The Journal of Mind and Behavior
Summer 1989, Volume 10, Number 3
Pages 303–306
ISSN 0271–0137

Induction: Processes of Inference, Learning and Discovery. J.H. Holland, K.J. Holyoak, R.E. Nisbett, and P.R. Thagard. Cambridge: The MIT Press, 1986, 385 pages, $24.95 hard.

Reviewed by David Leiser, Ben-Gurion University of the Negev

Holland, Holyoak, Nisbett, and Thagard set their aims very high in their book: presenting "an exploratory framework for understanding inductive reasoning and learning in organisms and machines, from rat conditioning to scientific creativity." They stake out a vast domain, bringing work from a wide range of disciplines to bear on a central issue: How is induction computationally possible for lower organisms, higher machines, and average humans? And if endorsements and number of printings are any indication, this is indeed an important work.

While the importance of the enterprise is without dispute, I wish to present some sobering observations about what was actually accomplished. The authors' central thesis is that knowledge consists of a collection of condition-action rules. Functioning is described as parallel activation of several rules, which compete to be allowed to do their thing; activation is by a process of directed spread of activation of associates. There is much in common with Anderson's ACT* theory: production rules and spread of activation, the place given to declarative versus production knowledge. The authors also share many concepts with Newell, in particular the idea of parallelism, which Newell limits by introducing variable binding.

Learning and development are explained by differential reinforcement of rules, according to their contributions to attaining goals. The authors introduce Holland's classifiers algorithm, which is a well worked-out model of learning by systems of rules. In it, rules are reinforced both for leading to a successful action and for appropriate stage-setting activity. Over time the collection of rules should structure itself, and emergent levels of organization (schemata, theories, paradigms, etc.) are expected to emerge, by generalization, specialization, and tying together of individual rules. The authors apparently see this model as a way of saying something definite, rather than as a definite claim about how rules are modified by experience, since they describe a second computational model (PI), different from and far less worked-out than Holland's work, yet interesting as far as it goes.

What is the substance of this claim? That there are rules? But rules for whom: for the subject, or for us? Since production rules are a universal formalism, that claim

---

I wish to thank Susan Carey for many useful discussions regarding this review. The opinions expressed engage but myself. Requests for reprints should be sent to David Leiser, Ph.D., Department of Behavioral Sciences, Ben-Gurion University of the Negev, P.O. Box 653, Beer Sheva 84105 Israel.

cannot be of psychological interest. But if the claim is that rules are "in" the subject, there are many passages where it is hard to defend. A rat may press a bar when hungry and, when administered a shock, it may drop everything and try to escape. Surely, this doesn't imply (as claimed, p. 153) that there is a rule, stating: "IF in a box, and hungry, and a shock is experienced, THEN do not press bar." The distinction is so basic and well-known that it is unforgivable to be unclear about it. The same type of confusion occurs at the other, conceptual end: from the existence of some high-order consistency in behavior, can one conclude that rules are encoded and followed (e.g., Kepler's abstract models, p. 325)?

The authors' model is extremely reductionistic, restricting everything to condition/action pairs. The authors ought therefore to pay more attention to what are, prima facie, higher-order structures—how they are represented, and whether they are at all likely to emerge from the algorithms as given. I had hoped to find what is clearly required to complement Anderson's ACT* model: a theory of emergence. The authors are aware of the need but do not deliver. One author who has a lot to say on this, Piaget, is very grossly misrepresented (pp. 226 and 354) and evidently not well known to the writers. Top-down, knowledge- or theory-driven development is given very short shrift, as is the account of the development of logical structures (Piaget's "logico-mathematical abstraction"), and of substantive theory formation in general (his "empirical abstraction").

The authors equivocate on many of the key issues. Are higher-order structures explicit or implicit? The authors' wavering on this point prevents them—and us—from focusing on the transition from the one to the other (e.g., ISA links, p. 133; paradigms, concepts, bundles of rules). After reading the book and the authors' discussion of the richness of higher-order structures, one remains unconvinced that their few operational rules are enough to generate all that has to be generated, in the way of such structures (although rules may implement the whole thing once developed, but that is very different).

The authors make some vague pronouncements on the importance of "causal" theory formation, i.e., going beyond observing surface regularities to some deeper understanding of their causes, only to refer back, two sections later, to the "causal mechanisms" allegedly discussed earlier. Which mechanism (pp. 137, 191, 240)? The sections on abduction (pp. 89 and 136) are similarly exercises in mutual passing of the buck.

Is knowledge specific or abstract and general? The empirical conclusion of their review of the literature is that it is specific in physics, social science, and logic (p. 205), but general in statistics (p. 256). This should give one pause, but no, the authors blithely assert of *both* conclusions that they follow from their model. In fact, in the case of statistics, the authors waver, and we are left with no conclusion at all.

The authors claim not to be interested in coherence—which is unrealistic in the case of systems composed of numerous modular rules—then insist on its importance (pp. 38 versus 139). True, it is success that is reinforced by the algorithm; at any given time, consistency is not an overt requirement. Children are notoriously inconsistent. However, they gradually outgrow this failing. Success in a structured world implies coherence and, therefore, assuming their system does go somewhere useful, coherence is a driving force, an "attractor," a "final force." Whatever the terminology, if it is not, theirs cannot be a satisfactory account of development.

Are concepts internally meaningful, or just contentless patterns of 0s and 1s ? They seem to be meaningful in the PI model but not in the classifiers model, yet the two models are presented side by side. What is the authors' position?

Integration between authors is very loose: we are treated to some of Nisbett's work, some of Holyoak's, some of Thagard's, and some of Holland's. Much of this research is genuinely interesting; but at times, the parts are at odds (p. 325) with one another, and relations to the worked-out model are mostly strained.

The book is constructed according to the overused rhetorical move: present a well worked-out model in detail, complete with mathematical Appendix, then use it as metaphor of whatever may be going on at a much more complex level, while still benefiting from the aura of technicality evoked by the worked-out model.

Actual integration of the range of material discussed would have been priceless. What we find is superficial treatment of most issues, cavalier dismissal of philosophical puzzles that have exercised some of the best professional minds, and nothing like the grand synthesis promised by the rave reviews printed on the back.

Is it a Bad Book then? Certainly not. Doing better is not going to be easy. The book is a fine, up-to-date compendium of facts and ideas relevant to the computational problem of inductions in humans. But an integrated theory (*pace* the reviewers) it is not.