feature discovery in both frameworks. Representational assumptions can be hidden in both approaches, but both frameworks can also make them explicit, and thus subject to modification.

Another issue that crosses the connectionist/AI boundary concerns the grounding of symbols (Harnad 1990). Many systems in *both* frameworks use symbols or nodes that have no connection to sensory input or motor output. AI systems seldom ground symbols like "blue" in perceptually-based definitions, but the same holds true for most network "input" nodes that respond to "blue" stimuli. Similarly, a few researchers in *both* frameworks have started to focus on this important topic, linking learning systems to sensori-motor components that make contact with an external environment (Iba & Langley 1987; Laird et al. 1989). Networks no more force one in this direction than do symbols; rather, the driving goal is a complete theory of cognition and perception, which might be cast in either formalism.

Finally, H&B argue that the connectionist approach leads to "surprisingly elegant and powerful models of memory, perception, motor control, categorization, and reasoning" (abstract), yet they provide no convincing evidence for this claim. Research in "symbolic" approaches to learning has addressed all of these topics, yet models of many complex cognitive phenomena handled in this framework remain beyond the reach of current connectionist systems.

In summary, H&B present a useful analysis of representational issues that arise in connectionist systems, describing some novel tools for their study. However, they also pepper their article with unsubstantiated claims about the superiority of this framework over "symbolic" methods, ignoring important advances in machine learning that address all of the issues that they raise. They present convincing arguments that networks open a promising avenue to integrating representation and learning, but they give no evidence to support their contention that it "provides a unique approach" to this topic. If anything, the increasing amount of research on learning and representation in both frameworks actually bolsters Fodor and Pylyshyn's (1988) arguments about the intrinsic relation between these two approaches.

Clearly, interactions between learning and representation will be central to an integrated theory of cognition, and future research should give priority to this topic. Cognitive science and AI are better served, however, by substantive studies and by attempts to *unify* apparently disparate frameworks (Langley 1989) than by polemical statements about the superiority of one approach over another, as emphasized in the current target article.

On learnability, empirical foundations, and naturalness

W. J. M. Levelt

Max Planck Institute for Psycholinguistics Nijmegen, The Netherlands
Electronic mail: *pim@hnympi51.bitnet*

One must distinguish between formal systems and their potential applications in the modeling of empirical domains. At the level of formal systems, connectionist models (as in Hanson and Burr's [H&B's] taxonomy, Figure 1) play a role comparable to automata in classical computational theory. It is at this level that H&B approach the issues of learning and representation and their interrelations. It is therefore a category error to compare formal connectionist systems to empirical cognitive theories. Even if it were true, as H&B claim, that the latter had ignored the relations between learning and representation (which is by no means the case – at least not in my field, psycholinguistics, where this very issue is part and parcel of language-acquisition research), it is irrelevant to the question at hand. The point is whether automata theory has ignored the relation between learning and representation. And the answer is no. Learnability theory has always been at the heels of (representational) automata theory, both in its deterministic and its probabilistic versions (cf., Levelt 1974, vol. I, and a host of later publications). And there is, at that, an important qualitative advantage to classical learnability theory: It has provided formal proofs for the learnability of myriad triples of automaton, knowledge domain, and presentation schedule. No such proofs are available for connectionist systems. Case studies can never make up for formal proof in these matters.

Turning now to the connectionist modeling of empirical domains, I want to address H&B's presupposition that learning in nets is natural. ("This is one of the reasons why learning is so natural in nets," section 4.1). This must be a claim about the close compatibility between learning in nets and learning in human beings. It may or may not be the case that there is such a close compatibility for particular domains of knowledge acquisition. But, first, this is not yet known: There exists no empirical connectionist psychology worth speaking about. The typical end product in connectionist research is some working system, but such systems are almost never put to empirical tests, at least not to tests that meet the standards of present-day experimental psychology. Second, such close compatibility is certainly not a striking feature where typically human learning capabilities are concerned. Let me give two examples: learning without overriding old knowledge, and one-shot rule learning.

(1) Learning without overriding old knowledge. In acquiring arithmetic, children can, after having learned addition, learn multiplication without loosing their addition skills. The only way to do this in nets is to train them on both skills simultaneously. For children (as well as for adults) the typical learning situation, however, is one in which new knowledge is added without having to retrain all the old skills simultaneously. Here nets do not behave naturally at all; old knowledge typically gets overridden.

(2) One-shot rule learning. If I am told that the numbers in my telephone district are, as of today, extended by the initial digit 2, I can immediately apply that rule without having to be retrained on all my name/telephone number associations. Such one-shot rule learning is at the basis of our educational system. It is a characteristically human form of learning, and I have not yet come across any net that performs this feat.

Three conclusions: First, there is a formal learnability theory for classical architectures, but not for connectionist architectures. Second, there is as yet no empirical connectionist psychology to support empirical claims about learning and representation. Third, human learning displays characteristic features that are not captured by nets.

Toward a unification of conditioning and cognition in animal learning

William S. Maki

Departments of Psychology and Computer Science, North Dakota State University, Fargo, ND 58105
Electronic mail: *nu021116@ndsuvm1.bitnet*

Hanson & Burr's (H&B's) main claim that connectionism provides the motive and means for the unified study of learning and representation. If they are right, we ought to see signs that connectionism is narrowing gaps both between and within disciplines in the cognitive sciences. A quick survey of the recent literature on "neural networks" elicits a strong impression that connectionism is providing a point of contact between the disciplines of neuroscience, computer science, and psychology. Here I review the present and potential impact of connectionism in the parts of psychology known as animal learning and