

## 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](mailto:d.lagnado@ucl.ac.uk)

[http://www.psychol.ucl.ac.uk/people/profiles/lagnado\\_david.htm](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 precepts 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  $\rightarrow$  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  $\leftarrow$  experimenter  $\rightarrow$  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](mailto:nanxin.li@yale.edu)    <http://pantheon.yale.edu/~nl238>

### **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](mailto:evanl@psych.usyd.edu.au)

<http://www.psych.usyd.edu.au/staff/evanl>

[justinh@psych.usyd.edu.au](mailto:justinh@psych.usyd.edu.au)

<http://www.psych.usyd.edu.au/staff/justinh>