Commentary on Mitchell, C. J., De Hower, J. & Lovibond, P. F. (2009) The propositional nature of human associative learning. *Behavioral and Brain Sciences* 32:219—20.

The New Enlightenment Hypothesis: all learners are rational

Rita 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.

References

Myung I. J. & Pitt, M. A. (2002) Mathematical modeling. In: /Stevens'

Handbook of Experimental Psychology/, vol. 4, ed. J. Wixsted. John Wiley.