

## 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<sup>a</sup> and James L. McClelland<sup>b</sup>

<sup>a</sup>Department of Psychology, Stanford University, Stanford, CA 94305;

<sup>b</sup>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