how intelligent behavior is implemented neurally. But Mitchell et al.'s is not the only challenge, and it may not be the most ground shaking.

Gallistel and Gibbon (2001) challenged associationism by demonstrating that it could not account for basic conditioning phenomena (in rats and pigeons) such as *time-scale invariance* or the observation that "neither the delay of reinforcement nor the ratio of reinforced to unreinforced presentation of the conditioned stimulus affects rates of acquisition and extinction" (p. 146). If cells that fire together do wire together, then intermittent conditioning should show less resistance to extinction than *fixed* schedules of conditioning; but this is not the case in classical and operant conditioning (Gallistel & Gibbon 2001; Skinner 1953). Hence, the target article should be seen as a red flag that can no longer be ignored – a wake-up call to begin to carefully take stock of what is actually known regarding the basis of human cognition. Moreover, if the time has come to rebuild the foundations of cognitive science, perhaps it is also worthwhile to clean up our terms and reexamine the value of the notion of "mental representation," an ambiguous term that has come to mean all kinds of things to all kinds of researchers.

Unlike consciousness, which has been regarded as the main "unsolved anomaly within the domain of the [scientific] approach" (Shallice 1972, p. 383), the neural mechanisms proposed to underlie conditioning reflect our intuitive understanding of how the mind/brain *should* work: Fear conditioning is mediated by the amygdala (LeDoux 2000) in a manner that is consonant with our switchboard intuition regarding nervous function. However, despite our intuitions, once Tolman (1948) demonstrated that there is reason to doubt that learning is due to simple stimulus-response (SR) models (e.g., by showing that rats could solve mazes without relying on any external or priopriceptive cues; see also Lashley 1951; Terrace 2005), doubt should have fallen over all SR explanations of behavior. But it did not. Complex behaviors continued to be explained by complex mechanisms (e.g., cognitive maps, reasoning), and simple behaviors continued to be explained by simple mechanisms (e.g., SR strength).

When explaining how a pigeon pecks a button for food, one appeals to the principles of operant conditioning, but when explaining how wasps and pigeons are able to find their home in the absence of external cues, one invokes the term "cognitive map" (Gallistel 1990), which depends on neural machinery that, for some reason, is believed to not be at play during button-pecking. It is seldom appreciated that, if all an animal possesses happens to be a sophisticated faculty of navigation, then this faculty will be used even for button pressing. The hands of evolution have been seen as behaving economically, using simple mechanisms for simple behaviors but reserving complex mechanisms for complex behaviors, which is a wrong way to think about evolution (cf. de Waal 2002). The fallacy has persisted even though, as noted by Mitchell et al., there has always been more evidence for the existence of high-level representational mechanisms than for simple mechanisms (e.g., SR strength).

It remains an empirical question whether, in humans, there are no instances in which associative learning is instantiated by the kinds of synaptic-level changes that have been identified in animals (see Phelps & LeDoux 2005). If so, what *is* the neural foundation of human cognition? Perhaps the time has come to examine whether conditioning is mediated by higher-level processes such as "interregional synchrony" or "neural coherence" (cf., Buzsáki 2006; Fries 2005), processes believed to serve an important role in communication and in "binding" representations (Hummel & Gerloff 2005).

Are *mental representations* those tokens used by the propositional system proposed by Mitchell et al.? If so, Gallistel (2001) defined a mental representation as "a system of symbols isomorphic to some aspect of the environment, used to make behavior-generating decisions that anticipate events and relations in that environment . . . [and, cognitive] psychology is the study of mental representations" (p. 9691). However, the nature of the isomorphism to the world remains unclear with respect to many "representational" processes, such as *non-intentional* states (e.g., moods or the experience of holding one's breath; Gray 2004). More concretely, in light of Gallistel's definition, it remains unclear to what the pungent flavor of hydrogen peroxide is isomorphic. This chemical differs molecularly from water only by the addition of a single oxygen atom, but few would perceive it as "water with a little too much oxygen." Instead, the toxic chemical is perceived (or "represented") as something that should be violently expelled from the body. This may lead one to hypothesize that, as with subjective urges (Morsella 2005) and percepts (Sperry 1964), the representation of H_2O_2 is isomorphic with respect to *how one should respond to the stimulus*, but this is not in line with the traditional view (based on the cognitive map) of what a mental representation is (Hommel et al. 2001). Hence, a more precise term is needed for the tokens that furnish the contents of the propositional reasoning system proposed by Mitchell et al. Cognitive science may be far from developing its periodic table, but it can still be rigorous about delineating what is known and what is not yet known.

What is the link between propositions and memories?

doi:10.1017/S0140525X09001083

Ben R. Newell

School of Psychology, University of New South Wales, Sydney 2052, Australia.
ben.newell@unsw.edu.au
http://www2.psy.unsw.edu.au/Users/BNewell/

Abstract: Mitchell et al. present a lucid and provocative challenge to the claim that links between mental representations are formed automatically. However, the propositional approach they offer requires clearer specification, especially with regard to how propositions and memories interact. A definition of a system would also clarify the debate, as might an alternative technique for assessing task "dissociations."

Propositions, memories, and their interaction. Mitchell et al. use the simple example of learning that a bell signals food to illustrate the differences between the propositional and dual-system approaches. Although useful, the simplicity of the example is potentially deceptive. Focusing on a situation in which the organism learns about a *single* binary cue (bell rings/does not ring) and a *single* binary outcome (food presented/not presented) potentially leaves out some of the devilish details of how organisms learn in *multiple*-cue environments. Learning to form the proposition: "When the bell rings, I expect food" does not appear too arduous (for humans and some other species at least); but is the same kind of propositional statement the *only* form of knowledge learned when situations become more complex?

Imagine an environment where cues and outcomes are continuous and relations are probabilistic. In such an environment, do organisms form propositions of the kind: "When the bell rings for more than 5 seconds (but not over 15 seconds), the green light is at 50% brightness, and the red light is off, I expect food approximately 80% of the time"? Research into multiple-cue-probability learning (Enkvist et al. 2006; Juslin et al. 2003), multi-attribute judgment (Newell & Bröder 2008), and categorization and concept learning (e.g., Allen & Brooks 1991; Nosofsky et al. 1989) has suggested that humans might try to learn such propositional information (i.e., rules) up to a point, but if the environment is too complex (e.g., cue-outcome relations are non-linear), or feedback is insufficient or inappropriate, other forms of knowledge – principally stored instances – are relied upon.

Mitchell et al. acknowledge that instance memories play a role (sect. 3.1) but state that "recollections of past bell-food pairings alone cannot produce learning" (sect. 3.1, para.6). Such a conclusion implies that experiments demonstrating behaviour accounted for by an exemplar model (which relies exclusively on stored representations of stimuli; e.g., Juslin et al. 2003; Nosofsky et al. 1989) are not demonstrations of learning. This conclusion seems too extreme. Participants in such experiments have learned to classify particular objects as belonging to Category A and others to Category B – they have learned an association between a stimulus (the to-be-classified-object) and a response (the category label). But the content of this learning appears to be instances rather than a proposition (see also Shanks & St. John 1994). The interplay (and relative influence) of instances and propositions is somewhat underspecified in Mitchell et al's approach. However, the implication is that learning can *only* occur when propositions (rules) are formed. This seems a step too far, especially in situations with multiple, non-binary cues.

Systems, processes, and their interaction. Mitchell et al. note that many dual-system models do not specify how systems interact with each other. This is certainly true and, moreover, it appears that there is little consensus across different areas on how such interaction might occur. For example, an influential dual-system model of category learning, COVIS (Ashby et al. 1998), proposes an initial bias towards an explicit hypothesis testing system, which is then usurped by an automatic, procedural system when the explicit system fails to learn. In contrast, popular dual-system theories of reasoning (e.g., Evans 2008) suggest that the initial bias is towards the automatic, intuitive system, which is only corrected by the explicit system when things appear to go awry. Part of the problem in specifying these interactions is that it is often not clear what is meant by a system or a process – and whether these terms are interchangeable (Evans 2008).

Mitchell et al. make a distinction (sect. 3.1), stating that their propositional approach is *not* a dual-*system* approach but that there are two types of *processes* (automatic processes of perception and memory and non-automatic processes of reasoning) in their single *learning* system. Although such specification clearly distinguishes their approach from the link-mechanism theories they wish to challenge, it blurs the distinction with many of the dual-process approaches to "higher-order" cognition. In their footnote, Mitchell et al. contrast their approach to other "dual-process or dual-system" theories of reasoning by stating that such approaches focus on *performance*, not learning – but at the same time Mitchell et al. want to incorporate memory (i.e., performance) processes into their *learning system*. Perhaps some clarification could be achieved by defining exactly what Mitchell et al. mean by a system (cf. Sherry & Schacter 1987).

New techniques for old problems. Much of the evidence for and against the propositional and link approaches reviewed in the target article comes in the form of task dissociations; that is, situations in which a variable (e.g., cognitive load, instruction) is claimed to have an effect on one system but no effect on another. However, dissociations are unable to bear the inferential weight placed upon them for several reasons. One reason is that a simple dissociation requires that a variable have *no* effect on a particular behavioural measure, an assertion that is impossible in principle to verify (Dunn 2003). Hence, although dissociations may be found, they are neither necessary nor sufficient for drawing inferences about the number of processes or systems underlying observed behaviour (Newell & Dunn 2008). An alternative technique which avoids the flaws of dissociation logic is state-trace analysis (Bamber 1979). This technique has already been applied successfully to many areas of cognitive science (see Newell & Dunn 2008), and its application to the areas reviewed in the target article might prove fruitful.

The new enlightenment hypothesis: All learners are rational

doi:10.1017/S0140525X09001095

Rita Nolan

Department of Philosophy, State University of New York at Stony Brook, Stony Brook, NY 11794-3750.
rita.nolan@stonybrook.edu
http://www.stonybrook.edu/philosophy/faculty/index.html#Nolan

Abstract: The proposal to recruit available formal structures to build an algorithmic model of all learning falters on close examination of its essential assumption: that the input and output of the model are propositional in structure. After giving three framework considerations, I describe three possibly fatal problems with this assumption, concluding each with a question that needs answering to avoid fatality.

I applaud Mitchell et al.'s expanded emphasis on cognition in learning theory, for our understanding pervades all we do. Nevertheless, there are fundamental problems with the propositional approach they propose. The title bills a propositional approach to human associative learning, animal learning being tucked in later as an egalitarian gesture, but the model proposed would be a standard neo-classic account of human learning in terms of a representational theory of mind *except for* its universal extension to all learning, human and otherwise. Such neo-classic accounts deem it explanation enough of some human behavior to hypothesize rich formal structures of inference and sentence generation internal to the organism as causes of like changes in behavior. The hypothesized structures are extrapolated from formal linguistics and formal logic. Some have found such explanations useful – not surprisingly for computer modeling of human linguistic behavior – but the target article's bold step is to extend the neo-classic model to all animal learning.

Mitchell et al. propose an algorithmic-level propositional model for all organismic learning that is sandwiched between a functional-level model and an implementation-level model. Algorithmic models of formal systems of inferences over formal structures of propositions exist, so the question is not whether what is inside the algorithmic box can be built. These inferential structures transform a propositional input into a propositional output; and they are sensitive to different conditions as constraints. Because the sandwich isolates the algorithmic-level box from any existential referents, to determine the explanatory adequacy of the model we are led to focus on the input/output structures as the locus of the psychological part of the explanation.

Proposition is a term of art, a moveable vector, but there must be some retained minimal content for its artful use to be contentful. It cannot remain an undefined abstract term and bear explanatory weight. Perhaps it seems that propositional structure is a well-defined formal concept and that this is all that is required for the algorithmic model to have content. Even so, the viability of the model as psychologically explanatory still requires assessment of its assignment of propositional structure to the input and output of the algorithmic box.

Human language users have a range of generalized information-bearing structures available that can be mistaken for propositional structures when they are not. So, *seeing a cat up a tree* differs informationally from *seeing that a cat is up a tree*, as *seeing a red box* differs from *seeing that a box is red. Learning to recognize an elm* differs informationally from *learning that an elm has double-toothed, feather-veined leaves. Learning how to tie your shoe* differs informationally from *learning that to tie your shoe, you first hold the left lace in one hand* [and so forth]. The input/output assumptions of the model assimilate to the propositional all structures such as these that mark off different sorts of perceptual and procedural cognitive achievements from propositional learning. How is the explanatory value of the model enhanced by trading in these finer-grained informational structures for the merely available and smooth operations in the box?

Outside the algorithmic box, a key reason for hypothesizing propositions is that they are taken to be the unique bearers of truth values and this requires that they can be either true or false. Recognition of this property operates essentially in any task of drawing valid inferences since their special feature is that they preserve truth. Of course we informally anthropomorphize the mental lives of animals and certainly some analogue to belief is exhibited by them; perhaps some primal state that preceded language in Modern Humans. But to predicate any propositional attitude of an animal more strictly speaking, and particularly belief, requires that the animal can distinguish truth *of the proposition* from *its* falsehood. For to believe a proposition is to believe that *it* is true, for which feat one must be able to believe that *it*, one and the same proposition, is false. No explicit concept of truth is required for this ability, nor is it supposed that a belief must be an occurrent mental phenomenon. Granting that *belief* is used as a term of art in the description of the input/output of the algorithmic box and thus dispensing with some of its everyday content, can its content relative to its use as a propositional attitude for the central objects that take part in inferential operations – propositions – be dispensed with, when it is exactly that use which the model aims to capture?

The input and output of the algorithmic model proposed by the propositional approach exhibit the fine-grained information-bearing structures of linguistic vehicles of assertion; they are sentences of a language in the form of statements. With the resources of language at hand comes a powerful, productive vehicle for describing whatever we notice; a feature that may make the propositional approach initially attractive for representing the cognitive changes of learning for all species. But the power and productivity of language can also pose a direct challenge to the requirement of falsifiability for a model. Language makes possible a vast number of available alternative propositional descriptions of any event and any belief content, even to a limiting case of [x believes that] *something happened*. This feature of the propositional approach allows very high flexibility in describing the input and the output. If a model is meant to explain anything then it must admit of falsification, but it is hard to see what could falsify it given this degree of flexibility. If some result appears to falsify the model, one can always redescribe the input and output, trying out different descriptions until hitting upon ones that work. Does this high flexibility make the model merely a redescription, not an explanation, of what it is meant to model (for a discussion, see Myung & Pitt 2002)?

It is essential to the proposed model that the input/output structures to the algorithmic box are propositional in structure, for these alone are the domain of inferential relations and the aim of the model is to construe all learning as inferential.

Is cultivating "biological blindness" a viable route to understanding behavioral phenomena?

doi:10.1017/S0140525X09001101

Andreas Olsson and Ame Öhman

Department of Clinical Neuroscience, Karolinska Institute, Psychology Section, Stockholm 17177, Sweden.

andreas.olsson@ki.se

http://www.psych.nyu.edu/phelpslab/pages/andreas.html

Arne.Ohman@ki.se

http://ki.se/ki/jsp/polopoly.jsp?d=15053&a=13512&l=sv

Abstract: Mitchell et al. propose that associative learning in humans and other animals requires the formation of propositions by means of conscious and controlled reasoning. This approach neglects important aspects of current thinking in evolutionary biology and neuroscience that support the claim that learning, here exemplified by fear learning, neither needs to be conscious nor controlled.

In an era characterized by a growing convergence among evolutionary biology, neurobiology, and behavioral sciences, Mitchell et al. make a bold claim. The authors argue that in humans (and all other animals with the explicit exception of *Aplysia*), learning to associate stimuli requires the formation of propositions (symbolic representations with truth-value) by means of conscious and controlled reasoning. Although swimming towards the (main)stream is sometimes necessary to reach the source, there are several reasons to believe that the authors are heading in the wrong direction.

In principle, the proposal that the formation of propositions is necessary for learning to occur – a process subserved by protein synthesis – is not different from the absurd claim that successful protein syntheses in the digestive systems must also be preceded by propositional representations. In fact, given our current knowledge of biological systems, it is highly unlikely that humans, not to speak about other animals, are conscious or in control of the majority of processes underlying learning and memory formation. Nonetheless, they may be both (i.e., conscious and in control) in regards to the *expression* of the acquired learning.

In an attempt to avoid linking their proposal to established biological principles of learning, the authors point towards the distinction between the psychological and the neural level of explanation and argue that their thesis applies only to the former. However, this illustrates a general problem inherent in the presented approach. By limiting the phenomena under investigation by either defining out of existence critical aspects of associative learning (e.g., their biological principles) or neglecting several lines of existing research (e.g., fear conditioning and lesions studies in humans), the authors end up proving little more than their assumptions. It is also noteworthy that the notoriously problematic terms "conscious" and "awareness," although central to the argument, are not explicated (except that the reader is reassured that "*Aplysia* do not have conscious beliefs"; target article, sect. 6.3, para. 3).

Here, we highlight three more specific problems with the approach presented by Mitchell et al.

1. The notion of a unitary type of associative learning resting on conscious awareness sits very uncomfortably with established ideas in evolutionary biology. Evolution is commonly conceived as a slow accumulative process, building layer upon layer of brain tissue that incorporates successful adaptations at one level into more complex functions at higher levels. As a consequence, we share many behavioral systems and their associated neural circuitry with our primitive predecessors, unlikely candidates for using awareness as their primary principle of learning. Nor does it seem a likely evolutionary feat to have reorganized the human brain for exclusive use of this principle to modify behavior. Rather, from the evolutionary perspective, many different forms of learning would be expected, as elaborated by, for example, Gregory Razran (1971). The MacLean (1993) concept of a "triune brain" is one, often discussed, example of layered evolution of this kind, which directly implies that there are at least three levels of behavioral organization, each of which may incorporate associative learning: one concerned with reflexes and instincts (brain stem and striatum), a second that incorporates emotion and autonomic control (the limbic brain), and a third level concerned with instrumental behavior and cognition (thalamus and the cerebral cortex).

2. Related to the lack of compatibility with evolutionary thinking is the omission of several lines of research within the neurosciences. A contemporary version of the MacLean concept is the model of rodent fear conditioning by LeDoux (1996), Davis (1992), Fanselow (1994), Maren and Quirk (2004), among others, which has been confirmed in human brain imaging studies (Morris et al. 1998). Because this model posits that the input to and output from the central hub in the fear network do not necessarily have to go through the cortex, it strongly implies that the fear network and its modification through fear conditioning are independent of conscious awareness. Therefore, this model (and its elaboration for human fear conditioning by Öhman & Mineka 2001) clearly implies two levels of learning that are partially independent but also interacting. This model provides an articulated version of dual-process theory that integrates neuroscience and behavior and is now supported by a host of both behavioral (Hamm & Vaitl 1996; Öhman & Soares 1998) and imaging work (Critchley et al. 2002), showing that conscious awareness of the associated stimuli or their contingency is not necessary for learning to be acquired and expressed.

Providing further support for the independence of (at least) two kinds of learning is the work on patients with lesions on the hippocampus, a structure known to be critical for the formation of declarative memories. Following fear conditioning, these patients fail to report the contingency between two associated stimuli (e.g., a neutral tone or image and an aversive shock) in a fear conditioning paradigm, but they show a normal conditioned response as measured by the skin conductance response, SCR (Bechara et al. 1995). In contrast, patients with lesions to the amygdala, a key player in the brain's fear network and known to be necessary for the implicit expression of learned fear, display the opposite response pattern with intact declarative memory, but an impaired conditioned response (spared conditioned response or SCR) (Bechara et al. 1995; Weike et al. 2005). These findings show a striking dissociation between explicitly (propositionally) and implicitly (SCR) expressed emotional learning. Further support along the same lines is the demonstration of fear conditioning to unseen visual stimuli in a cortically blind patient with bilateral lesions to the primary visual cortex (Hamm et al. 2003).

Taken together, the findings listed above should make it clear that the psychological and neural levels of explanations are tightly coupled and that psychological models of learning can benefit tremendously by drawing from what is known in the neurosciences.

3. In their argumentation against the "dual-process" theory of learning, Mitchell et al. build a straw man around the claim that the SCR as a measure of learning is unaffected by conscious and controlled cognitions. This purported claim, ascribed to the dual-process camp, is then used to refute that emotional learning can occur without conscious and controlled reasoning. Indeed, some work has shown that the SCR can be used to index learned responses to both consciously and non-consciously perceived stimuli (Öhman & Mineka 2001; Olsson & Phelps 2004). However, the authors neglect the literature which claims that the potentiation of the startle reflex may be less affected by propositional and declarative processes. Whereas the SCR has been shown to be more sensitive to cognitive processes, such as propositional reasoning, which is likely to be cortically mediated, the potentiating of the startle reflex and eyeblink conditioning draw mainly on subcortical and cerebellar mechanisms, respectively (Clark & Squire 1998; Davis 2006). This is supported by the accumulating evidence of a dissociation between the startle response and SCR, in which the SCR tracks conscious awareness of stimuli contingencies, whereas the startle response tracks non-conscious learning (Hamm & Vaitl 1996; Weike et al. 2007).

To sum up, the propositional account of associative learning proposed by Mitchell et al. may be parsimonious, but it critically lacks the compatibility with current evolutionary biology and neuroscience. Cultivating this form of "biological blindness" will not advance our understanding of behavioral phenomena, such as associative learning.

There is more to thinking than propositions

doi:10.1017/S0140525X09001113

Derek C. Penn,[a] Patricia W. Cheng,[a] Keith J. Holyoak,[a] John E. Hummel,[c] and Daniel J. Povinelli[b]

[a]*Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095;* [b]*Cognitive Evolution Group, University of Louisiana at Lafayette, New Iberia, LA 70560;* [c]*Department of Psychology, University of Illinois, Champaign, IL 61820.*

dcpenn@ucla.edu **http://www.cognitiveevolutiongroup.org/**
cheng@lifesci.ucla.edu **http://reasoninglab.psych.ucla.edu/**

holyoak@lifesci.ucla.edu http://reasoninglab.psych.ucla.edu/
jehummel@cyrus.psych.uiuc.edu
http://www.psych.uiuc.edu/~jehummel/ ceg@louisiana.edu
http://www.cognitiveevolutiongroup.org/

Abstract: We are big fans of propositions. But we are not big fans of the "propositional approach" proposed by Mitchell et al. The authors ignore the critical role played by implicit, non-inferential processes in biological cognition, overestimate the work that propositions alone can do, and gloss over substantial differences in how different kinds of animals and different kinds of cognitive processes approximate propositional representations.

All the co-authors of this commentary believe that associative learning theory is dead in the water. Penn and Povinelli (2007) have argued that associative learning alone is unable to account for causal reasoning in nonhuman or human animals. Hummel and Holyoak (1997; 2003) have argued that associationist-style representations are inadequate for modeling human relational reasoning. Cheng (1997) showed that both human and nonhuman causal learning involves rational inferences that go far beyond keeping track of covariation information. And Penn et al. (2008) have recently argued that animals of many taxa employ relationally structured, functionally compositional representations and that human minds, in particular, closely approximate the higher-order features of a classical "language of thought." Indeed, one of us actually anticipated the thesis proposed by Mitchell et al. more than a decade ago: "Is it possible," Waldmann and Holyoak (1992) asked in the final sentence of their paper, "that lower-order associative learning should be reduced to higher order causal induction, rather than vice versa?" (p. 235).

So, at first glance, it would seem that we would be strong supporters of the "propositional approach" proposed by this target article. We are not. *Some* cognitive processes in both human and nonhuman animals involve controlled, effortful inferences operating over highly structured relational representations, but Mitchell et al.'s claim that *all* learning is effortful, conscious, and propositional is unfounded and implausible.

Not all learning is conscious and inferential. Certainly, *some* learning involves hypothesis testing and conscious propositional beliefs – at least in humans. But there is overwhelming evidence that many forms of learning are implicit and non-inferential. To cite just the most obvious examples: priming, motor-skill learning, fear conditioning, and implicit category learning. Mitchell et al.'s thesis completely fails to account for these processes.

Mitchell et al. also miss the critical distinction between conscious awareness of a cognitive process and conscious awareness of the *output* of the process (e.g., Nisbett & Wilson 1977), a central feature of most dual-process models of cognition (Evans 2008). Instead, Mitchell et al.'s thesis seems to resurrect Descartes' notion of a transparent mind fully conscious of its own reasoning processes.

We need more processes, not fewer. Like Mitchell et al., we are not fans of dual-process theories. But our problem is the opposite of Mitchell et al.'s: In our view, there are many more than two kinds of processes involved in human and nonhuman cognition. As Evans (2008) shows, the idea that there is one system that is purely conscious, inferential, and propositional, and one other that is purely automatic, implicit, and associative, is no longer sustainable. Mitchell et al.'s one-process theory fails to explain all the phenomena that motivated dual-process models to begin with (see Evans 2008).

Propositions are not enough. We are big fans of propositions. But propositions *alone* are ill-suited for many aspects of biological cognition – for example, pattern matching and completion, graded semantic flexibility, and stimulus generalization. Any plausible model of biological cognition must incorporate *both* the structural features of propositional representations *and* the automatic, graded flexibility of distributed, associative representations (Hummel & Holyoak 1997).

Propositions are not just structured relations. Mitchell et al. briefly mention that propositions specify "the way in which events are related" (sect. 1, para. 5). While true, this is just the first step down a long path towards full-fledged propositions. There are many other critical features of propositions that Mitchell et al. omit, such as the capacity to systematically represent types, variables, roles, and higher-order relations, and to perform rule-governed operations over these representations in an inferentially coherent fashion (Hummel & Holyoak 1997).

Crucially, these propositional features do not form a package by nomological necessity (cf. Fodor & Pylyshyn 1988). In our view, non-human animals approximate certain features of propositions and not others (Penn et al. 2008). By reducing propositions to structured relations, Mitchell et al. gloss over all the interesting computational and comparative challenges.

Causal learning is not monolithic. We agree that *causal* learning is not purely associative in either human or nonhuman animals (Cheng 1997; Penn & Povinelli 2007; Waldmann & Holyoak 1992). But this does not mean that all kinds of causal learning or all kinds of animals employ the same degree of propositional sophistication (Penn et al. 2008). There is good evidence that nonhuman animals employ structured representations and are capable of first-order causal inferences. But this does not mean that rats employ the "higher-order reasoning processes" employed by humans (see Penn & Povinelli 2007). In the case of Beckers et al. (2006), for example, the rats' inferences can be modeled as a kind of sequential causal learning that does not require higher-order relational representations (Lu et al. 2008).

Rescorla-Wagner is not propositional. Mitchell et al. claim that the Rescorla-Wagner model can be thought of as a "simple mathematical model of propositional reasoning" (sect. 6.1, para. 10). Yes, the Rescorla-Wagner model could be implemented symbolically; but that does not make it a rational model. The Rescorla-Wagner model assumes a linear generating function and lacks a representation of causal power (Cheng 1997). Ironically, Mitchell et al. miss why associationist theories fail as rational models of causal reasoning.

PDP models are better and worse than Mitchell et al. claim. Mitchell et al. have an idiosyncratic view of parallel distributed processing (PDP) models. On the one hand, they claim that traditional PDP models can account for the propositional capabilities of humans. On the other hand, they claim that "a single node in a PDP model does not represent anything" (sect. 6.2, para. 3). They are wrong on both accounts. PDP models are incapable of representing the structured relations that Mitchell et al. claim are the sine qua non of learning (Hummel & Holyoak 1997). But this does not mean they represent nothing at all. Every node in a PDP network has some equivalence class, and this equivalence class is precisely what it represents. Just because this equivalence class does not correspond to something one can point at does not mean it does not exist.

Darwin and neuroscientists are not all wrong. Mitchell et al. admit that the neuroscientific evidence provides little support for their claim that all learning is propositional. But they dismiss this evidence as inconclusive. They take a similarly dismissive attitude towards the comparative evidence. They admit that *Aplysia* do not employ propositions. But they have no evolutionary explanation for what happened to the *Aplysia*'s primordial associative learning mechanisms in more sophisticated creatures such as rats and humans. Did all this pre-propositional baggage simply shrivel up and die? We think not.

Many nonhuman animals have cognitive capabilities that go far beyond the automatic formation of simple links. But the degree to which propositional mechanisms are employed differs between different kinds of animals and between different kinds of cognitive process within a given individual. There are some ways in which humans are still very much like *Aplysia*. There are other ways in which we are unique.

Like Darwin, Mitchell et al. overestimate the propositional capabilities of nonhuman animals (Penn et al. 2008). But worse, Mitchell et al. ignore the incremental and cumulative fashion in which evolution crafted the various kinds of minds on this planet. Darwin did not make that mistake.

The computational nature of associative learning

doi:10.1017/S0140525X09001125

N. A. Schmajuk and G. M. Kutlu

Department of Psychology and Neuroscience, Duke University, Durham, NC 27516.

nestor@duke.edu **gunes.kutlu@duke.edu**

Abstract: An attentional-associative model (Schmajuk et al. 1996), previously evaluated against multiple sets of classical conditioning data, is applied to causal learning. In agreement with Mitchell et al.'s suggestion, according to the model associative learning can be a conscious, controlled process. However, whereas our model correctly predicts blocking following or preceding subadditive training, the propositional approach cannot account for those results.

In their target article, Mitchell et al. point out that, in contrast to the propositional approach, associative models cannot explain some causal learning results. Here we show that an attentional-associative model, previously evaluated against multiple sets of classical conditioning data, provides explanations for causal learning experiments.

An attentional-associative model of conditioning. Schmajuk et al. (1996; henceforth SLG) proposed a neural network model of classical conditioning (see also, Larrauri & Schmajuk 2008; Schmajuk & Larrauri 2006). The network incorporates (a) an attentional mechanism regulated not only by novelty (difference between actual and predicted magnitude) of the unconditioned stimulus (US) as in the Pearce and Hall (1980) model, but also by novelty of the conditioned stimuli (CSs) and the context (CX); (b) a network in which associations are controlled by a modified, moment-to-moment (vs. trial-to-trial) constrained version of the Rescorla and Wagner (1972) competitive rule; and (c) feedback from the associative network to the input. The attentional mechanism was designed to explain latent inhibition (Lubow & Moore 1959), and the feedback loop was included to describe inferential processes such as sensory preconditioning.

Gray et al. (1997) showed that the SLG model also describes automatic (or unconscious) and controlled (or conscious) processing (Pearce & Hall 1980; Schneider & Shiffrin 1977). In the framework of the model, stimulus X might be processed in controlled or conscious mode when environmental novelty and the representation of X, X_X, are large; and in automatic or nonconscious mode when novelty and X_X are small. Therefore, in agreement with Mitchell et al.'s position, the SLG model suggests conditioning occurs mostly consciously. However, according to the model, in the case of latent inhibition, a preexposed X with a small X_X remains unconscious. Therefore, in line with Mitchell et al.'s reference to the effects of masking on learning processes, the SLG model suggests that X pre-exposure reduces conscious processing of the X but conditioning still occurs at a slower pace.

Causal learning. Several studies on causal learning were concerned with the effect of additivity information on blocking and backward blocking (e.g., Beckers et al. 2005). Blocking refers to the fact that a potential cause X is not considered a cause of a given outcome (OUT, represented by "+") when it is presented together with another potential cause A, if A had been previously shown to be a cause of that US (A+, AX+). Two potential causes, G and H, are additive if, when presented together OUT is equal to the sum of their OUTs when presented separately (this is represented as G+, H+, GH++). When the joint OUT of G and H is less than the sum of their individual OUTs, the causes are subadditive (G+, H+, GH+). Beckers et al. (2005) demonstrated that additivity pre-training resulted in stronger blocking than subadditivity pretraining (Experiment 2); additivity pre-training resulted in stronger backward blocking than subadditivity pre-training (Experiment 3); additivity post-training resulted in stronger blocking than subadditivity post-training (Experiment 4); and blocking is stronger when OUT is weaker than the maximum OUT experienced by the subjects (Experiment 1). According to Beckers et al. (2005), their results can be explained in inferential terms: blocking is not present if either the additivity premise or the submaximal premise is not satisfied. In the following paragraphs, we describe how the model addresses two of these experimental results.

Additivity training preceding blocking. Like the Rescorla-Wagner model, the SLG model explains blocking because, at the time of the presentation of X, A already predicts the OUT. According to the model, the compound stimulus (C) activated by G and H and associated with OUT during pre-training, is fully activated by A and X. This association, together with the blocking stimulus A, contributes to predict the OUT, thereby increasing blocking. Because the C-OUT association acquired during pre-training is stronger in the additive than in the subadditive case, blocking is stronger in the former than in the latter case (see Fig. 1, Left Panels).

We assumed generalization between compounds GH and AX to be strong based in Young & Wasserman's (2002) experimental data showing that generalization between elements is much smaller than generalization between compounds. In addition, the model implements generalization among elements and between elements and compounds through the presence of a common contextual stimulus that is always active.

Additivity training following blocking. As Mitchell et al. correctly observe in the target article, in the absence of pre-training, the C compound is already associated with OUT during post-training and, therefore, increased C-OUT associations cannot be used to explain increased blocking. Interestingly, the SLG model provides an attentional interpretation for the result. In terms of the model, during the AX+ phase of blocking, OUT-X and C-X associations are formed. During the subsequent additivity post-training, OUT-X and C-X associations predict X, but X is not there. In the additive case, the stronger OUT extinguishes its OUT-X association faster than the weaker non-additive OUT does. During additivity post-training, presentation of the novel stimuli G and H, as well as the absence of stimuli A and X, increases novelty. Thus, because the representation of X is weaker in the additive case, attention to X increases less, and blocking is stronger than in the subadditive case (see Fig. 1, Right Panels).

Conclusion. In agreement with Mitchell et al.'s position, the SLG model suggests that associative learning can be a conscious, controlled process related to higher-order cognition. Furthermore, in addition to the above experiments, computer simulations show that the SLG model describes (a) the facilitatory effect of additivity training before backward blocking (Beckers et al. 2005), (b) maximality effects (Beckers et al. 2005, Experiment 1), (c) the facilitatory effect of subtractivity pre-training results on backward blocking (Mitchell et al. 2005), and (d) higher-order retrospective revaluation (De Hower & Beckers 2002). Interestingly, whereas the propositional approach predicts no blocking following subadditive pre- and post-training (see Beckers et al. 2005, pp. 241, 246), the SLG model can account for those results. Furthermore, these finely graded results are also present in the model description of latent inhibition, in which weaker conditioned responding is observed.

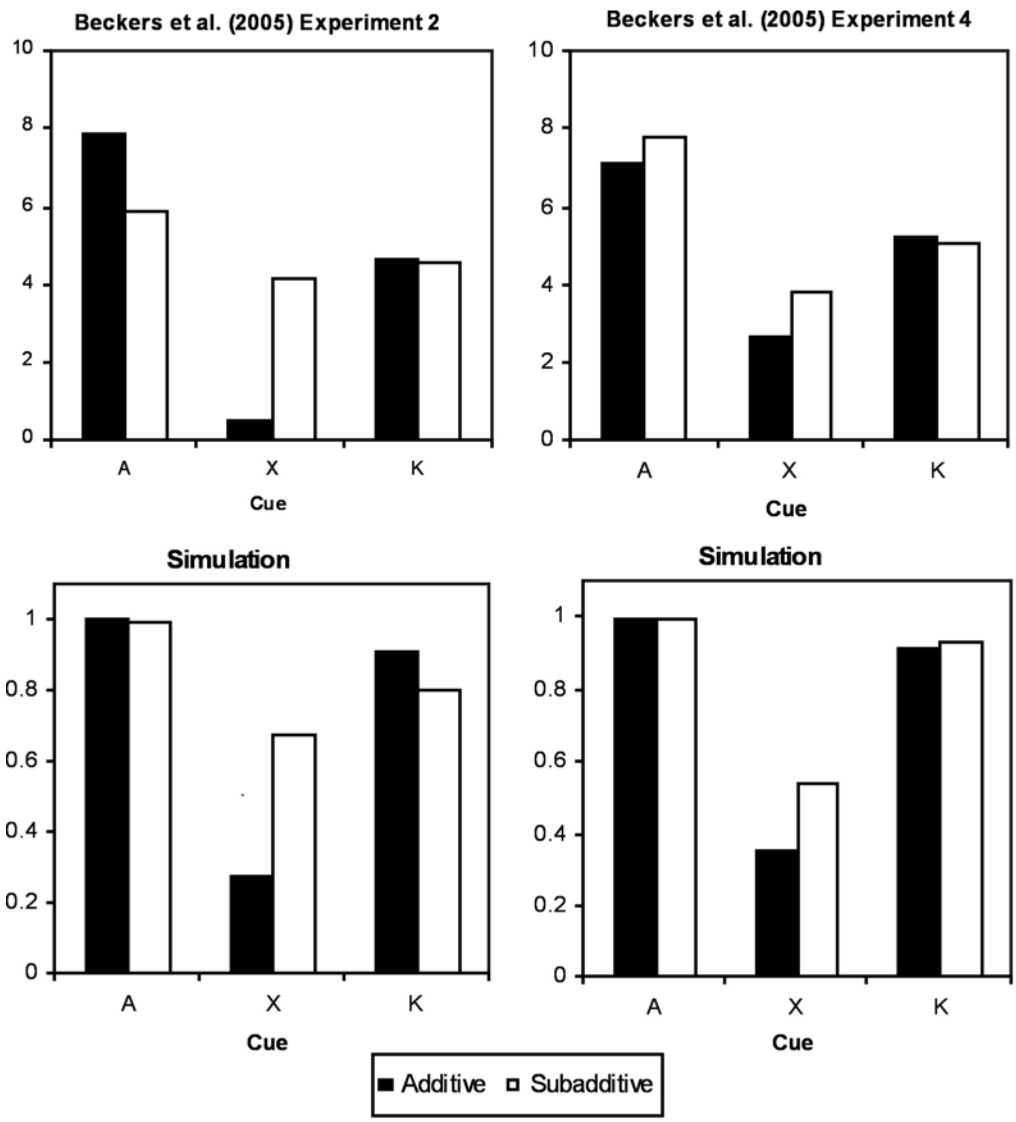

Figure 1 (Schmajuk & Kutlu). *Left Upper Panel*: Mean causal ratings for cue *A*, *X*, and *K* in a blocking experiment following additivity pre-training; data from Beckers et al. (2005, Experiment 2). *Right Upper Panel*: Same blocking experiment followed by additivity post-training, data from Beckers et al. (2005, Experiment 4). *Lower Panels*: Corresponding simulations with the SLG model. Parameters used were analogous to those used in previous papers. The Outcome is represented as a CS. Rating is given by the sigmoid $R = \text{Predicted Outcome}^6/(\text{Predicted Outcome}^6 + \beta^6)$, where $\beta =$ Average of predicted outcomes for all *A*, *X*, and *K*. Simulations for Experiment 2 included 90 additive or subadditive pre-training, 20 *A*+, 40 *AX*+ trials, and 40 *KL*+ trials. Simulations for Experiment 4 included 20 *A*+, 20 *AX*+, 20 KL+, and 90 additive or subadditive post-training trials. Stimulus duration was 10 time units, stimulus intensity was .6, *C* intensity was 1, OUT was 2 for additivity and 1 for subadditive training.

Of mice and men: Revisiting the relation of nonhuman and human learning

doi:10.1017/S0140525X09001137

Holger Schultheis[a] and Harald Lachnit[b]

[a]*Transregional Collaborative Research Center Spatial Cognition, University of Bremen, 28359 Bremen, Germany;* [b]*Department of Psychology, Philipps-University Marburg, 35032 Marburg, Germany.*

schulth@sfbtr8.uni-bremen.de

http://www.cosy.informatik.uni-bremen.de/staff/schultheis/

lachnit@staff.uni-marburg.de

http://www.staff.uni-marburg.de/~lachnit/

Abstract: To support their main claim, Mitchell et al. broach the issue of the relationship between the learning performance of human and nonhuman animals. We show that their argumentation is problematic both theoretically and empirically. In fact, results from learning studies with humans and honey-bees strongly suggest that human learning is not entirely propositional.

Mitchell et al. argue that learning relies on propositional reasoning in humans, as well as in certain other animals such as rats. They support their view with similar results for both species in certain learning studies. At the same time, they admit that not all nonhuman animals (e.g., *Aplysia*) draw on propositional processes. Taking their argument seriously, we will show that their stance creates grave problems in accounting for human learning exclusively by propositions.

The first problem arises from the fact that Mitchell et al.'s view effectively partitions all animals into two groups; those, most notably humans, whose learning relies entirely on propositional reasoning and those whose learning is completely without propositions (e.g., *Aplysia*).

Such partitioning implicitly establishes a "magical" point in human phylogenesis where animals stopped being equipped with non-propositional learning. But why should evolution favor individuals who throw away mechanisms that have proven adaptive for generations of ancestors? It seems much more likely that, instead of disposing of the mechanism, individuals may have complemented it by acquiring other mechanisms such as propositional reasoning. Hence, it seems more reasonable to assume that humans are equipped with two, rather than only one, learning system. Against this background, any account which claims that human learning is entirely propositional is only satisfactory if it also explains why previously existing learning systems have been shut down during phylogenesis. However, such an explanation is not provided by Mitchell et al.

The second problem arises from the methodology used to establish the point of propositional learning in non-human animals. Based on behavioral similarities of nonhuman animals and humans, the learning mechanisms assumed for humans are also thought to be at work in the nonhuman animals. This method for ascribing learning mechanisms, however, cuts both ways. If nonhuman animals and humans behave similarly in learning, one can as well ascribe the nonhuman learning mechanisms to humans. Either Mitchell et al. are not aware of the two-edged nature of their method or they are not aware of results on nonhuman learning which shed doubt on their main claim.

Some of these results have been obtained using honey-bees as subjects. Komischke et al. (2002), for example, trained bees on reversal problems, where a previously reinforced conditioned stimulus (CS) was subsequently not reinforced, and vice versa. The bees showed improved reversal learning performance, depending on the number of previously encountered reversals. Thus, one could argue that the bees acquired some kind of a (propositional) reversal rule.

In another series of experiments, Lachnit et al. (2007) and Deisig et al. (2007) investigated the effects of trial spacing on learning performance in humans and bees, respectively. Importantly, the two studies yielded comparable results. For both, humans and bees, increasing the inter-trial interval led to a general increase of the conditioned response (CR) and an improved ability to differentiate between CS+ and CS- in learning positive patterning (PP) and negative patterning (NP) discrimination problems.

Further similarities between the learning of humans and bees have been demonstrated by Kinder and Lachnit (2003) and Deisig et al. (2003). In these studies, subjects were presented with an extended NP discrimination ($A/B/C+$, $AB/AC/BC+$, $ABC-$). Contrary to major associative learning theories available at that time, both bees and humans showed no CR difference for the single- and double-compound CSs.

These five studies show substantial convergence of learning in humans and bees. Direct comparisons indicate that humans and bees behave similarly regarding several learning problems and experimental manipulations. Consequently, following the method of Mitchell et al., humans and bees can be assumed to draw on similar learning mechanisms. Mitchell et al. probably would argue that propositional reasoning is at work, because human learning relies completely on propositional reasoning. Hence, according to Mitchell et al., learning in bees also relies on propositional reasoning.

Yet, the results of Komischke et al. (2003) clearly contradict this conclusion. Bees had to solve NP and PP discriminations under various conditions. In one condition, the CSs of the NP problem were presented to both antennae of the bees, whereas in another condition, the CSs were presented to only one of the antennae. A third condition simultaneously presented PP to one antenna and NP to the other antenna (with identical CSs for both problems): (a) NP was learned only in condition 1 and 3; (b) the PP discrimination was hampered in condition 2.

These results contradict the idea that bees' learning relies on propositional reasoning. According to the propositional account, learning should degrade in dual-task situations. This was observed for PP; NP, however, improved under the dual-task condition. Hence, bees' learning did not rely on propositional reasoning. If learning in bees is not propositional and learning in bees and humans is comparable, using the method of Mitchell et al., one can only conclude that humans – at least sometimes – also do not learn propositionally.

Interestingly, this conclusion regarding non-propositional learning in humans on PP and NP tasks is further corroborated by Lachnit and Kimmel (1993) and Lachnit et al. (2002). Lachnit and Kimmel (1993), employing a design similar to that of Shanks and Darby (1998), observed an asymmetric relation of PP and NP in transfer situations: Learning PP hampers NP, but not vice versa. However, utilizing different response systems (spared conditioned response [SCR] and eye-blink) for the different patterning problems, Lachnit et al. (2002) found no such asymmetry; PP and NP were solved equally well instead. If learning to solve PP and NP discrimination problems relies on the construction of consciously available general rules, why then should the effect of learning these rules be response-specific?

In light of the evidence presented here, the account of Mitchell et al. in the target article seems untenable. Not only do the authors fail to present important information to justify their account, but empirical evidence on human and nonhuman learning also contradicts their claim that human learning is entirely propositional.

The associative nature of human associative learning

doi:10.1017/S0140525X09001149

David R. Shanks
Division of Psychology and Language Sciences, University College London, London WC1H 0AP, United Kingdom.
d.shanks@ucl.ac.uk **www.ucl.ac.uk/psychlangsci**

Abstract: The extent to which human learning should be thought of in terms of elementary, automatic versus controlled, cognitive processes is unresolved after nearly a century of often fierce debate. Mitchell et al. provide a persuasive review of evidence against automatic, unconscious links. Indeed, unconscious processes seem to play a negligible role in any form of learning, not just in Pavlovian conditioning. But a modern connectionist framework, in which "cognitive" phenomena are emergent properties, is likely to offer a fuller account of human learning than the propositional framework Mitchell et al. propose.

We should not be too harsh on ourselves for having failed, after a century of study, fully to have worked out the basic nature (cognitive or automatic) of human learning. Psychologists have been struggling with this paradox ever since Thorndike first formulated the law of effect (Thorndike 1931; cf. Postman 1947; Spence 1950).

The paradox is highlighted by the following facts:

1. As Mitchell et al. rightly point out, awareness appears to be a necessary condition for learning. Their review focuses on conditioning, but the point holds for many other forms of learning such as speeded responding to structured materials (Perruchet & Amorim 1992; Shanks et al. 2003), context-guided visual search (Smyth & Shanks 2008), grammar learning (Tunney & Shanks 2003), decision making (Maia & McClelland 2004), and many others.

2. Learning is not an automatic process. It is controlled by both bottom-up influences (by the stimuli and their relationships) and by top-down ones (how the stimuli are perceived; attention; expectancies; working memory capacity; and so on). How could learning be automatic given the evidence that

stimuli are not even perceived when attention is fully diverted elsewhere (Macdonald & Lavie 2008)?

3. Many aspects of learning seem to involve reasoning. For instance, after learning that two cues predict an outcome (*AT*+), presentation of evidence that one of them alone predicts the outcome (*A*+) causes less predictive influence to be assigned to *T* (Shanks 1985; Van Hamme & Wasserman 1994).

4. If by "reasoning" one means the manipulation of symbolic representations, then much of cognition does not appear to be well described as reasoning. The embodied cognition movement has made it clear that many aspects of behaviour traditionally interpreted in terms of inferences over amodal symbolic representations are better explained via notions of mental simulation (Barsalou et al. 2003; Niedenthal 2007). Moreover, logic-based accounts of reasoning have been subjected to severe criticism even in such central domains as reasoning about conditional statements ("If *A* then *B*") (Oaksford & Chater 2007).

5. Extraordinarily rich explanations of learning phenomena have been achieved by models built out of automatic link machinery (i.e., connectionism). Such models demonstrate massive "emergentism," in that processes that seem cognitive and high-level emerge from the operations and interactions of very elementary processing units. Indeed, these models can often be viewed as operating in optimal (Bayesian) ways.

One way to resolve the paradox is to ignore (4) and (5) and argue, as Mitchell et al. do, that the basic processes of associative learning intrinsically embody the principles of reasoning. Indeed, it is easy to combine a logic-based system (based on a computer programming language for symbolic reasoning) with a Rescorla-Wagner-like rule governing belief strength (Shanks & Pearson 1987), such that inference over propositions yields behaviour with the appropriate level of strength.

Yet such a propositional framework for learning only scores 3 out of 5 on the list above. An alternative resolution which scores rather better begins by noting that many things that are true of automatic links are not necessarily true of larger-scale connectionist models. Unlike automatic links, for instance, connectionist models can represent semantic information. Indeed, if there has been a single goal behind the connectionist movement, it has been to emphasize this fact. Such models can "reason." A simple connectionist model described by Ghirlanda (2005) explains the retrospective revaluation effect described in (3) above, and some of the other reasoning-like effects described by Mitchell et al. are beginning to be modelled in connectionist systems (e.g., Schmajuk & Larrauri 2008). Unlike links, processing in connectionist models is often assumed to be related to awareness (states of settled activity – attractor states – may be just those states of which we are conscious). Unlike links, connectionist models have no difficulty in binding top-down and bottom-up influences. Many models incorporate pathways for top-down attentional control. And so on. There is a long way to go, but it is not inconceivable that such an approach will eventually make the paradox of learning dissolve.

Close examination of the empirical data also adds weight to the view that at least some aspects of learning emerge from elementary link processes and questions the propositional reasoning account. Quasi-rational behaviour, such as blocking, occurs not only in intentional learning situations, but also in incidental ones in which it seems very unlikely that the individual would be motivated to "reason." For example, in speeded reaction time tasks in which some structural property is informative about a target's location, cue-competition effects are observed (Endo & Takeda 2004). Such effects are well modelled in connectionist systems (Cleeremans 1993).

Further evidence for link-like processes emerges in experiments in which individuals judge event probabilities after exposure to a cue-learning task. As Mitchell et al. explain, such studies show that cue-outcome contingency has an impact on probability estimates even when variations in contingency do not affect the objective probabilities. Hence, if the probability of an outcome given a cue, $P(O|C)$, is say 0.75, participants' judgments will be greater when the probability of the outcome in the absence of the cue, $P(O|{\sim}C)$, is 0 rather than 0.75 (Lagnado & Shanks 2002). Mitchell et al. argue (sect. 5.2) that such effects arise because of participants' confusion or uncertainty about the term "probability" in the experimental instructions. But several studies (Nosofsky et al. 1992; Shanks 1990) show the same bias in "implicit" probability estimates when probability language is absent. In these conditions, participants choose which of two outcomes is the correct diagnosis for a patient with a certain symptom pattern. The word "probability" is not even employed – participants are asked to choose which outcome they think is correct. Such studies also challenge Mitchell et al.'s suggestion that the effect is due to confusion about the absence of other cues, as it also emerges when binary dimensions are used in which there are no absent cues. These biases, which fall naturally out of link-based models, are hard to reconcile with propositional reasoning accounts.

How do we get from propositions to behavior?

doi:10.1017/S0140525X09001150

Daniel A. Sternberg[a] and James L. McClelland[b]

[a]*Department of Psychology, Stanford University, Stanford, CA 94305;*
[b]*Department of Psychology and the Center for Mind, Brain, and Computation, Stanford University, Stanford, CA 94305.*

sternberg@stanford.edu **jlm@psych.stanford.edu**
http://psych.stanford.edu/~jlm

Abstract: Mitchell et al. describe many fascinating studies, and in the process, propose what they consider to be a unified framework for human learning in which effortful, controlled learning results in propositional knowledge. However, it is unclear how any of their findings privilege a propositional account, and we remain concerned that embedding all knowledge in propositional representations obscures the tight interdependence between learning from experiences and the use of the results of learning as a basis for action.

Mitchell et al. have made a number of important contributions to our understanding of human contingency learning. They cite many compelling studies that demonstrate the strong influences of verbal instruction and processing resources on learning. Along the way they have become committed to a particular representational format that may limit their ability to explain the breadth of learned human behavior.

We have two main concerns with their proposal that learning is the result of controlled reasoning processes that operate on propositional representations. First, it is unclear that many of the empirical findings they describe automatically follow from a propositional approach. These would instead seem to depend on the particular models of learning and memory that one employs, not on the use of a propositional format per se. Second, if Mitchell et al. are indeed arguing that *all* learned knowledge is encoded as propositions with attached levels of belief – a very strong claim – they need to provide a convincing account of how we could learn to perform complex, context-sensitive behaviors using processes that operate only on propositions and beliefs. Without such an explanation, their theory can only speak to a restricted set of data drawn from standard human contingency learning paradigms, and not to human and animal learning in its broader sense. We elaborate these points below.

In their discussion of cognitive load, Mitchell et al. describe a series of studies by De Houwer and Beckers (2003) who found a reduction of forward blocking in a standard contingency learning task when a moderately engaging secondary task was introduced. At the same time, participants continued to demonstrate robust learning of the explicit contingencies they experienced during training. Although Mitchell et al. do not provide an explanation as to why blocking should require more cognitive resources than learning all of the other cue-outcome relationships in the task, we have been interested in whether the ability to retrieve relevant memories may play a role in generating these effects.

We have recently explored cue-competition effects in two tasks using common stimuli and abstract contingencies but different task demands: a self-paced prediction task that is similar to a standard contingency learning experiment, and a fast-paced, cued reaction task (Sternberg & McClelland, in preparation). In both tasks, participants saw one or two cue objects on each trial, some of which were usually followed by the appearance of a dot. In the prediction task, participants had unlimited time to observe the cue objects before pressing a key to predict whether the dot would appear on that trial. After making their prediction, the actual outcome appeared and feedback was given. In the reaction task, on trials where the dot occurred, it appeared 350 msec after the cue object(s), and participants had to press a key within 275–400 msec (this deadline decreased at a constant rate during training). On trials where the dot did not appear, participants had to refrain from responding. Both groups were able to learn the explicit contingencies they experienced during training, as revealed by reaction times (RTs) and contingency ratings in the reaction task, and by test-trial predictions and contingency ratings for the prediction task. However, while prediction participants showed clear competition effects, evidence for competitions effect in the reaction task was scant.

Mitchell et al.'s view that knowledge of the contingencies is stored in the form of propositions does not, in our view, shed light on the differences between these conditions. The form of storage could be propositions, images, or something else – regardless, the explanation could hinge on constraints on the time available for retrieval of relevant prior episodes from memory. On this view, blocking would depend on considering both the relevant singleton and the related pair at the same time, even though these events are not presented together, whereas learning the direct events would depend only on recalling past experiences with the presented stimulus. Because the reaction task reduces the time available for recollection, we would expect to see a reduction in cue-competition effects in this task compared to those observed in the prediction task.

We also consider it to be uncontroversial that the operation of any learning process should depend on the availability of cognitive resources such as attention and memory. There have been many theories of multiple *interacting* learning processes that make no claims about a completely autonomous and resource-free associative learning system (e.g., McClelland et al. 1995). As it is likely that a number of interacting neural systems can spring into action when a novel or familiar stimulus is encountered, it will be important to continue to study in more detail the role of attention (see Kruschke & Blair 2000; Kruschke et al. 2005) and memory (McClelland & Thompson 2007; Vandorpe et al. 2007b) in learning and reasoning.

The propositional account that Mitchell et al. advocate also may run into difficulties addressing details of situation-specific response behaviors. The same contingency may be learned in different behavioral contexts, and require different kinds of responses, and the responses may not transfer even if the contingencies remain the same. For example, we would not expect that participants who had learned cue-outcome contingencies in our prediction task would immediately show fast RTs to relevant items if tested using our reaction task, even if they were verbally instructed that the same contingencies they had previously learned still held. It seems likely that something more like a sensory-motor skill of responding to particular stimuli has been acquired by these participants. Such skills, and not simply verbal propositions, seem fairly clearly to underlie abilities like playing an instrument or driving a car.

Finally, Mitchell et al.'s current theory sheds little light on the *process* of learning complex behaviors. Consider Cleeremans and McClelland (1991), who trained participants to make button responses to sequences of visual stimuli that were generated by a moderately complex finite state grammar. They found that participants show graded sensitivity to the conditional probabilities between items, and that this sensitivity developed gradually across training sessions. They were able to capture the learning trajectories in the human data using a simple recurrent network (SRN) model. While the structure of the information in the task could certainly be represented in a propositional form (as it was represented in this way in the experiment script), the SRN provided an account of the process of learning through time, in addition to its final state. The key ingredient here appears to be some sort of gradual strengthening process, rather than the presence or absence of a propositional statement of contingency.

Mitchell et al. conclude by presenting us with a tall order. Because they find little evidence for an automatic link-formation system, they suggest that we should recast all of the existing literature on human learning as evidence about the operation of a single propositional learning system. This conclusion seems to us to present a false dichotomy. The studies the authors present on cognitive load and verbal instruction manipulations are indeed important and useful challenges to simple associative accounts of contingency learning, but by embedding all knowledge in a propositional form, Mitchell et al. may sacrifice an account of how we learn complex behaviors.

Automatic (spontaneous) propositional and associative learning of first impressions

doi:10.1017/S0140525X09001162

James S. Uleman
Department of Psychology, New York University, New York, NY 10003.
jim.uleman@nyu.edu **http://www.psych.nyu.edu/uleman**

Abstract: Contrary to the target article's claims, social cognition research shows considerable learning (about other people) that is relatively automatic. Some of this learning is propositional (spontaneous trait inferences) and some is associative (spontaneous trait transference). Other dichotomies – for example, between learning explicit and implicit attitudes – are also important. However conceived, human conditioning is not synonymous with human learning.

Alice solved the mystery halfway through the book. And Bob kicked the puppy out of his way while crossing campus.

Of course, you have no idea who Alice and Bob are, and probably have no interest in finding out. But if you are like most participants in our studies, you have already inferred that Alice is clever and Bob is cruel. Furthermore, you made these inferences without realizing it, and would likely deny them if asked.

The extensive research on such spontaneous trait inferences (STIs) and related phenomena supports some of Mitchell et al.'s arguments but challenges others. STI research does not use the framework or procedures of classical conditioning, so many of the terms in the target article (unconditioned stimulus [US], conditioned response [CR], blocking, overshadowing) are not applicable. But others are – for example, learning, propositions, associations, automatic and controlled processes, conscious and unconscious processes. So STI findings have direct bearing on the target article's more general claims.

Automatic propositional learning about others occurs. In a typical study of spontaneous inferences, participants read a series of trait-implying statements about other people, with an explicit goal *other than* forming impressions of them. The participants may be asked to memorize the material, or just familiarize themselves with it. Then they are tested in indirect ways to see whether they made trait inferences. Cued-recall, a favorite test in early research, also showed that participants were unaware of inferring anything about the targets (see Uleman et al. 1996b, for a review). That is, these inferences are unconscious. Although certainly aware of the stimuli, participants are not aware of their inference processes or outcomes.

The best evidence that STIs occur at encoding, rather than at retrieval, comes from on-line measures of concept activation. Zárate et al. (2001) had participants read trait-implying sentences for a subsequent memory test and simultaneously do a lexical decision task. Participants were quicker to identify as words implied traits that immediately followed trait-implying sentences than control sentences. Uleman et al. (1996a) used a slightly different task. Participants had to indicate, as quickly as possible, whether probe test words that immediately followed each sentence had literally been in the sentence. As predicted, participants were slower to correctly say "No" when sentences implied the trait probes than when they did not. So even though trait inferences interfered with optimal task performance, participants made them anyway. STIs are uncontrollable, as well as unintended.

Bargh (1994) identified efficiency as the fourth criterion of automaticity, and there is good evidence that STIs are highly efficient. Todorov and Uleman (2003) used a false recognition paradigm in which participants read 60 pairs of photos and sentences for a subsequent memory test. Then they judged whether particular traits had appeared in the sentences paired with particular photos. False recognition of implied traits was higher when traits were paired with actors' photos than with other photos, even when participants only saw each photo-sentence pair for two seconds, or viewed them to count the number of nouns in each sentence (not for a subsequent memory test), or viewed them while concurrently rehearsing a six-digit number. Thus, STIs are highly efficient, occurring under speeded, and shallow processing, and concurrent cognitive load conditions.

These false recognition results also show that STIs are about actors, and not just inferences about the behaviors. (See also Carlston & Skowronski 1994, for evidence from a different paradigm.) That is, STIs represent learned propositional knowledge, acquired relatively automatically (unconsciously, unintentionally, etc.). According to the target article (especially sect. 3), STIs cannot occur. Yet the evidence for spontaneous social inferences is extensive (Uleman et al. 2008).

Automatic associative learning about others also occurs. Suppose you read that Carol said Dan returned the lost wallet with all the money still in it, either for a subsequent memory test or merely to familiarize yourself with such information. When this information is paired with a photo of Carol but not Dan, you are likely to (unintentionally and unconsciously) associate honesty with Carol, the communicator (Skowronski et al. 1998; Todorov & Uleman 2004). Extensive research shows that this spontaneous trait transference (STT) does not occur because you confuse Carol and Dan, or intentionally draw conclusions about Carol. Instead, the activated trait is simply associated with the communicator, affecting subsequent trait ratings and other responses to her. Carlston and Skowronski (2005) proposed that STTs represent mere associations between actors and traits, whereas STIs represent attributional inferences about actors.

Mitchell et al. emphasize the centrality of truth value to propositional knowledge. This suggests that if STI involves inferring propositions and STT does not, then making participants suspicious of the veracity of behavior descriptions paired with actors or communicators should interfere with STI but not with STT. This is exactly what Crawford et al. (2007) found. These results not only support the propositional status of and relevance of truth value to STI, but they also provide another reason to distinguish between propositional and associative learning.

Explicit learning is dissociable from implicit learning about others. This dissociation challenges Mitchell et al.'s broad claims. Rydell et al. (2006) asked participants to form an impression of Bob (so these are not spontaneous inferences). They saw 100 positive (or negative) statements describing Bob; then described how much they liked him; then saw 100 negative (or positive) statements; and then gave their explicit attitudes again. Uninterestingly, the participants liked Bob more (less) after positive (negative) descriptions, and changed their attitudes after reading inconsistent information. More interestingly, they were also subliminally primed on each trial with negative (or positive) words (e.g., love, cancer), and their implicit attitudes toward Bob were measured with a modified Implicit Association Test (IAT; Greenwald et al. 1998) after each series of 100 trials. The valence of the explicit and subliminal information always differed. After receiving positive explicit and negative subliminal information, the participants' explicit attitude was positive and their implicit attitude was negative. The second series of trials reversed these attitudes.

So, two kinds of evaluative learning occurred simultaneously about the same object (Bob): one intentional, conscious, and deliberate; the other, unintentional and unconscious. These dichotomies seem essential for describing the acquisition of explicit and implicit attitudes, even though Mitchell et al. claim that they are only useful for describing perceptual and performance processes.

These comments do not challenge the target article's claim that most, if not all, human *conditioning* is best understood as conscious, intentional, and effortful learning of propositions about the world. But they do challenge the authors' claim that all human learning can be subsumed under the rich and time-honored paradigms of conditioning. These comments also challenge the conflation of conscious, intentional, effortful, and controllable processes with each other and with all propositional learning – and of their "opposites" with each other and associative learning processes. There are varieties of both human propositional *and* associative learning that the target article neglects.

ACKNOWLEDGMENTS
Thanks to students in my lab for suggestions: SoYon Rim, Laura Kressel, and Justin Mahalak.

A one-system theory that is not propositional

doi:10.1017/S0140525X09001174

James E. Witnauer,[a] Gonzalo P. Urcelay,[b] and Ralph R. Miller[a]

[a]*Department of Psychology, State University of New York–Binghamton, Binghamton, NY 13902-6000;* [b]*Department of Experimental Psychology and Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge CB2 3EB, United Kingdom.*

jwitnau1@binghamton.edu **gu203@cam.ac.uk**
rmiller@binghamton.edu
http://www2.binghamton.edu/psychology/
http://www.psychol.cam.ac.uk/

Abstract: We argue that the propositional and link-based approaches to human contingency learning represent different levels of analysis because propositional reasoning requires a basis, which is plausibly provided by a link-based architecture. Moreover, in their attempt to compare two general classes of models (link-based and propositional), Mitchell et al. refer to only two generic models and ignore the large variety of different models within each class.

Mitchell et al. depict propositional and associative approaches to human contingency learning as incompatible with each other. Based on a comparison between generic link-based and propositional models, Mitchell et al. conclude that the propositional approach is superior to the link-based approach. We will argue that: (1) propositional and link-based accounts are not incompatible and are concerned with separate levels of analysis; and (2) Mitchell et al. complicate their analysis by comparing two broad families of models, which has important implications for evaluating these families.

We assert that the propositional and link-based approaches are concerned with different levels of analysis. Note that our argument is different from Mitchell et al.'s argument that specific link-based models speak to two different levels of analysis. The propositional approach argues that humans and animals use propositional reasoning to guide judgments about outcomes in Pavlovian and human contingency learning situations. This approach is silent concerning the cognitive architecture that supports propositional reasoning. In contrast, the link-based approach is concerned with the extent to which one representation can activate another representation and is relatively silent on the way in which animals and humans use associations.

For some of the reasons outlined by Mitchell et al., one might argue that the link-based level of analysis is not useful for understanding behavior. However, we contend that, to the extent that we are able to assess changes in associations through Pavlovian conditioning and human contingency learning, the associative level of analysis is helpful in understanding many aspects of human and animal behavior. Connectionist models have been used to describe phenomena in divergent areas of cognitive psychology. Aside from the obvious examples of Pavlovian conditioning and contingency learning, connectionism has been highly influential throughout cognitive psychology, including perception, categorization, language, memory, attention, social cognition, and cognitive pathology.

Evidence that these approaches are concerned with different levels of analysis comes from the literature concerning connectionist models of language processing. Connectionist models assume that connections (analogous to links or associations) between processing units provide the foundation for complex information processing. In these systems, weighted connections allow activation to pass between units and learning is presumably driven by changes in the strengths of the weights between processing units. Connectionist models of language are link-based models that can represent and process propositional knowledge. A second notable example of propositional logic being based on link-based knowledge is provided by Wynne (1995) in his associative account of transitive inference. Therefore, the existence of several link-based accounts of propositional reasoning suggests that propositional reasoning can be explained at the associative level of analysis by reductionism.

Moreover, associative theories have informed us about the way the brain organizes and processes information. According to the Rescorla and Wagner (1972) model, a discrepancy between the strengths of the outcome experienced and the outcome expected based upon all cues present is necessary for changes in the strength of a CS-US association. The results of electrophysiological and neuroimaging studies suggest that the brain generates a signal that encodes the discrepancy between expected and experienced outcomes and that this signal is correlated with learning at a behavioral level (e.g., Corlett et al. 2004; Schultz 1998). Mitchell et al. aptly note that in many connectionist-like models of cognition (and in the brain), stimulus representations are distributed, meaning that units in these systems do not carry specific representational value and that information is represented by patterns of activity across arrays of units. This does not necessarily undermine the link-based level of analysis because, in systems of distributed representation, weighted links function to bind arrays of units that together represent stimuli. Also, the results of modeling studies suggest that individual neurons are highly variable in the extent to which they locally encode information, such that some neurons function like grandmother cells (i.e., a single neuron more directly represents a stimulus) and others are more broadly tuned (Verhagen & Scott 2004).

Mitchell et al. argue that the link-based approach does not explain behavior as well as the propositional approach. They base this argument on a comparison between a generic associative and a generic propositional model. This strategy has the unfortunate consequence of ignoring the great variety of associative and inferential models available. For example, Mitchell et al. point out that the propositional approach is better equipped than associative models to account for data indicating that awareness is related to learning. However, the prediction that awareness will be related to learning is not a necessary prediction from a propositional model. Bayesian models are similar to the propositional approach but do not assert that awareness is necessary for learning. It is also conceivable that an associative model might argue that awareness is necessary for learning.

Similarly, data that uniquely support a specific associative model cannot be interpreted as inconsistent with the general propositional approach. The sometimes competing retrieval model (SOCR; Stout & Miller 2007) uniquely anticipated that when a target stimulus is conditioned in compound with two blocking cues, responding to the target stimulus is greater than when it is conditioned in compound with only one blocking cue (Witnauer et al. 2008). Despite SOCR being an associative model, these data do not allow us to conclude that the link-based approach is superior to the general propositional approach. In fact, Witnauer et al.'s data were problematic for many associative models and might be consistent with revised propositional models (see Miller & Escobar [2001] for a similar argument concerning ill-conceived comparisons between generic acquisition- and performance-focused models). Differentiation between models requires identification of specific models, and the results are applicable only to the actual models that are compared.

The predictions of a generic model (such as the one outlined by Mitchell et al.) or a family of models (such as acquisition-focused models of Pavlovian learning) are necessarily less precise (and less testable) than the predictions of a specific model (e.g., SOCR). This is evident in Mitchell et al.'s application of their generic model to the effect of cognitive load on learning phenomena. In section 4.2, Mitchell et al. assert that the studies of the effect of cognitive load on learning agree with the predictions of the propositional approach. Propositional reasoning presumably requires cognitive resources and, consistent with this view, manipulations that diminish the availability of cognitive resources diminish learning. In section 4.4, they assert that propositional reasoning can occur in highly complex (demanding) situations. If it is assumed that task complexity is directly related to cognitive load, then the propositional approach (as outlined in sect. 4.2) predicts that learning should not be observed in highly complex situations. This inconsistency (and others) in Mitchell et al.'s propositional approach is the result of comparing generic models of learning rather than fully specified models.

ACKNOWLEDGMENTS

This research was supported by National Institute of Mental Health Grant 33881. Inquiries concerning all aspects of this research should be addressed to Ralph R. Miller, Department of Psychology, SUNY–Binghamton, Binghamton, NY 13902-6000, USA; e-mail: rmiller@binghamton.edu.

Authors' Response

Link-based learning theory creates more problems than it solves

doi:10.1017/S0140525X09001186

Chris J. Mitchell,[a] Jan De Houwer,[b] and Peter F. Lovibond[a]

[a]*School of Psychology, University of New South Wales, Kensington 2052, Australia;* [b]*Department of Psychology, Ghent University, Henn Dunantlaan 2, B-900, Ghent, Belgium.*

chris.mitchell@unsw.edu.au Jan.dehouwer@ugent.be p.lovibond@unsw.edu.au
http://www.psy.unsw.edu.au/profiles/cmitchell.html
http://users.ugent.be/~jdhouwer/
http://www.psy.unsw.edu.au/profiles/plovibond.html

Abstract: In this response, we provide further clarification of the propositional approach to human associative learning. We explain why the empirical evidence favors the propositional approach over a dual-system approach and how the propositional approach is compatible with evolution and neuroscience. Finally, we point out aspects of the propositional approach that need further development and challenge proponents of dual-system models to specify the systems more clearly so that these models can be tested.

In our target article, we put forward the claim that all associative learning in humans is mediated by the non-automatic truth evaluation of propositions about relations in the world. This implies that there is no need to postulate a second learning mechanism that is based on the unconscious formation of links between symbolic representations. One positive outcome of our target article was that several commentators expressed their agreement with the basic tenet of our position (**Beckers & Vervliet**; **Chater**; **Morsella, Riddle, & Bargh** [**Morsella et al.**]; **Gopnik**; **Greenwood**; **Lagnado**; **Li**; **Lyn & Rumbaugh**; and **Newell**). However, it is clear that we were unable to convince everybody. Many commentators explicitly expressed their belief in a dual-system model of learning in which both propositional and link-formation processes can produce learning (**Baeyens, Vansteenwegen, & Hermans** [**Baeyens et al.**]; **Boakes**; **Dawson & Schell**; **Dwyer, Le Pelley, George, Haselgrove, & Honey** [**Dwyer et al.**]; **Gawronski & Bodenhausen**; **Livesey & Harris**; **Matute & Vadillo**; **McLaren**; **Miles, Proctor, & Capaldi** [**Miles et al.**]; **Olsson & Öhman**; **Schultheis & Lachnit**; **Uleman**). This, too, is a positive outcome of our target article, because it confirms that we are not fighting a straw man, but a widely held position in psychology. Few researchers had previously made explicit in writing their belief in a dual-system model of learning.

The arguments made by the commentators for rejecting the propositional approach can be grouped into three categories: (1) the evidence that supports the propositional approach can also be explained by link-formation models; (2) there are empirical findings that contradict the propositional approach and thus support the idea of a second, link-based, learning mechanism; and (3) there are conceptual grounds for rejecting the possibility that all learning is propositional. Some of the arguments presented in the commentaries are based on an interpretation of the existing empirical evidence with which we disagree. Others seem to reflect a misunderstanding of the propositional approach, which suggests that we were not completely clear in the target article. From our analysis, we conclude that although the current propositional approach needs further refinement, the basic assumption that all associative learning is mediated by propositions is conceptually sound and is supported by the available data.

R1. Can evidence for the propositional approach also be explained by link-formation models?

In section 3 of the target article, a large number of empirical findings were presented that confirm predictions of the propositional approach. We first describe exactly why the propositional approach predicts these findings, and then look at the extent to which these findings can be explained by link-formation models. The core assumption of the propositional approach is that associative learning depends on the non-automatic truth evaluation of propositions about relations in the world. Some commentaries revealed a misunderstanding of what we meant by "propositions" and how propositions differ from associative links. Propositions are statements about the events in the world and how they are related. These statements may be quite specific as to the nature of the relation between events (e.g., "*X* causes *Y*"). However, they can be quite general; they may specify only that particular events are related, but not how they are related. We agree with **Gawronski & Bodenhausen** that the crucial distinction between propositions and links is that only propositions imply a truth value. People can thus differ in the degree to which they believe that statements about the world are true or false. Although propositions are described as statements, they do not, as **Shanks** seems to imply, necessarily involve abstract representations. As Barsalou (2008) pointed out, propositions can also involve embodied, grounded representations. Propositions are very different from associative links. Associative links are not statements about the world. They are assumed to be states of the world, namely, states through which one representation can activate another representation.

The idea that associative learning is a function of the non-automatic truth evaluation of propositions leads to two main predictions. First, everything that influences truth evaluation can influence associative learning. This means that apart from the actual occurrence of events, factors such as prior knowledge, instructions, intervention, and deductive reasoning also matter. Second, because truth evaluation depends on non-automatic processes (in the sense that they are dependent on awareness, cognitive resources, time, and goals), associative learning should also be non-automatic. There is substantial evidence to support these predictions (see our target article and De Houwer [2009] for reviews). Because of this evidence, many commentators accept the conclusion that some associative learning phenomena are the result of propositional processes (e.g., **Baeyens et al.**; **Boakes**). However, they do not agree with the more far-reaching conclusion that all learning is propositional. They argue that there are some "true" forms of learning that are driven by link formation. Although it is not clear to us what the defining characteristic of "true" learning is, we do agree that evidence in support of the propositional approach is also compatible with dual-system models. Thus, such findings do not exclude the possibility that link formation exists as a second learning mechanism that supplements propositional learning. The reasons for why many commentators maintain a belief in the existence of link formation will be analyzed in the next section. In the remainder of this section, we will discuss a second response to the evidence in support of the propositional approach, namely, that link formation models can also account for (some of) that evidence.

A first group of commentators (**Dwyer et al.**; **Livesey & Harris**; **Witnauer, Urcelay, & Miller** [**Witnauer et al.**]; see also section 2.1 of the target article) points out that link formation is not necessarily a fully automatic process. For instance, influential link-formation models do postulate that attention modulates the formation of links (e.g., Mackintosh 1975; Pearce & Hall 1980). One could also argue that link formation leads to awareness of relations in the world (e.g., Davey 1992). Hence, evidence for the non-automatic nature of associative learning could in principle be accommodated by link-formation models by adding assumptions about the way in which link formation is non-automatic. However, one of the important reasons why many researchers believe in the existence of a link-formation mechanism is precisely because it might account for seemingly automatic, irrational types of learning. This is probably why many commentators explicitly characterize link formation as primitive and automatic. Another problem with the argument that link formation might be non-automatic is that there is no a priori reason for this assumption. There are, however, good reasons to assume that the truth evaluation of propositions is a largely non-automatic process. Hence, the failure to observe automatic learning is a prediction of, and therefore supports, the propositional approach.

Other commentators reject the dual-system idea, but argue that the single system is based on link formation. They suggest that the evidence for the propositional approach can be accounted for by (link-based) connectionist models of learning. In this way, propositional processes can be reduced to link formation. We did foresee this argument in our target paper when we discussed the nature of connectionist models (sect. 6.2). Our main point was that connectionist models differ from link-formation models because they involve sub-symbolic, rather than symbolic, representations. The nodes in a connectionist model do represent information (as **Penn, Cheng, Holyoak, Hummel, & Povinelli** [**Penn et al.**] point out), but they do not represent a discrete stimulus or event in the world (as they do in link-formation models). In other words, the information represented by symbolic nodes (representations) is specified by the model, but the information represented by sub-symbolic nodes (in connectionist networks) needs to be inferred from how the system operates. Our quarrel is only with symbolic link-formation models. We are aware of the fact that the brain resembles a sub-symbolic connectionist network. In fact, connectionist models can be regarded as simulated miniature brains (Clark 1990). It is also obvious that the truth evaluation of propositions must somehow be implemented in the brain. Hence, we do not exclude the possibility that truth evaluation might be implemented in a connectionist model (although current models still seem to have a long way to go before this promise can be realized). Such an implementation would not, however, constitute evidence that propositional processes can be reduced to symbolic link formation or that there is a need to assume the existence of symbolic link formation.

R2. Are there empirical findings that contradict the propositional approach?

A second category of arguments focuses on evidence for associative learning that cannot be explained by propositional models. The evidence within this category can itself be divided into three classes. The first class concerns evidence for associative learning in the absence of relevant propositions (e.g., learning without awareness). Second, several commentators refer to learned behavior that is arational in that it cannot be inferred logically from the propositions that people entertain. They claim that the propositional approach does not provide an explanation for the production of Pavlovian conditioned responses (CRs). The third class includes evidence for arational beliefs about relations in the world. Some beliefs seem to result from factors other than truth evaluation and, it is argued therefore, constitute evidence for the formation of links. In this section, we explain why we are not convinced by these arguments.

R2.1. Evidence for learning in the absence of relevant propositions

A number of commentators cited studies that they considered to be demonstrations of unconscious learning (**Boakes**; **Dawson & Schell**; **Dwyer et al.**; **Gawrosnki & Bodenhausen**; **Livesey & Harris**; **McLaren**; **Miles et al.**; **Morsella et al.**; **Olsson & Öhman**; **Penn et al.**; **Uleman**). The majority of these studies have already been considered in the several reviews we cited in our target article, and shown to be artefactual, unreplicable, or subject to alternative interpretations. In many cases, direct supporting evidence for specific alternative interpretations has been reported. Other studies nominated by the commentators have been published subsequent to these reviews, but have used procedures subject to the same criticisms as the earlier studies. The commentary process provides a unique insight into the results that are considered to provide the most robust evidence for unconscious learning by dual-system theorists; therefore, we consider it important to briefly review the main difficulties with these studies and provide references with additional detail for the interested reader.

Before turning to the specific research procedures, it is important to first emphasize one general point about this line of inquiry. Apparent unaware conditioning can be generated simply by the use of an insensitive or noisy measure of awareness, because some aware participants will be misclassified as being unaware. As a consequence, studies that successfully show a relationship between awareness and conditioning should be given greater weight than those that fail to find such a relationship, unless there is independent evidence that the null result was not due to lack of sensitivity of the awareness measure (Lovibond et al., submitted).

Studies involving fear conditioning to backwardly masked (or subliminal) fear-relevant stimuli (cited by **Dawson & Schell**, **Olsson & Öhman,** and **Penn et al.**) suffer from three primary limitations. The first two are insensitive measures of awareness and the use of fixed stimulus durations which are suprathreshold for some participants. The third is a little more subtle. Sometimes the measurement of awareness relies on report of stimulus identity. However, participants who cannot identify the stimuli may nevertheless be aware that CS+ and CS- differ in some way – this is all that is required for differential conditioning to be observed (Cornwell et al. 2007; Lovibond & Shanks 2002; Pessoa et al. 2005). In addition, several early studies have since been shown to be an artifact of the experimental procedure; they could be explained by a specific correlated hypothesis account based on restrictions on trial order (Wiens et al. 2003). Finally, studies using a more sensitive measure of awareness, such as a trial-by-trial measure of subjective US expectancy, have shown a concordance between this measure and objective measures such as skin conductance (e.g., Öhman & Soares 1998). These findings directly support a single-system account. The same is true for the resistance to extinction observed with supraliminal fear-relevant stimuli (Dawson et al. 1986; Lovibond et al. 1993).

Studies involving anesthesia and spinal preparations (cited by **Dwyer et al.** and **Miles et al.**) have mostly been conducted in animals, so implications for the role of awareness are hard to draw. By contrast, there has been a great deal of research on learning in humans suffering from amnesia. Early reports of conditioning in amnesics (cited by Miles et al.) were largely anecdotal, with little detail regarding the nature of the assessment of conscious knowledge. Those studies which have been more fully reported (e.g., Bechara et al. 1995; Clark & Squire 1998) have often used a post-experimental assessment of awareness, which is particularly likely to underestimate contingency knowledge in amnesics at the time of conditioning (Lovibond & Shanks 2002). A recent study by Speekenbrink et al. (2008), however, was specifically designed to assess the relationship between task performance and relevant explicit knowledge *during* learning, in a probabilistic category learning task. They found a slightly slower learning rate in amnesics compared to control participants. But, importantly, there was no evidence for a qualitatively different learning process and no dissociation between task performance and explicit knowledge of cue-outcome association.

As noted by **Miles et al.** and **Olsson & Öhman**, there have been several reports that eyeblink conditioning with a delay conditioning procedure (in which CS and US overlap) is unrelated to awareness, unlike conditioning with a trace procedure (in which CS and US are separated by a brief interval; see **Li**'s commentary). However, there have been specific criticisms of some of these studies (Lovibond & Shanks 2002; Shanks & Lovibond 2002). Furthermore, other researchers have successfully demonstrated a relationship between awareness and differential eyeblink conditioning (with no evidence of conditioning in unaware participants), regardless of whether a trace or delay procedure is used (Knuttinen et al. 2001; Lovibond et al, submitted).

The use of startle modulation as an index of conditioning (cited by **Olsson & Öhman**) has not, in general, yielded evidence of unconscious conditioning. The early study by Hamm and Vaitl (1996) is limited by the fact that contingency awareness was assessed after an extinction period (Lovibond & Shanks 2002), and more recent studies by Weike et al. (2007) have been criticized by Dawson et al. (2007) as using an insensitive recall-based awareness measure. More importantly, however, several studies have shown a reliable relationship between contingency awareness and startle modulation, with no evidence of modulation in the unaware group (Dawson et al. 2007; Lipp & Purkis 2005; Purkis & Lipp 2001).

There have been numerous reviews of evaluative conditioning (cited by **Baeyens et al.**), and specifically the claim that such conditioning can occur without contingency awareness (see also **Bliss-Moreau & Barrett**). As noted in our target article, more recent studies have started to provide clearer evidence of an association between contingency awareness and evaluative conditioning, both for the picture-picture procedure (Pleyers et al. 2007) and the flavor-flavor procedure (Wardle et al. 2007). Other commentators (**Boakes**; **Livesey & Harris**) focused on the more general point that conditioning with tastes, odors, and internal bodily consequences, whether considered as an example of evaluative conditioning or not, may be independent of cognitive factors such as instruction and awareness. Again, much of this evidence in humans is anecdotal. Curiously, Boakes also cites placebo effects in this context, but of course this is an instructional effect that suggests bodily reactions are in fact sensitive to cognitive factors. One exception, as noted by Lovibond and Shanks (2002), is the odor-taste learning demonstrated by Stevenson et al. (1998). These findings are suggestive of an associative process that is independent of explicit contingency knowledge, and this avenue is worthy of further investigation to see if it supports the idea of a separate gustatory learning module isolated from cognitive processes.

Finally, many commentators (**Gawronski & Bodenhausen**; **Miles et al.**; **Morsella et al.**; **Penn et al.**; **Uleman**) raised studies from the cognitive and social literatures that do not involve direct recording of conditioned responses but that are, nonetheless, clearly associative in nature. Many of these studies report effects that have been labeled implicit learning, such as sequence learning, motor learning, and artificial grammar learning. The most comprehensive review of this literature was carried out by Shanks and St. John (1994), who reported that virtually all of the studies of implicit learning at that time failed either their sensitivity or their informational criterion, or both. Subsequent research has strengthened and extended Shanks and St. John's conclusion and extended it to other claims of implicit learning, including sequence learning (Wilkinson & Shanks 2004), continuous tracking (Chambaron et al. 2006), probabilistic category learning (Newell et al. 2007), and contextual cueing (Smyth & Shanks 2008).

An alternative approach to demonstrating non-propositional learning, taken by Perruchet (1985) and Perruchet et al. (2006), is to show effects of learning that directly contradict relevant propositions (see sect. 5.1 of the target article). Such evidence would be much more difficult to discount than the evidence for unaware learning, because demonstrating the presence of a proposition that contradicts the behavior is easier than demonstrating the absence of propositions altogether.

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

Some commentators argued that the Perruchet effect alone is sufficient evidence to postulate the existence of link formation as a second learning mechanism (**Dwyer et al.**). We believe, however, that the evidence is not yet conclusive. For instance, recent findings suggest that at least the reaction time version of the Perruchet effect does not result from the operation of two associative learning systems. Mitchell et al. (in revision) presented half of their participants with the same 50% tone-square partial reinforcement schedule used by Perruchet et al (2006). The remaining participants received the same schedule but with the tones removed. In their experiment, just as in the one Perruchet et al. (2006) conducted, reaction times were fastest following runs of square-present trials, and slowest following runs of square-absent trials. This was true for all participants, regardless of whether or not they had been presented with the tone (CS). This suggests that the pattern of reaction times does not indicate the operation of an associative learning process; it is simply the consequence of the recent square presentations. Because the behavioral effect is not due to the relation between the tone and the target event, there is no longer a contradiction between propositional beliefs about the tone-square relation and a behavioral effect of that relation. Therefore, the dissociation between RTs and expectancy cannot indicate the operation of two distinct associative learning systems.

The research on the Perruchet effect shows that the propositional approach can be falsified in principle. However, it also illustrates how one should be careful in drawing conclusions prematurely. Many years of research on learning in the absence of relevant propositional knowledge has led to few (if any) convincing findings. This calls for a cautious approach toward any new piece of evidence that claims to demonstrate learning in the absence of relevant propositional knowledge.

R2.2. Learning can result in arational behavior

Some commentators (e.g., **Baeyens et al.; Livesey & Harris**) point out that when participants do have relevant propositional knowledge about relations in the world, one cannot explain, on the basis of this knowledge, why those relations have the behavioral effect that they do. For instance, why does the proposition "the tone predicts the shock" lead to an increase in skin conductance (the CR) after the presentation of the tone? Why would a proposition such as "this flavor was sometimes followed by a bad aftertaste" lead to a disliking of the flavor when presented on its own? In the target article, we acknowledged that the propositional approach does not explain why propositions about relations in the world have certain effects but do not have others. This is because propositional models are not models of behavior. They are models of one determinant of behavior: associative learning. There are many other factors that go to determine the way in which learning translates to performance.

This line of reasoning may give the impression that the propositional approach is unfalsifiable; any behavior that does not seem to follow logically from the propositions entertained can be attributed to some performance factor or other. However, although we do not have a psychological model of this translation process, we do know quite a lot about the effects that particular propositions will have, and we do assume some stability in this process. For example, as **Olsson & Öhman** concede, we know that (all else being equal) the proposition "the tone signals shock" will lead to an expectancy of shock when the tone is presented, and this expectancy will generate the CR of increased skin conductance (see sect. 3.2 of the target article). At some point during evolution, it would appear that the cognitive expectancy of negative events became a trigger for the activation of genetically determined defensive response patterns that include an increase in skin conductance. We can, given that we know this process is in place, make predictions about the effects of propositions on behavior.

It is very important to note that, in fact, the link-formation models do not explain CR production any better than does the propositional approach. Most link-formation models remain silent about how links are expressed in behavior. Furthermore, the psychological mechanism provided by the link model for the production of CRs relies on the old idea that conditioned responses are simply indirectly activated unconditioned responses. Thus, for instance, when a tone is paired with a shock, it might evoke an increase in skin conductance because activation caused by the presentation of the tone spreads to a representation of the shock, which then activates the responses associated with shock. But, as we pointed out in our target article (see sect. 2.2), conditioned responses often differ substantially from unconditioned responses. And these differences cannot always be explained in terms of the nature of the CS. Link-formation models also do not explain why learning can be selective. For instance, why does a flavor that is paired with a bad aftertaste become negative, whereas a color paired with the same aftertaste does not (see comment of **Baeyens et al.**)? In sum, neither propositional models nor link-formation models provide a full account of learned behavioral responses. Hence, we see no reason why evidence for arational learned behavior should favor link-formation models.

R2.3. People sometimes have arational beliefs about relations in the world

The studies discussed in sections R2.1 and R2.2 of this rejoinder can be seen as attempts to discount propositional models of learning by showing behavioral effects of learning that cannot be due to relevant propositions, either because relevant propositions are absent (sect. 2.1) or because the behavioral effects of learning cannot be inferred logically from the relevant propositions (sect. 2.2). Some commentators (**Dwyer et al.**; **McLaren**) discuss a third set of studies that followed a different approach. They suggest that some propositional beliefs about relations in the world are arational, and, therefore, might result from the operation of a link-formation mechanism. The evidence presented in the target article on this issue (sect. 4.3.2) came from the studies of Le Pelley et al. (2005a) on unblocking, Karazinov and Boakes (2007) on second-order conditioning (see **Uleman** for very similar studies in the context of social psychology), and from Shanks's (2007) example of the judgment of event probabilities. Additional studies of this kind are presented in the commentaries (e.g., McLaren cites Spiegel & McLaren's [2006] work on sequence learning).

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

As we made clear in our target article (see sect. 5.2), evidence for seemingly arational beliefs about relations in the world does not convince us of the need for a separate link-formation mechanism. First, there is a very important general point that, even if certain findings cannot be explained by propositional models, this is not evidence for the existence of link-formation processes. Many other explanations are also possible. Second, detailed research is needed before one can exclude with confidence propositional processes as a source of the observed phenomenon. Human reasoning is not necessarily normatively correct. Hence, demonstrations of arational or irrational beliefs do not fall outside of the scope of truth evaluation on the basis of reasoning. Thus, we see no need to evoke the idea of link formation in order to account for such beliefs.

We do acknowledge that these concerns complicate attempts to provide evidence in support of the link-formation mechanism. Hence, we understand the complaint of certain commentators that the propositional approach runs the risk of becoming unfalsifiable (**McLaren**; **Nolan**). However, those who want to argue for the existence of link-formation processes should examine carefully all possible explanations that do not rely on link formation before they conclude that link-formation processes do operate. It is unlikely that such issues can be settled by isolated demonstrations of effects that at first sight appear to contradict the propositional approach, or (as **Witnauer et al.** point out) seem consistent with a particular associative model (see McLaren; **Dwyer et al**.). The burden of proof in the first instance is on those who want to claim the existence of link-formation processes to rule out all other possibilities. That being said, we do accept that once efforts have been made to rule out these other possibilities, proponents of propositional models cannot simply dismiss potentially problematic findings on the basis of the argument that *some kind of* propositional process *might* be crucial. They should always back up their alternative explanations with empirical evidence or sound arguments (see the work of Mitchell et al. [in revision] on the Perruchet effect). If they cannot, they should accept the conclusion that some findings fall beyond the scope of the propositional approach.

R3. Are there conceptual arguments for rejecting the propositional approach?

The evidence discussed in section R2 of this response relates to possible inconsistencies between the propositional approach and empirical knowledge about associative learning. In this section, we discuss comments that relate to possible inconsistencies between the propositional approach and our knowledge about evolution, the brain, and normative criteria that good theories should meet.

R3.1. Propositional models are not in line with the principles of evolution

Some commentators (**Miles et al.**; **Schultheis & Lachnit**) suggested that the propositional model of human learning requires postulation of a "magical cut-point" in evolution, when propositional learning appeared and reflexive learning disappeared. Similarly, these and other commentators (**Matute & Vadillo**; **Olsson & Öhman**; **Penn et al.**) asked why evolution would simply abandon an effective and adaptive reflexive mechanism. However, in our target article, we explicitly argued for a "continuum of cognitive complexity" (sect. 6.3) between primordial creatures, on the one hand, and humans, on the other hand. We suggested that early representational structures were likely S-R in nature, that these were elaborated to allow S-S structures that separated knowledge from action, and that further quantitative elaboration of knowledge structures provided the necessary foundation for abstract reasoning and artificial symbolic systems (language). This continuity position is entirely consistent with the known principles of evolution, which as noted by Olsson and Öhman is "a slow accumulative process . . . that incorporates successful adaptations at one level into more complex functions at higher levels." Therefore, our answer to the question of what happened to primitive reflexive mechanisms is not that they shriveled up and died (Penn et al.), but rather that they evolved into the rich representational system that humans possess. As **Morsella et al.** note in their commentary, this system is suitable for the solution of both complex and simple problems.

Furthermore, contrary to the claims of several commentators (e.g., **Hall; Matute & Vadillo**), the removal of a primitive learning mechanism during the course of evolution might well have been adaptive. A primitive mechanism that creates links whenever certain inputs co-occur might be adaptive for simple organisms that register only limited and simple inputs but maladaptive for more complex organisms that receive a lot of complex input. In more complex organisms, an automatic link system would lead to an overload of associations that would result in chaotic behavior (as in the case of schizophrenia; see Carson et al. 2003). This problem can be managed if propositional encoding needs to occur before a relation can be coded in memory and influence behavior. Hence, in addition to the benefits offered by a more complex, propositionally based learning mechanism, reducing the negative effects of a primitive learning mechanism provides a second evolutionary reason for the gradual evolution from primitive to more complex learning mechanisms.

Ironically, it is the dual-system model that suggests an implausible discontinuity in evolution, because it denies that earlier reflexive systems formed the precursor to representational abilities and, subsequently, to human reasoning and language. Instead, proponents of this model generally assume that reasoning emerged independently and late in mammalian evolution, for example, in primates. In the extreme case, those who propose that reasoning relies upon language (e.g., **Castro & Wasserman**) are forced to argue that reasoning is unique to humans (and perhaps, to a very limited extent, certain primates) and therefore emerged de novo at "one minute to midnight" in biological time. But emergence of a novel capacity in such a short time-frame is entirely implausible given the accumulative and opportunistic nature of evolution, which is much more likely to appropriate existing mechanisms to new purposes rather than develop a new one from scratch. We agree with Christiansen and Chater (2008) that it is more plausible to suppose that language was enabled by the prior evolution of a critical level of

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

cognitive capacity, and that language was, therefore, moulded by the operating characteristics of the cognitive system.

The only alternative for dual-system theorists is to concede that propositional reasoning, or the representational capacity that preceded it, developed well before the evolution of humans. But this strategy raises another set of problems for dual-system theorists. They must now confront seriously the role that the nascent cognitive system might play in learning and other tasks in the animal laboratory. According to the propositional model, this is exactly what animal researchers have been studying successfully for many years, and indeed they sometimes describe their work as "animal cognition." But from a dual-system perspective, it would now be necessary to force a distinction between cognitive processes and co-existing reflexive or link-formation processes. We have already reviewed the vanishingly small evidence base for such a distinction in humans, where some predictions of the dual-system model can be tested through verbal report. It seems unlikely that independent evidence will emerge from the animal laboratory that clearly supports a dual-system architecture. Rather, we consider that the learning capacities of particular species studied in the laboratory will reflect the nature and complexity of their cognitive system. Species more closely related to humans (e.g., other mammals and primates, in particular) will have cognitive systems and hence learning capacities correspondingly more similar to our own (see **Lyn & Rumbaugh**).

It is through this lens that we approach the comments of **Miles et al.** and **Schultheis & Lachnit** concerning the learning abilities of honeybees. We suggest that during human evolution there was a continuum of development of cognitive capacity from simple organisms to modern humans. However, not all of these earlier species survive today, and of those that are extant, many are on branches of the evolutionary tree that diverged quite early from human evolution. Therefore, we are not able to make strong predictions as to the cognitive abilities of a particular species. We would not be surprised if a species such as the honeybee showed some characteristics indicative of capacity to represent the environment, such as cognitive maps, but we would also not be surprised to find strong limitations to such capacities. It is also likely that species with substantially different nervous systems will have solved the same problem in different ways, due to convergent evolution. For example, different species may show blocking on the basis of different learning mechanisms, just as they achieve other capacities such as locomotion through different mechanisms (e.g., flying, swimming, walking).

Some commentators made the case for preservation of a reflexive learning system on the basis that this system is fast and automatic (**Matute & Vadillo**), and that evolution favors the concurrent existence of multiple, complementary systems with different strengths (**Hall**; **Olsson & Öhman**; **Schultheis & Lachnit**). The implication of this argument is that the cognitive system is slow and effortful and hence will be out-performed by a reflexive system in situations requiring rapid reactions. However, it is not at all clear that a reflexive system, if it existed in humans, would in fact be faster in such situations. This is because situations requiring rapid responses (e.g., predator avoidance) are ones that depend on *performance* based on prior learning, not on new learning per se. As we noted in our target article, performance based on existing propositional knowledge is fast and automatic – regardless of whether the knowledge is acquired through direct contingency exposure or is learned symbolically. Furthermore, when it comes to acquisition of new learning, it is debatable whether reflexive learning would be faster than propositional learning. Many S-R theories propose a relatively slow and gradual process of establishing and strengthening of links, in contrast to the rapid "insight" learning associated with propositional knowledge. What is clear, however, is that there would be considerable selection pressure to favor the ability to represent the environment in richer ways and to put together pieces of existing knowledge to draw new conclusions – precisely the characteristics we ascribe to the human learning system.

R3.2. Propositional models are not in line with what is know about the brain

Several commentators considered that we had ignored or understated the importance of biological data, in particular brain data. It is true that the primary conclusions we reached were based on behavioral data. However, it is not the case that we ignored brain data. Rather, we argued that at our present state of neuroscientific understanding, the brain data are not definitive with regard to the central debate between single- and dual-system models. That is, the available brain data are to some extent open to interpretation. Of course, neuroscientists have had much more experience in fitting their data to the prevailing dual-system model of learning, so it is perhaps not surprising that several critiques were based on long-standing assumptions regarding the mapping between brain structures and psychological mechanisms.

The most common assumption implicit in the present commentaries was that consciousness and higher cognitive functions are exclusively mediated by the cerebral cortex (**Hall**; **Miles et al.**; **Olsson & Öhman**). Such a "corticocentric perspective" has been challenged recently in this journal by Merker (2007). He reviews evidence that natural or experimental loss of cortical tissue does not eliminate conscious and goal-directed behavior, and points out that mid-brain structures conform more closely than cortical structures to the limited-capacity, executive role associated with reasoning and decision making. In general terms, the graceful degradation of performance that is observed with increasing damage to brain tissue is strongly suggestive of distribution, rather than localization, of function. And, as **Chater** notes in his commentary, many dissociations observed in lesion and activation studies are not inconsistent with a single-processing system. There is certainly little reason to believe that the functions of cortical and subcortical structures map neatly onto the cognitive and reflexive systems proposed in dual-system models of learning. Our single-system model makes no specific claims regarding particular brain mappings. What our model does predict, however, is that those neural systems underlying reasoning (whether they are organized on a topological, chemical, or some other basis) also subserve associative learning. None of the data put forward by the commentators, either from lesion studies in animals or neuropsychological case studies in humans, contradicts this prediction.

R3.3. Propositional models do not meet the normative criteria that good models should meet

Some commentators argue that good models should be falsifiable, parsimonious, and formalized and that propositional models do not possess these characteristics. In this section, we explain why we believe that propositional models are falsifiable and parsimonious. They lack formalization, but that does not impede their heuristic and predictive value.

R3.3.1. Propositional models are not falsifiable. As stated in section 2.1 of this response, reliable demonstrations of associative learning in the absence of relevant propositions (i.e., propositions about the relations in the world that drive learning) would raise serious doubts about the validity of the propositional approach (but would not necessarily provide evidence for link formation). We noted that such demonstrations are only convincing if sensitive measures are used to probe all relevant propositional knowledge that participants might have. Most past research does not satisfy this criterion. A more powerful approach, taken by Perruchet (1985) is to demonstrate that participants have relevant propositional knowledge that contradicts the observed learning effects. The Perruchet effect, if it is demonstrated to be an effect of associative learning (and not due to a performance effect or some other factor) could not be accommodated within the propositional approach.

Tests of the propositional approach must also satisfy one other requirement. As pointed out in section 3.1 of the target article, because propositions can be stored in memory after their truth has been evaluated, they can influence behavior as the result of memory activation even at times when their truth is not evaluated. Hence, in order to demonstrate learning in the absence of relevant propositional knowledge, one should take into account not only propositions that are truth evaluated at the time the learned behavior is emitted, but also those evaluated earlier in training, before the behavior is observed. **Baker, Baetu, & Murphy** [**Baker et al.**] argue that, because propositions can be coded in memory and can then influence later behavior in an automatic way, the propositional approach is rendered unfalsifiable. However, carefully controlled experiments can be set up in which new relations are introduced and propositions about those relations are tested online, at regular intervals during the experiment. The available evidence suggests that learned behavior will be observed only after participants report propositions about the relations that were implemented (e.g., Dawson & Biferno 1973). When retrospective rather than online measures of propositional knowledge are used, researchers should make sure that the measures are sensitive enough to pick up knowledge that participants might have held in the past. This can be done by minimizing the delay between learning and testing.

Finally, in contrast to **Livesey & Harris**, we do believe that research on learning in the absence of relevant propositional knowledge (e.g. outside of awareness) is crucial in testing the propositional account. If convincing evidence for this type of learning can be found, it would raise serious doubts about models in which propositional processes are the only ones that can support associative learning.

R3.3.2. Propositional models are not parsimonious. In our target article (sect. 6.1), we argued that the propositional approach to associative learning is more parsimonious than a dual-process approach because "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, some commentators have still suggested that the propositional approach is not parsimonious.

Dwyer et al. argue that the propositional approach lacks parsimony because it needs additional systems in order to account for phenomena such as habituation and perceptual learning. We agree that, within the propositional approach, learning cannot take place by propositional processes alone, but also requires perception and memory. However, a very similar argument seems also to apply to link models. There may still be some researchers who assume that link formation can provide a full account of learning or behavior. It does seem more realistic, though, to assume that link models, like propositional models, need to be extended by other processes in order to explain (non-associative influences on) behavior. These necessary additional processes of perception and memory provide mechanisms for the phenomena referred to by Dwyer et al.

Livesey & Harris argue that a propositional approach to learning is not parsimonious because propositional processes are effortful. A system in which all learning depends on effortful processes might end up using more resources than a system in which learning is sometimes outsourced to an effortless link-formation process. There are, however, several counterarguments to this position. First, as we pointed out earlier in this response (see sect. R3.1), having a primitive, effortless link-formation process that registers relations in the world could lead to an overload of information, resulting in chaotic behavior. It could also lead to action tendencies that are contrary to those arising from the cognitive system. The propositional encoding of relations in the world could provide the necessary buffer against such chaotic and hence wasteful behavior. Second, one can overestimate the expenditure of resources that propositional learning processes would require. Although truth evaluation is an effortful process, it can also be conducted in a "quick and dirty" manner. When there is little opportunity or motivation, not all information that is relevant for truth evaluation will be taken into account, thus minimizing the resources that are used to evaluate propositions about relations in the world. Moreover, once a truth evaluation has been completed, the resulting proposition can be stored in memory. On later occasions, this proposition can influence behavior without being evaluated again. In sum, propositional processes could provide the right balance between learning what needs to be learned and expenditure of cognitive resources.

R3.3.3. Propositional models are not formalized. There are two reasons why we do not mind the current lack of formalization of propositional models. First, we do not

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

believe that formalization is a normative criterion that all models should meet. Second, those who prefer formalized models are free to formalize the aspects of propositional models that can be formalized. With regard to the first argument, the quality of a model cannot be judged by its degree of formalization. It is easy to imagine a fully formalized model that is unable to account for, or correctly predict, any instance of associative learning. The quality of a model is determined by its heuristic and predictive value, that is, its ability to (a) account for and organize existing knowledge and (b) to correctly predict new empirical results. Formalization can help achieve these aims, but it is neither necessary nor sufficient.

Witnauer et al. argue that, because propositional models are not formalized, precise predictions are difficult to derive from propositional models. In response, we would like to point out that formalized (link) models of associative learning also do not necessarily allow for precise, unequivocal predictions. Most of these models include a variety of free parameters. Various predictions can be derived from a model by varying the value of parameters or by adding new parameters.

Despite these arguments, we acknowledge that formalization can also have benefits such as making explicit the assumptions and inferences that are made when accounting for or predicting a certain finding. It is therefore important to realize that various aspects of propositional models can and have been formalized. Formalization can be coined both in terms of logical operations (e.g., the modus tolens inference; see Beckers et al. 2005; Mitchell & Lovibond 2002) or mathematical expressions. As we pointed out in our target article, mathematical models of link formation such as the Rescorla-Wagner model can be regarded as mathematical formalizations of the operating principles of propositional processes. As **Beckers & Vervliet** correctly argue, such mathematical models have a heuristic and predictive value. What is important to realize, however, is that these models retain their value even when it is assumed that they describe propositional processes rather than link formation (see sect. 6.1 of the target article).

Other ways to formalize propositional models is by means of connectionist models and Bayes nets. It should be noted, however, that at present none of the available avenues for formalization is able to capture all aspects of propositional processes. For instance, as **Gopnik** correctly points out, Bayes nets and Baysian inferences do not model the fact that truth evaluation is a non-automatic process. Some commentators (e.g., **Baker et al.**; **Schmajuck & Kutlu**; **Shanks**; **Witnauer et al.**) are optimistic that future connectionist models will be able to mimic all aspects of propositional processes, but a brief look at the current status of connectionist models suggests that much progress still needs to be made.

R4. Conclusion

In this rejoinder paper, we have attempted to address the main arguments that the commentators have raised against propositional models. In our opinion, the assumption that all associative learning is mediated by propositional processes is supported by the great majority of the available data, and is not significantly undermined by the conceptual arguments that have been raised against it. We recognize that we have not responded to all of the points made in the commentaries. The diversity of views expressed, and the different disciplines from which they come (philosophy, cognitive psychology, animal learning, social psychology, and neuroscience), are testament to the central importance of learning to understanding human and animal behavior. But this diversity also means that many very interesting comments and suggestions are beyond the scope of the present response. We do believe, however, that we have addressed the main arguments raised against the central claims made in the target article, that learning is propositional in nature and there is little reason to postulate the existence of an additional link-formation mechanism.

We also acknowledge that many aspects of the propositional approach require further development. As **Newell** and **Sternberg & McClelland** point out (see also **Mandler**), the way in which propositional processes interact with perception and memory, and whether memory processes can as such support associative learning, needs clarification. Evidence that memory processes can produce learning independently would undermine the propositional approach. However, such evidence would not constitute evidence for a link-formation mechanism. There are also other aspects of the propositional approach that need refinement. We agree with Sternberg & McClelland that research on human learning would benefit greatly from integration with reasoning research.

Just as was the case with the target article, we anticipate that this response article will not have persuaded everyone. We would like to finish, therefore, by outlining what we see as some significant future challenges to the dual-system view. It is clear from the commentaries provided here that the postulated link-formation mechanism has a variety of forms. At one end of the spectrum, it is a simple S-R system that we share with *Aplysia* (**Hall**; **Matute & Vadillo**; **Penn et al.**; **Schultheis & Lachnit**). At the other end of the spectrum is quite a different mechanism, which is affected by attention and seems to be comparatively rich and complex (e.g., **Dwyer et al.**; **Livesey & Harris**; **McLaren**). It is sometimes advantageous to propose a simple mechanism (often on grounds of parsimony). At other times it is advantageous to propose a more complex mechanism (perhaps because the data demand it). But the system in which the links are formed cannot have the characteristics of both the *Aplysia*-like S-R mechanism and those of the more complex, attention-demanding, S-S mechanism – these characteristics are incompatible. Proponents of the dual-system view must, therefore, decide on the nature of the link-formation mechanism; otherwise, they may be accused of wanting to "have their cake and eat it too."

Those who propose the *Aplysia*-like system must then explain, for example, why learning does not occur outside of awareness. A different problem arises for those who opt for the more complex link-formation mechanism described by the models of Rescorla and Wagner (1972), Mackintosh (1975), Pearce and Hall (1980), Wagner (1981), and Miller and Matzel (1998). This version of the dual-system approach proposes that two very complex learning systems, which have many characteristics in common and which are sensitive to the same

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

environmental variables, operate side by side. Clarification is needed as to the crucial differences between the link-formation system and the propositional system. It is not at all clear why both of these systems are needed.

Whatever the characteristics of the link mechanism, clarification is also required as to the way in which the link system and the propositional system interact. That is, how is conflict between the output of these systems resolved? In our target article, we noted the lack of detail regarding conflict resolution in dual-system theories (although see Gawronski & Bodenhausen [2006] for an exception), and regrettably many of the present commentaries continue this noncommittal tradition. For example, **Olsson & Öhman** stated that the two levels of learning are "partially independent but also interacting." Lack of testability therefore remains a fundamental weakness of the dual-system approach. We reiterate our challenge to dual-system theorists to clarify the laws or mechanisms by which they consider the output of the two systems to be gated to produce observed behavioral outcomes. We believe that such an exercise will reveal that major concessions are needed to allow dual-system models to fit with existing data (e.g., regarding awareness, cognitive load, and verbal instruction) and that, in this light, the coordinated propositional system we propose might be viewed more favorably.

We would like to end with a comment on the evidence for the link-formation mechanism. As we pointed out in the target article and reaffirmed in this response to commentaries, a close examination of the data reveals only one or two isolated phenomena that might indicate the presence of a non-propositional (perhaps link-based) learning mechanism. These include the eyeblink version of the Perruchet effect (Perruchet 1985) and the odor-taste learning work of Stevenson et al. (1998). As **Dwyer et al.** concede, evidence for the link-formation mechanism is not widespread. Thus, even the proponents of the dual-system approach accept that the link mechanism is of somewhat limited explanatory value. It seems to us that, if we do indeed possess two separate learning mechanisms, then we should see evidence for both mechanisms everywhere. Why, therefore, is the evidence for the second mechanism so weak and so vanishingly small? We keep an open mind, but there seems to be an obvious and almost unavoidable conclusion, that no such mechanism exists.

ACKNOWLEDGMENTS
Preparation of this paper was supported by a Discovery Project grant DP0774395 to Chris Mitchell and Peter Lovibond from the Australian Research Council, and Grant BOF/GOA2006/001 of Ghent University and Travel Grant K.1.424.06.N.01 of the Research Foundation – Flanders (FWO – Vlaanderen, Belgium) to Jan De Houwer. Thanks also goes to Gavan McNally and Fred Westbrook for discussions on issues in neuroscience.

References

[The letters "a" and "r" before author's initials stand for target article and response article references, respectively.]

Adams, C. D. & Dickinson, A. (1981) Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology* 33B:109–21. [aCJM]
Allen, S. W. & Brooks, L. R. (1991) Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General* 120:3–19. [BRN]
Amsel, A. (1989) *Behaviorism, neobehaviorism, and cognitivism in learning theory: Historical and contemporary perspectives*. Erlbaum. [JDG]
Anderson, J. R. & Lebiere, C. J. (1998) *Hybrid modeling of cognition: Review of the atomic components of thought.* Erlbaum. [AGB]
Arcediano, F., Matute, H., Escobar, M. & Miller, R. R. (2005) Competition between antecedent and between subsequent stimuli in causal judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31:228–37. [DMD]
Asch, S. E. (1962) A problem in the theory of associations. *Psychologische Beiträge* 6:553–63. [GM]
(1969) A reformulation of the problem of association. *American Psychologist* 24:92–102. [GM]
Asch, S. E. & Ebenholtz, S. M. (1962) The principle of associative symmetry. *Proceedings of the American Philosophical Society* 106:135–63. [GM]
Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U. & Waldron, E. M. (1998) A neuropsychological theory of multiple systems in category learning. *Psychological Review* 105:442–81. [BRN]
Baetu, I. & Baker, A. G. (in press) Human judgments of positive and negative causal chains. *Journal of Experimental Psychology: Animal Behavior Processes*. [AGB]
Baeyens, F., Crombez, G., De Houwer, J. & Eelen, P. (1996a) No evidence for modulation of evaluative flavor-flavor associations in humans. *Learning and Motivation* 27:200–41. [FB]
Baeyens, F., Eelen, P., Crombez, G. & De Houwer, J. (2001a) On the role of beliefs in observational flavor conditioning. *Current Psychology* 20:183–203. [FB]
Baeyens, F., Eelen, P. & Van den Bergh, O. (1990a) Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning. *Cognition and Emotion* 4:3–18. [AGB, TB, aCJM]
Baeyens, F., Eelen, P., Van den Bergh, O. & Crombez, G. (1990b) Flavor-flavor and color-flavor conditioning in humans. *Learning and Motivation* 21:434–55. [FB]
Baeyens, F., Hendrickx, H., Crombez, G. & Hermans, D. (1998) Neither extended sequential nor simultaneous feature positive training result in modulation of evaluative flavor conditioning in humans. *Appetite* 31:185–204. [FB]
Baeyens, F., Vansteenwegen, D., De Houwer, J. & Crombez, G. (1996b) Observational conditioning of food valence in humans. *Appetite* 27:235–50. [FB]
Baeyens, F., Vansteenwegen, D., Hermans, D. & Eelen, P. (2001b) Chilled white wine, when all of a sudden the doorbell rings: Mere reference and evaluation versus expectancy and preparation in human Pavlovian learning. In: *Advances in psychology research, vol. 4*, ed. F. Columbus, pp. 241–77. Nova Science. [FB]
Baker, A. G. & Mackintosh, N. J. (1979) Pre-exposure to the CS alone, US alone, or CS and US uncorrelated: Latent inhibition, blocking by context or learned irrelevance? *Learning and Motivation* 10:278–94. [AGB]
Baker, A. G., Murphy, R. A. & Mehta, R. (2003) Learned irrelevance and retrospective correlation learning. *Quarterly Journal of Experimental Psychology* 56:90–101. [AGB]
Bamber, D. (1979) State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology* 19:137–81. [BRN]
Bar, M. (2003) A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience* 15:600–09. [EB-M]
(2007) The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences* 11(7):280–89. [EB-M]
Bargh, J. A. (1994) The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In: *Handbook of social cognition, vol. 1*, 2nd edition, ed. R. S. Wyer, Jr. & T. K. Srull, pp. 1–40. Erlbaum. [JSU]
Bargh, J. A. & Chartrand, T. L. (1999) The unbearable automaticity of being. *American Psychologist* 54:462–79. [NC]
Barrett, L. F. & Bar, M. (in press) Seeing it with feeling: Affective predictions about "what" and "where" in the human brain. *Proceedings of the Royal Society of London: Biological Sciences*. [EB-M]
Barrett, L. F. & Bliss-Moreau, E. (in press) Affect as a psychological primitive. *Advances in Social Psychology*. [EB-M]
Barsalou, L. W. (2008) Grounded cognition. *Annual Review of Psychology* 59:617–45. [EB-M, rCJM]
Barsalou, L. W., Simmons, W. K., Barbey, A. K. & Wilson, C. D. (2003) Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences* 7:84–91. [DRS]
Batsell, W. R., 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. [DMD, aCJM]
Bechara, A., Tranel, D, Damasio, H., Adolphs, R., Rockland, C. & Damasio, A. R. (1995) Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science* 269:1115–18. [rCJM, AO]

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

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, and Cognition* 31:238–49. [arCJM, NAS]
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. [TB, LC, aCJM, DCP]
Benedetti, F., Pollo, A., Lopiano, L., Lanotte, M., Vighetti, S. & Rainero, I. (2003) Conscious expectation and unconscious conditioning in analgesic, motor, and hormonal placebo/no placebo responses. *Journal of Neuroscience* 23:4315–23. [RAB]
Bernstein, I. L. (1985) Learned food aversions in the progression of cancer and its treatment. In: *Experimental assessments and clinical applications of conditioned food aversions*, ed. N. S. Braveman & P. Bronstein, pp. 365–80. New York Academy of Sciences. [RAB]
Bernstein, I. L. & Webster, M. M. (1980) Learned taste aversions in humans. *Physiology and Behavior* 25:363–66. [DMD, EJL]
Beylin, A. V., Gandhi, C. C., Wood, G. E., Talk, A. C., Matzel, L. D. & Shors, T. J. (2001) The role of the hippocampus in trace conditioning: temporal discontinuity or task difficulty? *Neurobiology of Learning and Memory* 76(3):447–61. [NL]
Blaisdell, A., Sawa, K., Leising, K. J. & Waldmann, M. R. (2006) Causal reasoning in rats. *Science* 311:1020–22. [LC, aCJM]
Blaser R. E., Couvillon, P. A. & Bitterman, M. E. (2004) Backward blocking in honeybees. *Quarterly Journal of Experimental Psychology B* 57:349–60. [JDM]
Blough, D. S. (1975) Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes* 1:3–21. [DMD]
Boakes, R. A. (1989) How one might find evidence for conditioning in adult humans. In: *Aversion, avoidance and anxiety: Perspectives on learning and memory*, ed. T. Archer & L.-G. Nilsson, pp. 381–402. Erlbaum. [aCJM]
Bolles, R. C. (1993) *The story of psychology: A thematic history.* Brooks-Cole. [JDG]
Bolles, R. C. & Fanselow, M. S. (1980) A perceptual-recuperative model of fear and pain. *Behavioral and Brain Sciences* 3:291–323. [aCJM]
Bouton, M. E. (1993) Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin* 114:80–99. [aCJM]
(2004) Context and behavioral processes in extinction. *Learning and Memory* 11:485–94. [AGB]
Bower, G. H. (1970) Organizational factors in memory. *Cognitive Psychology* 1:18–46. [GM]
Bower, G. H. & Bryant, D. J. (1991) On relating the organizational theory of memory to levels of processing. In: *Memories, thoughts, and emotions*, ed. W. Kessen, A. Ortony & F. Craik, pp. 149–68. Erlbaum. [GM]
Braine, M. D. S. (1978) On the relation between the natural logic of reasoning and standard logic. *Psychological Review* 85:1–21. [LC, aCJM]
Braine, M. D. S. & Rumain, B. (1983) Logical reasoning. In: *Handbook of child psychology: Cognitive development, vol. 3*, ed. J. H. Flavell & E. M. Markman. Wiley. [LC]
Brandon, S. E., Vogel, E. H. & Wagner, A. R. (2000) A componential view of configural cues in generalization and discrimination in Pavlovian conditioning. *Behavioural Brain Research* 110:67–72. [DMD]
Brewer, W. F. (1974) There is no convincing evidence for operant or classical conditioning in adult humans. In: *Cognition and the symbolic processes*, ed. W. B. Weimer & D. S. Palermo, pp. 1–42. Erlbaum. [AGB, RAB, aCJM]
Buchel, C., Dolan, R. J., Armony, J. L. & Friston, K. J. (1999) Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. *Journal of Neuroscience* 19(24):10869–76. [NL]
Burns, P. & McCormack, T. (under review) Temporal information and children's and adults' causal inferences. *Thinking and Reasoning*. [DAL]
Buzsáki, G. (2006) *Rhythms of the brain.* Oxford University Press. [EM]
Carew, T. J., Hawkins, R. D. & Kandel, E. R. (1983) Differential classical conditioning of a defensive withdrawal reflex in *Aplysia californica*. *Science* 219:397–400. [GH]
Carey, M. P. & Burish, T. G. (1988) Etiology and treatment of the psychological side effects associated with cancer chemotherapy: A critical review and discussion. *Psychological Bulletin* 104:307–25. [EJL]
Carlston, D. E. & Skowronski, J. J. (1994) Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology* 66:840–56. [JSU]
(2005) Linking versus thinking: Evidence for the differential associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology* 89:884–98. [BG, JSU]
Carson, S. H., Peterson, J. B. & Higgins, D. M. (2003) Decreased latent inhibition is associated with increased creative achievement in high-functioning individuals. *Journal of Personality and Social Psychology* 85:499–506. [rCJM]
Carter, R. M., Hofstotter, C., Tsuchiya, N. & Koch, C. (2003) Working memory and fear conditioning. *Proceedings of the National Academy of Sciences USA* 100(3):1399–404. [NL]
Chambaron, S., Ginhac, D., Ferrel-Chapus, C. & Perruchet, P. (2006) Implicit learning of a repeated segment in continuous tracking: A reappraisal. *The Quarterly Journal of Experimental Psychology* 59:845–54. [rCJM]
Chambers, L., Mobini, S. & Yeomans, M. R. (2007) Caffeine deprivation state modulates expression of acquired liking for caffeine-paired flavours. *Quarterly Journal of Experimental Psychology* 60:356–66. [RAB]
Chater, N. (2003) How much can we learn from double dissociations? *Cortex* 39:167–69. [NC]
(in press) Rational and mechanistic perspectives on reinforcement learning. *Cognition*. [NC, aCJM]
Chater, N. & Manning, C. D. (2006) Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences* 10:335–44. [AG]
Chater, N., Tenenbaum, J. B. & Yuille, A. (2006) Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Science* 10:287–91. [AG, aCJM]
Cheng, D. T., Disterhoft, J. F., Power, J. M., Ellis, D. A. & Desmond, J. E. (2008) Neural substrates underlying human delay and trace eyeblink conditioning. *Proceedings of the National Academy of Sciences USA* 105(23):8108–13. [NL]
Cheng, P. W. (1997) From covariation to causation: A causal power theory. *Psychological Review* 104:367–405. [AGB, aCJM, DCP]
Christiansen, M. H. & Chater, N. (2008) Language as shaped by the brain. *Behavioral and Brain Sciences* 31(5):489–558. [rCJM]
Christiansen, M. H., Dale, R. A. C., Ellefson, M. R. & Conway, C. M. (2002) The role of sequential learning in language evolution: Computational and experimental studies. In: *Simulating the evolution of language*, ed. A. Cangelosi & D. Parisi, pp. 165–88. Springer-Verlag. [LC]
Clancy, P., Jacobsen, T. & Silva, M. (1976) The acquisition of conjunction: Across linguistic study. *Stanford Papers and Reports on Child Language Development* 12:71–80. [LC]
Claparède, E. (1907) Expériences sur lamémoire dans un cas de psychose de Korsakoff. *Revue Médicale de la Suisse Romande* 27:301–03. [JDM]
Clark, A. (1990) Connectionism, competence, and explanation. *British Journal of Philosophical Science* 41:195–222. [rCJM]
Clark, D. M. (2004) Developing new treatments: On the interplay between theories, experimental science and clinical innovation. *Behavior Research and Therapy* 42:1089–104. [aCJM]
Clark, R. E., Manns, J. R. & Squire, L. R. (2002) Classical conditioning, awareness, and brain systems. *Trends in Cognitive Sciences* 6(12):524–31. [NL]
Clark, R. E. & Squire, L. R. (1998) Classical conditioning and brain systems: The role of awareness. *Science* 280(5360):77–81. [NL, rCJM, AO]
(1999) Human eyeblink classical conditioning: Effects of manipulating awareness of the stimulus contingencies. *Psychological Science* 10(1):14–18. [NL]
Clayton, N. S. & Dickinson, A. (1998) Episodic-like memory during cache recovery by scrub jays. *Nature* 395:272–74. [aCJM]
Cleeremans, A. (1993) *Mechanisms of implicit learning.* MIT Press. [DRS]
Cleeremans, A. & McClelland, J. L. (1991) Learning the structure of event sequences. *Journal of Experimental Psychology: General* 120(3):235–53. [DAS]
Cobos, P. L., Lopez, F. J., Cano, A., Almaraz, J. & Shanks, D. R. (2002) Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology: Animal Behavior Processes* 28:331–46. [DMD]
Cohen, N. J. & Squire, L. R. (1980) Preserved learning and pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science* 210:207–10. [JDM]
Colgan, D. M. (1970) Effect of instructions on the skin conductance response. *Journal of Experimental Psychology* 86:108–12. [aCJM]
Collins, D. J. & Shanks, D. R. (2002) Momentary and integrative response strategies in causal judgment. *Memory and Cognition* 30:1138–47. [DMD]
Cook, S. W. & Harris, R. E. (1937) The verbal conditioning of the galvanic skin reflex. *Journal of Experimental Psychology* 21:202–10. [aCJM]
Corlett, P. R., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Honey, G. D. Honey, R. A. E., Robbins, T. W., Bullmore, E. T. & Fletcher, P. C. (2004) Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron* 44:877–88. [JEW]
Cornwell, B. R., Echiverri, A. M. & Grillon, C. (2007) Sensitivity to masked conditioned stimuli predicts conditioned response magnitude under masked conditions. *Psychophysiology* 44:403–406. [rCJM]
Courville, A. C., Daw, N. D. & Touretzky, D. S. (2006) Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences* 10:294–300. [NC]

Coutureau, E. & Killcross, S. (2003) Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behavioural Brain Research* 146:167–74. [NC]

Crawford, M. T., Skowronski, J. J., Stiff, C. & Scherer, C. R. (2007) Interfering with inferential, but not associative, processes underlying spontaneous trait inference. *Personality and Social Psychology Bulletin* 33:677–90. [JSU]

Critchley, H. D., Mathias, C. J. & Dolan, R. J. (2002) Fear conditioning in humans: The influence of awareness and autonomic arousal on functional neuroanatomy. *Neuron* 33:653–63. [AO]

Culler, E. & Mettler, F. A. (1934) Conditioned behavior in a decorticate dog. *Journal of Comparative Psychology* 18:291–303. [JDM]

Cunningham, W. A. & Zelazo, P. D. (2007) Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences* 11:97–104. [NC]

Davey, G .C. L. (1992) Classical conditioning and the acquisition of human fears and phobias: A review and synthesis of the literature. *Advances in Behaviour Research and Therapy* 14:29–66. [rCJM]

Davis, M. (1992) The role of the amygdala in fear and anxiety. *Annual Review Neuroscience* 15:353–75. [AO]

(2006) Neural systems involved in fear and anxiety measured with fear potential startle. *American Psychologist* 61:741–56. [AO]

Dawson, M. E. (1970) Cognition and conditioning: Effects of masking the CS-UCS contingency on human GSR classical conditioning. *Journal of Experimental Psychology* 85:389–96. [aCJM]

Dawson, M. E. & Biferno, M. A. (1973) Concurrent measurement of awareness and electrodermal classical conditioning. *Journal of Experimental Psychology* 101:55–62. [MED, arCJM]

Dawson, M. E. & Furedy, J. J. (1976) The role of awareness in human differential autonomic classical conditioning: The necessary-gate hypothesis. *Psychophysiology* 13:50–53. [MED]

Dawson, M. E., Rissling, A. J., Schell, A. M. & Wilcox, R. (2007) Under what conditions can human affective conditioning occur without contingency awareness? Test of the evaluative conditioning paradigm. *Emotion* 7:755–66. [rCJM]

Dawson, M. E. & Schell, A. M. (1985) Information processing and human autonomic classical conditioning. In: *Advances in psychophysiology, vol. 1*, ed. P. K. Ackles, J. R. Jennings & M. G. H. Coles, pp. 89–165. JAI Press. [aCJM]

(1987) Human autonomic and skeletal classical conditioning: The role of conscious cognitive factors. In: *Cognitive processes and Pavlovian conditioning in humans*, ed. G. Davey, pp. 27–55. Wiley. [aCJM]

Dawson, M. E., Schell, A. M. & Banis, H. T. (1986) Greater resistance to extinction of electrodermal responses conditioned to potentially phobic CS: A noncognitive process? *Psychophysiology* 23:552–61. [MED, rCJM]

Dawson, M. E., Schell, A. M., Beers, J. R. & Kelly, A. (1982) Allocation of cognitive processing capacity during human autonomic conditioning. *Journal of Experimental Psychology: General* 111:273–95. [MED]

Dayan, P. & Abbott, L. F. (2001) *Theoretical neuroscience*. MIT Press. [NC]

De Houwer, J. (2007) A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology* 10:230–41. [aCJM]

(2009) The propositional approach to associative learning as an alternative for association formation models. *Learning and Behavior* 37:1–20. [DAL, arCJM]

De Houwer, J. & Beckers, T. (2002) Higher-order retrospective revaluation in human causal learning. *The Quarterly Journal of Experimental Psychology* 55B:137–51. [NAS]

(2003) Secondary task difficulty modulates forward blocking in human contingency learning. *Quarterly Journal of Experimental Psychology* 56B:345–57. [AGB, LC, aCJM, DAS]

De Houwer, J., Beckers, T. & Glautier, S. (2002) Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology* 55A:965–85. [aCJM]

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* 127:853–69. [aCJM]

De Houwer, J., Vandorpe, S. & Beckers, T. (2005) On the role of controlled cognitive processes in human associative learning. In: *New directions in human associative learning*, ed. A. J. Wills, pp. 41–63. Erlbaum. [aCJM]

(2007) Statistical contingency has a different impact on preparation judgements than on causal judgements. *Quarterly Journal of Experimental Psychology* 60:418–32. [aCJM]

de Waal, F. B. M. (2002) Evolutionary psychology: The wheat and the chaff. *Current Directions in Psychological Science* 11:187–91. [EM]

Deisig, N., Lachnit, H., Sandoz, J.-C., Lober, K. & Giurfa, M. (2003) A modified version of the unique cue theory accounts for olfactory compound processing in honeybees. *Learning and Memory* 10:199–208. [HS]

Deisig, N, Sandoz, J.-C., Giurfa, M. & Lachnit, H. (2007) The trial spacing effect in olfactory patterning discriminations in honeybees. *Behavioural Brain Research* 176(2):314–22. [HS]

Dickinson, A. (1980) *Contemporary animal learning theory.* Cambridge University Press. [aCJM]

(1989) Expectancy theory in animal conditioning. In: *Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory*, ed. S. B. Klein & R. R. Mowrer, pp. 279–308. Erlbaum. [aCJM]

(2001) Causal learning: An associative analysis. *The Quarterly Journal of Experimental Psychology* 54B:3–25. [aCJM]

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

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

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

Diez-Chamizo, V., Sterio, D. & Mackintosh, N. J. (1985) Blocking and overshadowing between intra-maze and extra-maze cues: A test of the independence of locale and guidance learning. *The Quarterly Journal of Experimental Psychology* 37B:235–53. [AGB]

Dulany, D. E. (1968) Awareness, rules, and propositional control: A confrontation with S-R behavior theory. In: *Verbal behavior and general behavior theory*, ed. T. R. Dixon & D. C. Horton. Prentice-Hall. [JDG]

Duncker, K. (1926) A qualitative study of productive thinking. *Pedagogical Seminary* 33:642–708. [GM]

Dunn, J. C. (2003) The elusive dissociation. *Cortex* 39:177–79. [BRN]

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

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* 6:175–87. [DMD, aCJM]

Endo, N. & Takeda, Y. (2004) Selective learning of spatial configuration and object identity in visual search. *Perception and Psychophysics* 66:293–302. [DRS]

Enkvist, T., Newell, B. R., Juslin, P. & Olsson, H. (2006) On the role of causal intervention in multiple-cue judgment: Positive and negative effects on learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:163–79. [BRN]

Epstein, S. (1994) Integration of the cognitive and psychodynamic unconscious. *American Psychologist* 49:709–24. [NC]

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. [aCJM]

Estes, W. K. (1950) Towards a statistical theory of learning. *Psychological Review* 57:94–107. [DMD]

Esteves, F., Parra, C., Dimberg, U. & Öhman, A. (1994) Non-conscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. *Psychophysiology* 31:375–85. [MED]

Evans, J. St. B. T. (2003) In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences* 7:454–59. [aCJM]

(2008) Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology* 59(1):255–78. [EB-M, aCJM, BRN, DCP]

Evans, J. St. B. T. & Over, D. E. (1996) *Rationality and reasoning*. Psychology Press. [aCJM]

Falmagne, R. J. (1975) Overview: Reasoning, representation, process, and related issues. In: *Reasoning: Representation and process in children and adults*, ed. R. J. Falmagne. Erlbaum. [LC]

Fanselow, M. S. (1994) Neural organization of the defensive behavior system responsible for fear. *Psychonomic Bulletin and Review* 1:429–38. [AO]

Fazio, R. H. & Olson, M. A. (2003) Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology* 54:297–327. [BG]

Fodor, J. A. (2003) *Hume variations*. Oxford University Press/Clarendon Press. [TB, aCJM]

Fodor, J. A. & Pylyshyn, Z. W. (1988) Connectionism and cognitive architecture. *Cognition* 28:3–71. [DCP]

Fries, P. (2005) A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences* 9:474–80. [EM]

Fulcher, E. P. & Hammerl, M. (2001) When all is revealed: A dissociation between evaluative learning and contingency awareness. *Consciousness and Cognition* 10:524–49. [aCJM]

Gabrieli, J. D., Carrillo, M. C., Cermak, L. S., McGlinchey-Berroth, R., Gluck, M. A. & Disterhoft, J. F. (1995) Intact delay-eyeblink classical conditioning in amnesia. *Behavioral Neuroscience* 109:819–27. [JDM]

Gallistel, C. R. (1990) *The organization of learning.* MIT Press. [EM]

(2000) The replacement of general-purpose learning models with adaptively specialized learning modules. In: *The new cognitive neurosciences*, ed. M. S. Gazzaniga, pp. 1179–91. MIT Press. [HM]

(2001) Mental representations, psychology of. In: *International encyclopedia of the social and behavioural* sciences, ed. N. J. Smelser & P. B. Baltes, pp. 9691–95. Elsevier. [EM]

Gallistel, C. R., Balsam, P. D. & Fairhurst, S. (2004) The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences USA* 101:13124–31. [NC]

Gallistel, C. R. & Gibbon, J. (2001) Computational versus associative models of simple conditioning. *Current Directions in Psychological Science* 10:146–50. [EM]

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

Garcia, J. & Koelling, R. A. (1966) Relation of cue to consequence in avoidance learning. *Psychonomic Science* 4:123–24. [JDG]

Gardner, H. (1985) *The mind's new science: A history of the cognitive revolution*. Basic Books. [aCJM]

Gawronski, B. & Bodenhausen, G. V. (2006) Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin* 132:692–731. [BG, rCJM]

(2007) Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE Model. *Social Cognition* 25:687–717. [BG]

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B. & Strack, F. (2008) When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology* 44:370–77. [BG]

Gawronski, B. & LeBel, E. P. (2008) Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology* 44:1355–61. [BG]

Gawronski, B. & Strack, F. (2004) On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology* 40:535–42. [BG]

Gawronski, B. & Walther, E. (2008) The TAR effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin* 34:1276–89. [BG]

Ghirlanda, S. (2005) Retrospective revaluation as simple associative learning. *Journal of Experimental Psychology: Animal Behavior Processes* 31:107–11. [DRS]

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

Glymour, C. (2007) Statistical jokes and social effects: Intervention and invariance in causal relations. In: *Causal learning: Psychology, philosophy, and computation*, ed. A. Gopnik & L. Schulz, pp. 294–300. Oxford University Press. [DAL]

Goldstone, R. L. & Barsalou, L. W. (1998) Reuniting perception and conception. *Cognition* 65:231–62. [LC]

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T. & Danks, D. (2004) A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review* 111:1–32. [AG, DAL]

Gopnik, A. & Schulz, L. (2004) Mechanisms of theory-formation in young children. *Trends in Cognitive Science* 8:8. [AG]

eds. (2007) *Causal learning: Psychology, philosophy, and computation*. Oxford University Press. [NC, AG]

Gopnik, A. & Tenenbaum, J., eds. (2007) Special issue of *Developmental Science* on Bayes net and Bayesian methods in cognitive development. *Developmental Science* 10. [AG]

Gray, J. A. (2004) *Consciousness: Creeping up on the hard problem.* Oxford University Press. [EM]

Gray, J. A., Buhusi, C. V. & Schmajuk, N. A. (1997) The transition from automatic to controlled processing. *Neural Networks* 10:1257–68. [NAS]

Greenspoon, J. (1955) The reinforcing effect of two spoken sounds on the frequency of two behaviors. *American Journal of Psychology* 68:409–16. [JDG]

Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. (1998) Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology* 74:1464–80. [JSU]

Greenwood, J. D. (2001) Understanding the "cognitive revolution" in psychology. *Journal of the History of the Behavioral Sciences* 35:1–22. [JDG]

(2008) *A conceptual history of psychology.* McGraw-Hill. [JDG]

Griffiths, T. L. & Tenenbaum, J. B. (2005) Structure and strength in causal induction. *Cognitive Psychology* 51:354–84. [DAL]

Guez, D. & Miller, R. R. (2008) Blocking and pseudoblocking: The reply of *Rattus norgevicus* to *Apis mellifera*. *Quarterly Journal of Experimental Psychology* 61:1186–98. [JDM]

Hall, G. (2002) Associative structures in Pavlovian and instrumental conditioning. In: *Stevens' handbook of experimental psychology, vol. 3*, 3rd edition, ed. C. R. Gallistel, pp. 1–45. Wiley. [GH]

Hallett, M. (2007) Volitional control of movement: The physiology of free will. *Clinical Neurophysiology* 117:1179–92. [EM]

Hamm, A. O. & Vaitl, D. (1996) Affective learning: Awareness and aversion. *Psychophysiology* 33:698–710. [rCJM, AO]

Hamm, A. O, Weike, A. I., Schupp, H. T., Treig, T., Dressel, A. & Kessler, C. (2003) Affective blindsight: Intact fear conditioning to a visual cue in a cortically blind patient. *Brain* 126:265–66. [AO]

Haselgrove, M. (under review) Reasoning rats or associative animals? Comment on Beckers, Miller, De Houwer & Urushihara (2006) *Journal of Experimental Psychology: Animal Behaviour Processes*. [LC]

Henson, R. (2006) Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in Cognitive Sciences* 10:64–69. [aCJM]

Hintzman, D. L. (1986) Schema abstraction in a multiple-trace memory model. *Psychological Review* 93:411–28. [aCJM]

Hoffmann, H., Janssen, E. & Turner, S. L. (2004) Classical conditioning of sexual arousal in women and men: Effects of varying awareness and biological relevance of the conditioned stimulus. *Archives of Sexual Behavior* 33:43–53. [EB-M]

Holland, P. C. (1990) Event representation in Pavlovian conditioning: Image and action. *Cognition* 37:105–31. [aCJM]

Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. (2001) The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences* 24(5):849–937. [EM]

Hummel, F. & Gerloff, C. (2005) Larger interregional synchrony is associated with greater behavioral success in a complex sensory integration task in humans. *Cerebral Cortex* 15:670–78. [EM]

Hummel, J. E. & Holyoak, K. J. (1997) Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review* 104:427–66. [DCP]

(2003) A symbolic-connectionist theory of relational inference and generalization. *Psychological Review* 110:220–64. [DCP]

Iselin-Chaves, I. A., Willems, S. J., Jermann, F. C., Forster, A., Adam, S. R. & Van der Linden, M. (2005) Investigation of implicit memory during isoflurane anesthesia for elective surgery using the process dissociation procedure. *Anesthesiology* 103:925–33. [DMD]

Jenkins, H. M. & Moore, B. R. (1973) The form of the auto-shaped response with food or water reinforcers. *Journal of the Experimental Analysis of Behavior* 20:163–81. [EJL]

Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V. & Graybiel, A. M. (1999) Building neural representations of habits. *Science* 286:1745–49. [aCJM, NC]

Johnson-Laird, P. N. (1983) *Mental models*. Cambridge University Press. [aCJM]

Jones, F. & McLaren, I. P. L. (1999) Rules and associations. *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*, ed. M. Hahn & S. C. Stoness. Erlbaum. [IPLM]

Juslin, P., Olsson, H. & Olsson, A. C. (2003) Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General* 132:133–56. [BRN]

Kahneman, D. & Frederick, S. (2002) Representativeness revisited: Attribute substitution in intuitive judgment. In: *Heuristics and biases*, ed. T. D. Gilovich, D. W. Griffin & D. Kahneman, pp. 49–81. Cambridge University Press. [aCJM]

Kamin, L. J. (1969) Predictability, surprise attention and conditioning. In: *Punishment and aversive behaviour*, ed. B. A. Campbell & R. M. Church, pp. 279–96. Appleton-Century-Crofts. [NC, JDG, aCJM]

Kandel, E. R. (2000) Learning and memory. In: *Principles of neural science, 4th edition*, ed. E. R. Kandel, J. H. Schwartz & T. M. Jessell, pp. 1227–46. McGraw-Hill. [EM]

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* 60:448–60. [RAB, arCJM]

Katkin, E. S., Wiens, S. & Öhman, A. (2001) Nonconscious fear conditioning, visceral perception, and the development of gut feelings. *Psychological Science* 12:366–70. [aCJM]

Katona, G. (1940) *Organizing and memorizing*. Columbia University Press. [GM]

Killcross, S. & Coutureau, E. (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex* 13:400–08. [NC]

Killeen, P. R. (1981) Learning as causal inference. In: *Quantitative analyses of behavior, vol. 1: Discriminative properties of reinforcement schedules*, ed. M. L. Commons & J. A. Nevins, pp. 89–112. Ballinger. [LC]

Kinder, A. & Lachnit, H. (2003) Similarity and discrimination in human Pavlovian conditioning. *Psychophysiology* 40:226–34. [HS]

Knight, D. C., Cheng, D. T., Smith, C. N., Stein, E. A. & Helmstetter, F. J. (2004) Neural substrates mediating human delay and trace fear conditioning. *Journal of Neuroscience* 24(1):218–28. [NL]

Knuttinen, M. G., Power, J. M., Preston, A. R. & Disterhoft, J. F. (2001) Awareness in classical differential eyeblink conditioning in young and aging humans. *Behavioral Neuroscience* 115(4):747–57. [NL, rCJM]

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

Komischke, B., Giurfa, M., Lachnit, H. & Malun, D. (2002) Successive olfactory reversal learning in honeybees. *Learning and Memory* 9:122–29. [HS]
Komischke, B., Sandoz, J.-C., Lachnit, H. & Giurfa, M. (2003) Non-elemental processing in olfactory discrimination tasks needs bilateral input in honeybees. *Behavioural Brain Research* 145:135–43. [HS]
Kruschke, J. K. (2006) Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review* 113:677–99. [aCJM]
Kruschke, J. K. & Blair, N. (2000) Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin and Review* 7(4):636–45. [DAS]
Kruschke, J. K., Kappenman, E. & Hetrick, W. (2005) Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(5):830–45. [DAS]
Kveraga, K., Ghuman, A. S. & Bar, M. (2007) Top-down predictions in the cognitive brain. *Brain and Cognition* 65:145–68. [EB-M]
Lachnit, H. & Kimmel, H. D. (1993) Positive and negative patterning in human classical skin conductance response conditioning. *Animal Learning and Behavior* 21:314–26. [HS]
Lachnit, H., Kinder, A. & Reinhard, G. (2002) Are rules applied in Pavlovian electrodermal conditioning with humans general or outcome specific? *Psychophysiology* 39:380–87. [HS]
Lachnit, H., Ludwig, I. & Reinhard, G. (2007) Responding in configural discrimination problems depends on density of reinforcement in time. *Experimental Psychology* 54:281–88. [HS]
Lagnado, D. A. & Shanks, D. R. (2002) Probability judgment in hierarchical learning: A conflict between predictiveness and coherence. *Cognition* 83:81–112. [aCJM, DRS]
Lagnado, D. A. & Sloman, S. A. (2004) The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30:856–76. [DAL]
(2006) Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:451–60. [DAL]
Lagnado, D. A., Waldmann, M. R., Hagmayer, Y. & Sloman, S. A. (2007) Beyond covariation: Cues to causal structure. In: *Causal learning: Psychology, philosophy, and computation*, ed. A. Gopnik & L. Schulz, pp. 154–72. Oxford University Press. [DAL, aCJM]
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* 24:1331–52. [aCJM]
Larrauri, J. A. & Schmajuk, N. A. (2008) Attentional, associative, and configural mechanisms in extinction. *Psychological Review* 115:640–76. [NAS]
Lashley, K. S. (1951) The problem of serial order in behavior. In: *Cerebral mechanisms in behavior. The Hixon symposium*, ed. L. A. Jeffress, pp. 112–46. Wiley. [EM]
Le Doux, J. E. (1996) *The emotional brain.* Simon & Schuster. [AO]
(2000) Emotion circuits in the brain. *Annual Review of Neuroscience* 23:155–84. [aCJM, EM]
Le Pelley, M. E., Oakeshott, S. M. & McLaren, I. P. L. (2005a) Blocking and unblocking in human causal learning. *Journal of Experimental Psychology: Animal Behavior Processes* 31:56–70. [DMD, arCJM, IPLM]
Le Pelley, M. E., Oakeshott, S. M., Wills, A. J. & McLaren, I. P. L. (2005b) The outcome-specificity of learned predictiveness effects: Parallels between human causal learning and animal conditioning. *Journal of Experimental Psychology: Animal Behavior Processes* 31(2):226–36. [IPLM]
LeDoux, J. E. (1996) *The emotional brain.* Simon & Schuster. [AO]
(2000) Emotion circuits in the brain. *Annual Review of Neuroscience* 23:155–84. [EM]
Leech, R., Mareschal, D. & Cooper, R. P. (2008) Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences* 31:357–78. [LC]
Leung, H. T. & Westbrook, R. F. (2008) Spontaneous recovery of extinguished fear responses deepens their extinction: A role for error-correction mechanisms. *Journal of Experimental Psychology: Animal Behavior Processes* 34:461–74. [TB]
Lipp, O. V. & Purkis, H. (2005) No support for dual process accounts of human affective learning in Pavlovian conditioning. *Cognition and Emotion* 19:269–82. [rCJM]
Livesey, E. J. & McLaren, I. P. L. (forthcoming) Discrimination and generalization along a simple dimension: Peak-shift and rule-governed responding. *Journal of Experimental Psychology: Animal Behavior Processes*. [IPLM]
López, F. J., Cobos, P. L. & Caño, A. (2005) Associative and causal reasoning accounts of causal induction: Symmetries and asymmetries in predictive and diagnostic inferences. *Memory and Cognition* 33:1388–98. [DMD, aCJM]
López, F. J., Cobos, P. L., Caño, A. & Shanks, D. R. (1998a) The rational analysis of human causal and probability judgment. In: *Rational models of cognition*, ed. M. Oaksford & N. Chater, pp. 314–52. Oxford University Press. [aCJM]
López, F. J., Shanks, D. R., Almaraz, J. & Fernández, P. (1998b) Effects of trial order on contingency judgments: A comparison of associative and probabilistic contrast accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:672–94. [DMD]
Lovibond, P. F. (1993) Conditioning and cognitive-behavior therapy. *Behavior Change* 10:119–30. [aCJM]
(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. [aCJM]
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* 31:133–42. [aCJM]
Lovibond, P. F., Loo, J. C. J., Weidemann, G. & Mitchell, C. J. (submitted) Awareness is necessary for differential trace and delay eyeblink conditioning in humans. [rCJM]
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. [DMD, DAL, EJL, HM, arCJM]
Lovibond, P. F., Siddle, D. A. T. & Bond, N. W. (1993) Resistance to extinction of fear-relevant stimuli: Preparedness or selective sensitization? *Journal of Experimental Psychology: General* 122:449–61. [rCJM]
Lu, H., Rojas, R. R., Beckers, T. & Yuille, A. (2008) Sequential causal learning in humans and rats. In: *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, ed. B. C. Love, K. McRae & V. M. Sloutsky, pp. 188–95. Cognitive Science Society. [DCP]
Lubek, I. & Apfelbaum, E. (1987) Neo-behaviorism and the Garcia effect: A social psychology of science approach to the history of a paradigm clash. In: *Psychology in twentieth century thought and science*, ed. M. G. Ash & W. R. Woodward. Cambridge University Press. [JDG]
Lubow, R. E. & Moore, A. U. (1959) Latent inhibition: The effect of non-reinforced preexposure to the conditional stimulus. *Journal of Comparative and Physiological Psychology* 52:415–19. [NAS]
Lyn, H. (2007) Mental representation of symbols as revealed by vocabulary errors in two bonobos *(Pan paniscus)*. *Animal Cognition* 10:461–75. [HL]
Lyn, H., Greenfield, P. M. & Savage-Rumbaugh, E. S. (2006) The development of representational play in chimpanzees and bonobos: Evolutionary implications, pretense, and the role of interspecies communication. *Cognitive Development* 21:199–213. [HL]
Lyn, H. & Savage-Rumbaugh, E. S. (2000) Observational word learning by two bonobos: Ostensive and non-ostensive contexts. *Language and Communication* 20:255–73. [HL]
MacCorquodale, K. & Meehl, P. E. (1954) Edward C. Tolman. In: *Modern learning theory*, ed. W. K. Estes, S. Koch, K. MacCorquodale, P. E. Meehl, C. G. Mueller Jr., W. N. Schoenfeld & W. S. Verplanck, pp. 177–266. Appleton-Century-Crofts. [aCJM]
Macdonald, J. S. P. & Lavie, N. (2008) Load induced blindness. *Journal of Experimental Psychology: Human Perception and Performance* 34:1078–91. [DRS]
Mackintosh, N. J. (1974) *The psychology of animal learning*. Academic Press. [aCJM]
(1975) A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review* 82(4):276–98. [arCJM]
(1978) Conditioning. In: *Psychology survey No. 1*, ed. B. M. Foss. George Allen and Unwin. [JDG]
(1983) *Conditioning and associative learning*. Oxford University Press. [NC]
MacLean, P. D. (1993) Cerebral evolution of emotion. In: *Handbook of emotions*, ed. M. Lewis & J. M. Haviland. Guilford Press. [AO]
Maia, T. V. & McClelland, J. L. (2004) A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task. *Proceedings of the National Academy of Sciences USA* 102:16075–80. [DRS]
Mandler, G. (1968) Association and organization: Facts, fancies and theories. In: *Verbal behavior and general behavior theory*, ed. T. R. Dixon & D. L. Horton, pp. 109–19. Prentice Hall. [GM]
(1979a) Organization and repetition: Organizational principles with special reference to rote learning. In: *Perspectives on memory research*, ed. L.-G. Nilsson, pp. 293–327. Erlbaum. [GM]
(1979b) Organization, memory, and mental structures. In: *Memory organization and structure*, ed. C. R. Puff. Academic Press. [GM]
(2007) *A history of modern experimental psychology: From James and Wundt to cognitive science*. MIT Press. [GM]
Mandler, G., Rabinowitz, J. C. & Simon, R. A. (1981) Coordinate organization: The holistic representation of word pairs. *American Journal of Psychology* 94:209–22. Available at: http://repositories.cdlib.org/postprints/2447. [GM]
Mandler, J. M. & Mandler, G. (1964) *Thinking: From association to Gestalt*. Wiley (Reprint edition: Greenwood Press, 1981). [GM]

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

Manns, J. R., Clark, R. E. & Squire, L. R. (2000) Parallel acquisition of awareness and trace eyeblink classical conditioning. *Learning and Memory* 7(5):267–72. [NL]

Maren, S. & Quirk, G. J. (2004) Neuronal signalling of fear memory. *Nature Reviews Neurosciences* 5:844–52. [AO]

Marinkovic, K., Schell, A. M. & Dawson, M. E. (1989) Awareness of the CS-UCS contingency and classical conditioning of skin conductance responses with olfactory CSs. *Biological Psychology* 29:39–60. [MED]

Marr, D. (1982) *Vision: A computational investigation into the human representation and processing of visual information*. Freeman. [AGB, aCJM]

Mauk, M. D. & Thompson, R. F. (1987) Retention of classically conditioned eyelid responses following acute decerebration. *Brain Research* 403:89–95. [GH]

McClelland, J., McNaughton, B. & O'Reilly, R. (1995) Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* 102(3):419–57. [DAS]

McClelland, J. & Thompson, R. (2007) Using domain-general principles to explain children's causal reasoning abilities. *Developmental Science* 10(3):333–56. [DAS]

McClure, S. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. (2004) Separate neural systems value immediate and delayed monetary rewards. *Science* 306:503–07. [NC]

McCulloch, W. S. & Pitts, W. H. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5:115–33. [AGB]

McEchron, M. D., Tseng, W. & Disterhoft, J. F. (2003) Single neurons in CA1 hippocampus encode trace interval duration during trace heart rate (fear) conditioning in rabbit. *Journal of Neuroscience* 23(4):1535–47. [NL]

McGlinchey-Berroth, R., Carrillo, M. C., Gabrieli, J. D., Brawn, C. M. & Disterhoft, J. F. (1997) Impaired trace eyeblink conditioning in bilateral, medial-temporal lobe amnesia. *Behavioral Neuroscience* 111(5):873–82. [NL]

McIntosh, A. R., Rajah, M. N. & Lobaugh, N. J. (2003) Functional connectivity of the medial temporal lobe relates to learning and awareness. *Journal of Neuroscience* 23(16):6520–28. [NL]

McLaren, I. P. L., Green, R. E. A. & Mackintosh, N. J. (1994) Animal learning and the implicit/explicit distinction. In: *Implicit and explicit learning of languages*, ed. N. C. Ellis, pp. 313–32. Academic Press. [aCJM]

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

Merker, B. (2007) Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences* 30(1):63–81. [rCJM]

Miller, R. R. & Escobar, M. (2001) Contrasting acquisition-focused and performance-focused models of acquired behavior. *Current Directions in Psychological Science* 10:141–45. [JEW]

Miller, R. R. & Matzel, L. D. (1988) The comparator hypothesis: A response rule for the expression of associations. In: *The psychology of learning and motivation, vol. 22*, ed. G. H. Bower, pp. 51–92. Academic Press. [rCJM]

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:311–29. [arCJM]

Mitchell, C. J., Lovibond, P. F. & Condoleon, M. (2005) Evidence for deductive reasoning in blocking of causal judgments. *Learning and Motivation* 36:77–87. [aCJM, NAS]

Mitchell, C. J., Lovibond, P. F., Minard, E. & Lavis, Y. (2006) Forward blocking in human learning sometimes reflects the failure to encode a cue-outcome relationship. *Quarterly Journal of Experimental Psychology* 59:830–44. [DMD]

Mitchell, C. J., Wardle, S., Lovibond, P. F., Weidemann, G. & Chang, B. (in revision) Is Perruchet's dissociation between cued reaction time and outcome expectancy evidence for two separate associative learning processes? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. [rCJM]

Moors, A. (2007) Can cognitive methods be used to study the unique aspect of emotion: An appraisal theorist's answer. *Cognition and Emotion* 21:1238–69. [aCJM]

Moors, A. & De Houwer, J. (2006) Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin* 132:297–326. [aCJM]

Morris, J. S., Buchel, C. & Dolan, R. J. (2001) Parallel neural responses in amygdala subregions and sensory cortex during implicit fear conditioning. *Neuroimage* 13:1044–52. [MED]

Morris, J. S., Öhman, A. & Dolan, R. J. (1998) Modulation of human amygdala activity by emotional learning and conscious awareness. *Nature* 393:467–70. [AO]

Morsella, E. (2005) The function of phenomenal states: Supramodular interaction theory. *Psychological Review* 112:1000–21. [EM]

Müller, G. E. (1911) Zur Analyse der Gedächtnistätigkeit und des Vorstellungsverlaufes, I. Teil. *Zeitschrift für Psychologie, Ergänzungsband* 5:xiv–403. [GM]

Murdock, B. B. J. (1966) Forward and backward associations in paired associates. *Journal of Experimental Psychology* 71:732–37. [GM]

Murphy, R. A. & Baker, A. G. (2004) A role for CS-US contingency in Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes* 30:229–39. [AGB]

Myung I. J. & Pitt, M. A. (2002) Mathematical modeling. In: *Stevens' handbook of experimental psychology, vol. 4*, ed. J. Wixsted. Wiley. [RN]

Neumann, D. L. & Waters, A. M. (2006) The use of an unpleasant sound as an unconditional stimulus in a human aversive Pavlovian conditioning procedure. *Biological Psychology* 73:175–85. [EB-M]

Newell, B. R. & Bröder, A. (2008) Cognitive processes, models and metaphors in decision research. *Judgment and Decision Making* 3:195–204. [BRN]

Newell, B. R. & Dunn, J. C. (2008) Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences* 12:285–90. [BRN]

Newell, B. R., Lagnado, D. A. & Shanks, D. R. (2007) Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin and Review* 14:505–11. [rCJM]

Nicolas, S. (1996) Experiments on implicit memory in a Korsakoff patient by Claparède (1907). *Cognitive Neuropsychology* 13:1193–99. [JDM]

Niedenthal, P. M. (2007) Embodying emotion. *Science* 316:1002–05. [DRS]

Nisbett, R. E. & Ross, L. (1980) *Human inference: Strategies and shortcomings in social judgment*. Prentice-Hall. [JDG]

Nisbett, R. E. & Wilson, T. D. (1977) Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84(3):231–59. [JDG, DCP]

Nosofsky, R. M., Clark, S. E. & Shin, H. J. (1989) Rules and exemplars in categorization, identification and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:282–304. [BRN]

Nosofsky, R. M., Kruschke, J. K. & McKinley, S. C. (1992) Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:211–33. [DRS]

Oaksford, M. & Chater, N. (2007) *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press. [DRS]

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

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* 102:121–32. [aCJM]

(1994) "Unconscious anxiety": Phobic responses to masked stimuli. *Journal of Abnormal Psychology* 103:231–40. [aCJM]

(1998) Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following non-recognized fear-relevant stimuli. *Journal of Experimental Psychology: General* 127:69–82. [MED, arCJM, AO]

Olson, M. A. & Fazio, R. H. (2001) Implicit attitude formation through classical conditioning. *Psychological Science* 12:413–17. [BG]

Olsson, A. & Phelps, E. A. (2004) Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. *Psychological Science* 15:822–28. [EB-M, AO]

Parton, D. A. & DeNike, L. D. (1966) Performance hypotheses of children and response to social reinforcement. *Journal of Personality and Social Psychology* 4:444–47. [aCJM]

Patterson, M. M., Cegavske, C. F. & Thompson, R. F. (1973) Effects of a classical conditioning paradigm on hind-limb flexor nerve response in immobilized spinal cats. *Journal of Comparative and Physiological Psychology* 84:88–97. [JDM]

Pavlov, I. (1927) *Conditioned reflexes*. Oxford University Press. [aCJM]

Pearce, J. M. (1987) A model for stimulus generalization in Pavlovian conditioning. *Psychological Review* 94:61–73. [aCJM]

Pearce, J. M. & Bouton, M. E. (2001) Theories of associative learning in animals. *Annual Review of Psychology* 52:111–39. [aCJM]

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

Pearl, J. (1988) *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann. [NC]

(2000) *Causality: Models, reasoning, and inference.* Cambridge University Press. [DAL]

Pearl, J. & Russell, S. (2001) Bayesian networks. In: *Handbook of brain theory and neural networks*, ed. M. Arbib. MIT Press. [DAL]

Penn, D. C., Holyoak, K. J. & Povinelli, D. J. (2008) Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences* 31(2):109–78. [DCP]

Penn, D. C. & Povinelli, D. J. (2007) Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology* 58:97–118. [DCP]

Perruchet, P. (1985) A pitfall for the expectancy theory of human eyelid conditioning. *Pavlovian Journal of Biological Sciences* 20:163–70. [EJL, arCJM]

Perruchet, P. & Amorim, M. A. (1992) Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:785–800. [DRS]

Perruchet, P., Cleeremans, A. & Destrebecqz, A. (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* 32:955–65. [arCJM]

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J. & Frith, C. D. (2008) Subliminal instrumental conditioning demonstrated in the human brain. *Neuron* 59:561–67. [EM]

Pessoa, L., Japee, S. & Ungerleider, L. G. (2005) Visual awareness and the detection of fearful faces. *Emotion* 5:243–47. [rCJM]

Phelps, E. A. & LeDoux, J. E. (2005) Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron* 48:175–87. [EM]

Pineño, O., Denniston, J. C., Beckers, T., Matute, H. & Miller, R. R. (2005) Contrasting predictive and causal values of predictors and of causes. *Learning and Behavior* 33:184–96. [aCJM]

Pizzagalli, D. A., Eden Evins, A., Cowman Schetter, E., Frank, M. J., Pajtas, P. E., Santesso, D. L. & Culhane, M. (2008) Single dose of a dopamine agonist impairs reinforcement learning in humans: Behavioral evidence from a laboratory-based measure of reward responsiveness. *Psychopharmacology* 196:221–32. [NC]

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, and Cognition* 33:130–44. [arCJM]

Postman, L. (1947) The history and present status of the law of effect. *Psychological Bulletin* 44:489–563. [DRS]

Price, P. C. & Yates, J. F. (1993) Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Memory and Cognition* 21:561–72. [aCJM]

Proust, M. (1913/1922) *Remembrance of Things Past: Swann's Way*, trans. C. K. Scott-Moncrieff, Henry Holt. (Original work published 1913). Retrieved online November 11, 2008, from Project Gutenberg, http://www.gutenberg.org/etext/7178). [HM]

Purkis, H. M. & Lipp, O. V. (2001) Does affective learning exist in the absence of contingency awareness? *Learning and Motivation* 32:84–99. [rCJM]

Pylyshyn, Z. W. (1973) What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin* 80:1–24. [AGB]

Razran, G. (1971) *Mind in evolution. An east-west synthesis of learned behavior and cognition*. Houghton Mifflin. [AO]

Regier, T. & Gahl, S. (2004) Learning the unlearnable: The role of missing evidence. *Cognition* 93:147–55. [AG]

Rescorla, R. A. & Durlach, P. J. (1981) Within-event learning in Pavlovian conditioning. In: *Information processing in animals: Memory mechanisms*, ed. N. E. Spear & R. Miller, pp. 81–111. Erlbaum. [DMD]

Rescorla, R. A. & Wagner, A. R. (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In: *Classical conditioning: Current theory and research, vol. 2*, ed. A. H. Black & W. F. Prokasy, pp. 64–99. Appleton-Century-Crofts. [LC, GH, arCJM, NAS, JEW]

Revuski, S. (1971) The role of interference in association over a delay. In: *Animal memory*, ed. W. K. Honig & P. H. R. James. Academic Press. [JDG]

Rumbaugh, D. M., King, J. E., Beran, M. J., Washburn, D. A. & Gould, K. L. (2007) A salience theory on learning and behavior: With perspectives on neurobiology and cognition. *International Journal of Primatology* 28:973–96. [HL]

Rumbaugh, D. M. & Washburn, D. A. (2003) The *intelligence of apes and other rational beings*. Yale University Press. [HL]

Rydell, R. J., McConnell, A. R., Mackie, D. M. & Strain, L. M. (2006) Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science* 17:954–58. [BG, JSU]

Savage-Rumbaugh, E. S., McDonald, K., Sevcik, R. A., Hopkins, W. D. & Rupert, E. (1986) Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General* 115:211–35. [HL]

Schacter, D. L. (1987) Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:501–18. [aCJM]

Schmajuk, N. A., Lam, Y. & Gray, J. A. (1996) Latent inhibition: A neural network approach. *Journal of Experimental Psychology: Animal Behavior Processes* 22:321–49. [NAS]

Schmajuk, N. A. & Larrauri, J. A. (2006) Experimental challenges to theories of classical conditioning: Application of an attentional model of storage and retrieval. *Journal of Experimental Psychology: Animal Behavior Processes* 32:1–20. [NAS]

(2008) Associative models can describe both causal learning and conditioning. *Behavioral Processes* 77:443–45. [aCJM, DRS]

Schneider, W. & Shiffrin, R. M. (1977) Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review* 84:1–66. [NAS]

Schultz, W. (1998) Predictive reward signal of dopamine neurons. *Journal of Neurophysiology* 80:1–27. [JEW]

Seligman, M. E. P. (1970) On the generality of the laws of learning. *Psychological Review* 77:406–18. [AGB]

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

Shallice, T. (1972) Dual functions of consciousness. *Psychological Review* 79:383–93. [EM]

(1988) *From neuropsychology to mental structure*. Cambridge University Press. [NC]

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

(1990) Connectionism and the learning of probabilistic concepts. *Quarterly Journal of Experimental Psychology* 42A:209–37. [DRS]

(2007) Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology* 60:291–309. [RAB, DMD, arCJM]

Shanks, D. R. & Darby, R. J. (1998) Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes* 24:405–15. [DMD, HS, aCJM]

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

Shanks, D. R. & Lovibond, P. F. (2002) Autonomic and eyeblink conditioning are closely related to contingency awareness: Reply to Wiens and Öhman (2002) and Manns et al. (2002). *Journal of Experimental Psychology: Animal Behavior Processes* 28:38–42. [TB, rCJM]

Shanks, D. R. & Pearson, S. M. (1987) A production system model of causality judgment. In: *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, pp. 210–20. Erlbaum. [DRS]

Shanks, D. R. & St. John, M. F. (1994) Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences* 17:367–447; discussion pp. 447–94. [BRN, arCJM, HM]

Shanks, D. R., Wilkinson, L. & Channon, S. (2003) Relationship between priming and recognition in deterministic and probabilistic sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:248–61. [DRS]

Sherry, D. F. & Schacter, D. L. (1987) The evolution of multiple memory systems. *Psychological Review* 94:439–54. [BRN]

Shors, T. J. (2004) Memory traces of trace memories: Neurogenesis, synaptogenesis and awareness. *Trends in Neurosciences* 27(5):250–56. [NL]

Shultz, T. R. (2003) *Computational developmental psychology.* MIT Press. [AGB]

Shultz, T. R., Mareschal, D. & Schmidt, W. C. (1994) Modeling cognitive development on balance scale phenomena. *Machine Learning* 16:57–86. [AGB]

Skinner, B. F. (1953) *Science and human behavior.* Macmillan. [EM]

(1985) Cognitive science and behaviorism. *British Journal of Psychology* 76:291–301. [JDG]

(1987) *Upon further reflection.* Prentice Hall. [JDG]

Skowronski, J. J., Carlston, D. E., Mae, L. & Crawford, M. T. (1998) Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology* 74:837–48. [BG, JSU]

Sloman, S. A. (1996) The empirical case for two systems of reasoning. *Psychological Bulletin* 119:3–22. [aCJM]

Sloman, S. A. & Lagnado, D. A. (2005) Do we "do"? *Cognitive Science* 29:5–39. [DAL]

Smyth, A. C. & Shanks, D. R. (2008) Awareness in contextual cuing with extended and concurrent explicit tests. *Memory and Cognition* 36:403–15. [rCJM, DRS]

Sobel, D. M. & Kirkham, N. Z. (2006) Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology* 42:1103–15. [LC]

Solomon, P. R., Van der Schaaf, E. R., Thompson, R. F. & Weisz, D. J. (1986) Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. *Behavioral Neuroscience* 100(5):729–44. [NL]

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

Song, S., Howard, J. H. & Howard, D. V. (2007) Implicit probabilistic sequence learning is independent of explicit awareness. *Learning and Memory* 14:167–76. [JDM]

Speekenbrink, M., Channon, S. & Shanks, D. R. (2008) Learning strategies in amnesia. *Neuroscience and Biobehavioral Reviews* 32:292–310. [rCJM]

Spence, K. W. (1950) Cognitive versus stimulus-response theories of learning. *Psychological Review* 57:159–72. [DRS]

Sperry, R. W. (1964) Neurology and the mind-brain problem. In: *Basic readings in neuropsychology*, ed. R. L. Isaacson, pp. 403–29. Harper & Row. [EM]

Spiegel, R. & McLaren, I. P. L. (2006) Associative sequence learning in humans. *Journal of Experimental Psychology: Animal Behavior Processes* 32(2):156–63. [IPLM, rCJM]

Spirtes, P., Glymour, C. & Schienes, R. (1993) *Causation, prediction and search.* Springer-Verlag. [DAL]

Stadler, M. A. & Frensch, P. A., eds. (1998) *Handbook of implicit learning.* Sage. [EM]

Stanovich, K. E. (1999) *Who is rational? Studies of individual differences in reasoning*. Erlbaum. [aCJM]

Sternberg, D. A. & McClelland, J. L. (in preparation) Situation matters: Task-specific constraints modulate cue competition effects in contingency learning. [DAS]

Stevenson, R. J. & Boakes, R. A. (2004) Sweet and sour smells: The acquisition of taste-like qualities by odours. In: *Handbook of multisensory processes*, ed. G. Calvert, C. Spence & B. E. Stein, pp. 69–83. MIT Press. [RAB]

Stevenson, R. J., Boakes, R. A. & Prescott, J. (1998) Changes in odor sweetness resulting from implicit learning of a simultaneous odor-sweetness association: An example of learned synesthesia. *Learning and Motivation* 29:113–32. [arCJM]

Stewart-Williams, S. & Podd, J. (2004) The placebo effect: Dissolving the expectancy versus conditioning debate. *Psychological Bulletin* 130:324–40. [RAB]

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J. & Blum, B. (2003) Inferring causal networks from observations and interventions. *Cognitive Science* 27:453–89. [DAL]

Stout, S. C. & Miller, R. R. (2007) Sometimes competing retrieval (SOCR): A formalization of the extended comparator hypothesis. *Psychological Review* 114:759–83. [JEW]

Strack, F. & Deutsch, R. (2004) Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review* 8:220–47. [BG, aCJM]

Swartzentruber, D. & Rescorla, M. E. (1994) Modulation of trained and extinguished stimuli by facilitators and inhibitors. *Animal Learning and Behavior* 22:309–16. [AGB]

Tangen, J. M., Allan, L. G. & Sadeghi, H. (2005) Assessing (in)sensitivity to causal asymmetry: A matter of degree. In: *New directions in human associative learning*, ed. A. J. Wills, pp. 65–93. Erlbaum. [DMD]

Tenenbaum, J. B., Griffiths, T. L. & Kemp, C. (2006) Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences* 10:309–18. [AG]

Terrace, H. S. (2005) The simultaneous chain: A new approach to serial learning. *Trends in Cognitive Science* 9:202–10. [EM]

Thompson, R. (1959) Learning in rats with extensive neocortical damage. *Science* 129:1223–24. [JDM]

Thompson, R. F. (2005) In search of memory traces. *Annual Review of Psychology* 56:1–23. [aCJM, EM]

Thorndike, E. B. (1898) Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review, Monograph Supplement* 2(8). [JDG]

Thorndike, E. L. (1931) *Human learning.* Century. [DRS]

Todorov, A. & Uleman, J. S. (2003) The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology* 39:549–62. [JSU]

(2004) The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology* 87:482–93. [JSU]

Tolman, E. C. (1948) Cognitive maps in rats and men. *Psychological Review* 55:189–208. [AGB, EM]

Tooby, J. & Cosmides, L. (1992) The psychological foundations of culture. In: *The adapted mind: Evolutionary psychology and the generation of culture*, ed. J. Barkow, L. Cosmides & J. Tooby, pp. 19–136. Oxford University Press. [HM]

(2005) Conceptual foundations of evolutionary psychology. In: *The handbook of evolutionary psychology*, ed. D. M. Buss, pp. 5–67. Wiley. [HM]

Tunney, R. J. & Shanks, D. R. (2003) Subjective measures of awareness and implicit cognition. *Memory and Cognition* 31:1060–71. [DRS]

Tversky, A. & Kahneman, D. (1983) Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90:293–315. [NC]

Uleman, J. S., Hon, A., Roman, R. & Moskowitz, G. B. (1996a) On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin* 22:377–94. [JSU]

Uleman, J. S., Newman, L. S. & Moskowitz, G. B. (1996b) People as flexible interpreters: Evidence and issues from Spontaneous Trait Inference. In: *Advances in experimental social psychology, vol. 28*, ed. M. P. Zanna, pp. 211–79. Academic Press. [JSU]

Uleman, J. S., Saribay, S. A. & Gonzalez, C. (2008) Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology* 59:329–60. [JSU]

Vadillo, M. A. & Matute, H. (2007) Predictions and causal estimations are not supported by the same associative structure. *Quarterly Journal of Experimental Psychology* 60:433–47. [aCJM]

Van Hamme, L. J. & Wasserman, E. A. (1994) Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation* 25:127–51. [DRS]

Vandorpe, S., De Houwer, J. & Beckers, T. (2007a) Outcome maximality and additivity training also influence cue competition in causal learning when learning involves many cues and events. *Quarterly Journal of Experimental Psychology* 60:356–68. [aCJM]

(2007b) The role of memory for compounds in cue competition. *Learning and Motivation* 38:195–207. [DAS]

Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A. & Eelen, P. (2006) Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology: Animal Behavior Processes* 32:71–79. [FB]

Verhagen, J. V. & Scott, T. R. (2004) Artificial neural network analysis of gustatory responses in the thalamic taste relay of the rat. *Physiology and Behavior* 80:499–513. [JEW]

Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D. & Helen, P. (2005) Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behaviour Research and Therapy* 43:357–71. [EB-M]

Wagner, A. R. (1981) SOP: A model of automatic memory processing in animal behavior. In: *Information processing in animals: Memory mechanisms*, ed. N. E. Spear & R. R. Miller, pp. 5–47. Erlbaum. [GH, arCJM]

Wagner, A. R. & Brandon, S. E. (1989) Evolution of a structured connectionist model of Pavlovian conditioning (AESOP). In: *Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory*, ed. S. B. Klein & R. R. Mowrer, pp. 149–89. Erlbaum. [aCJM]

Wagner, A. R. & Donegan, N. H. (1989) Some relationships between a computational model (SOP) and a neural circuit for Pavlovian (rabbit eyeblink) conditioning. In: *Computational models of learning in simple neural systems*, ed. R. D. Hawkins & G. H. Bower, pp. 157–203. Academic Press. [GH]

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. [aCJM]

Waldmann, M. R. & Hagmayer, Y. (2005) Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31:216–27. [DAL]

Waldmann, M. R. & Holyoak, K. J. (1992) Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General* 121:222–36. [DCP]

Waldmann, M. R. & Walker, J. M. (2005) Competence and performance in causal learning. *Learning and Behavior* 33:211–29. [AGB, aCJM]

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* 32:454–59. [aCJM]

Wardle, S. G., Mitchell, C. J. & Lovibond, P. F. (2007) Flavor evaluative conditioning and contingency awareness. *Learning and Behavior* 35:233–41. [arCJM]

Wasserman, E. A., Elek, S. M., Chatlosh, D. L. & Baker, A. G. (1993) Rating causal relations: Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:174–88. [AGB]

Wasserman, E. A. & Zentall, T. R., eds. (2006) *Animal cognition: Experimental exploration of animal intelligence.* Oxford University Press. [NC]

Weidemann, 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.* [aCJM]

Weike. A. I., Hamm. A. O., Schupp, H. T., Runge, U., Schroeder, H. W. & Kessler, C. (2005) Fear conditioning following unilateral temporal lobectomy: Dissociation of conditioned startle potentiation and autonomic learning. *Journal of Neuroscience* 25:11117–24. [AO]

Weike, A. I., Schupp, H. T. & Hamm, A. O. (2007) Fear acquisition requires awareness in trace but not delay conditioning. *Psychophysiology* 44(1):170–80. [NL, rCJM, AO]

Weiss, C., Bouwmeester, H., Power, J. M. & Disterhoft, J. F. (1999) Hippocampal lesions prevent trace eyeblink conditioning in the freely moving rat. *Behavioural Brain Research* 99(2):123–32. [NL]

Wertheimer, M. (1921) Untersuchungen zur Lehre von der Gestalt. I. *Psychologische Forschung* 1:47–58. [GM]

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press

Wheeler, D. S., Beckers, T. & Miller, R. R. (2008) The effect of subadditive pretraining on blocking. *Learning and Behavior* 36:341–51. [TB]

Whitfield, M. & Jordan, C. H. (submitted) Mutual influences of explicit and implicit attitudes. [BG]

Wiens, S., Katkin, E. S. & Öhman, A. (2003) Effects of trial order and differential conditioning on acquisition of differential shock expectancy and skin conductance conditioning to masked stimuli. *Psychophysiology* 40:989–97. [rCJM]

Wiens, S. & Öhman, A. (2002) Unawareness is more than a chance event: Comment on Lovibond and Shanks (2002) *Journal of Experimental Psychology: Animal Behavior Processes* 28:27–31. [TB]

Wilkinson, L. & Shanks, D. R. (2004) Intentional control and implicit sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30:354–69. [rCJM]

Wilson, T. D., Lindsey, S. & Schooler, T. Y. (2000) A model of dual attitudes. *Psychological Review* 107:101–26. [aCJM]

Witnauer, J. E., Urcelay, G. P. & Miller, R. R. (2008) Reduced blocking as a result of increasing the number of blocking cues. *Psychonomic Bulletin and Review* 15:651–55. [JEW]

Woodruff-Pak, D. S. & Disterhoft, J. F. (2008) Where is the trace in trace conditioning? *Trends in Neurosciences* 31(2):105–12. [NL]

Wynne, C. D. L. (1995) Reinforcement accounts for transitive inference performance. *Animal Learning and Behavior* 23:207–17. [JEW]

Xu, F. & Tenenbaum, J. B. (2007) Word learning as Bayesian inference. *Psychological Review* 114:245–73. [AG]

Young, M. E. & Wasserman, E. A. (2002) Limited attention and cue order consistency affect predictive learning: A test of similarity measures. *Journal of Experimental Psychology: Learning, Memory and Cognition* 28:484–96. [NAS]

Yuille, A. & Kersten, D. (2006) Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences* 10:301–308. [AG]

Zárate, M. A., Uleman, J. S. & Voils, C. I. (2001) Effects of culture and processing goals on the activation and binding of trait concepts. *Social Cognition* 19:295–323. [JSU]

Zinbarg, R. E. (1990) Animal research and behavior therapy part I: Behavior therapy is not what you think it is. *The Behavior Therapist* 13:171–75. [aCJM]

https://doi.org/10.1017/S0140525X09000855 Published online by Cambridge University Press