Normativity and Mind

Duality of the Mind: A Bottom Up Approach Toward Cognition
by Ron Sun
Mahwah, NJ: Erlbaum, 2002. 283 pp. ISBN 0-8058-3880-5. $59.95

Review by Mark H. Bickhard

As is apt for a work with a thematic focus on dualities, this book proceeds on two levels. The first is concerned with a computer model of problem solving and learning that the author has been developing for some years, called (in its current version) Connectionistic Learning with Adaptive Rule Induction ON-line, or CLARION. Grounding the concern with duality, this computer model consists of two strata, intending to capture implicit (lower stratum) and explicit (higher stratum) aspects of representation, learning, and problem solving. The second level of the book is concerned with broad and deep issues in the philosophy of mind, such as the purported Heideggerian nature of representation, the nature of consciousness, and sociocultural factors in cognition—and the implications and connections between the computational model and these broader issues.

Continuing the theme of duality, my review also begins with two parts: (a) the underlying computational model is impressive and among the best of the hybrid symbolic-network models, and (b) nevertheless, I have concerns about the ultimate adequacy of such approaches for capturing the nature of mind with respect to the deeper issues of the nature of representation, consciousness, and sociality. Finally, I will suggest that these deeper concerns introduce a thematic issue that is not specific to this book but is endemic to the current fields of psychology, artificial intelligence, and cognitive science: a fundamental perplexity regarding normativity.

CLARION embodies a basic architectural hypothesis about thinking: an architectural differentiation into implicit and explicit kinds of processing, with various interactions between the processes in these two architectures. The implicit part of CLARION is a network that develops skills via a Q-learning-back propagation algorithm, a form of reinforcement learning. The explicit part develops rules using the rule-extraction-revision algorithm, which operates on the underlying network to extract rules of action. This second level can also modularize its problem situations with spatial and temporal partitioning.

The implicit-explicit distinction serves protean functions in this work, but the core intent is to capture the distinction in research on human learning between implicit and explicit learning. One example of the aptness of the model is that skills in the bottom layer of CLARION are distributed in the network, whereas rules in the top layer are symbolically, therefore locally, encoded. This difference, then, claims to capture the human phenomenon of implicit learning being relatively inaccessible and explicit learning being relatively accessible.

RON SUN, Department of Computer Engineering and Computer Science, University of Missouri—Columbia.

MARK H. BICKHARD, Department of Psychology, Lehigh University, 17 Memorial Drive East, Bethlehem, PA 18015. E-mail: mhb0@lehigh.edu

CLARION is able to model impressive ranges of data from human experiments. These include modeling implicit and explicit learning separately, as well as manifesting the synergy between the two under appropriate conditions. Sun also makes numerous comparisons between CLARION and other models in the literature, arguing for a clear superiority for the phenomena that CLARION is intended to capture. Sun is one of the leaders in exploring the potentialities of hybrid systems, part network and part symbol manipulation, and CLARION illustrates why. For anyone interested in a summary of, or introduction to his work, these parts of the book are recommended.

It is clear, however, that a significant part of Sun's heart is in the broader and deeper issues addressed, and here I am less sanguine about success. Let me begin with issues of representation. All representation in CLARION is based on the encoded features that are input to the base level network. All learning is of "action" choices and sequences with respect to encoded "reinforcement" levels resulting from the attainment of certain states. I have long argued against the adequacy of any encoding-based model of representation (Bickhard, 2002), but for now I will simply point out that modeling representation in terms of encodings is modeling representation in terms of (a form of) representation. At a minimum, there is no account of representation that is genuine for the system itself, not just for the designer or user of that system. All such "encodings" are for the external observer only.

Furthermore, both the initial features and the reinforcements are hand-coded for each problem. There is no way to learn new basic features, new "atoms," of representation, nor of any basic norms and goals that are real for the system (Morrison, 1997). CLARION could not handle, for example, ill-defined problems in which the individual has to learn more about what the problem is—the normative constraints that define the problem—at the same time as learning more about how to solve it, such as a detective learning more about what kind of murder is involved and simultaneously thereby guiding in appropriate ways the search for the murderer (Christensen & Hooker, 2000). These points hold for virtually all computer models, and thus do not affect CLARION any more than others, but that does not mitigate the force of the points for the deeper aspirations of modeling the mind.

Sun contends that representation emerges in action and interaction in the world, in Heideggerian *comportment*. I am in strong agreement with this general orientation, but CLARION's desiccated model of such interaction does not seem like a good framework for addressing interactively emergent representation. The input representational features are not emergent in interaction, and, because all further representation is based on them, it is not clear how interaction plays any role in generating representation, except perhaps for selecting particular weight vectors on those features. There is no differentiation between internal system states and external environmental states, so the relationship between them is problematic. The "actions" are action encodings, with no connection to any real environment. The "reinforcements" are not learned or emergent. And the normativity of all representations (e.g., the ability to determine that misrepresentation has occurred) is resident only in the designer or user of the system, not for the system itself. This is particularly important in that (a) Heideggerian comportment is saturated with normativity and anticipation and (b) the power of interactively emergent representation is precisely that it can capture such normativity and anticipation. CLARION does not, and it is not at all clear that any such computer model ever could do more than to simulate such processes (because of failures to capture genuine normativity), but CLARION does not do that either. For example, no model of representational content is offered.

System detection of representational error guides both action and learning, so the inability to model such detection is a serious practical problem, as well as a theoretical and philosophical problem. CLARION learns with respect to preencoded reinforcement, not with respect to detection of its own representational error. A truly autonomous robot, for example, will have to do better.

Sun's discussion of consciousness is wide ranging, addressing both computational and neurological models. The implicit-explicit framework of CLARION is well suited to capture some issues regarding consciousness—the distinction itself is defined in terms of ease of conscious access to the representations involved. In practice, such access is usually manifested in the ability to verbalize the knowledge available. CLARION has no model of language, however, and the assumption seems to be that the local symbolic encodings in the "rules" layer capture easy access and that a straight readout of these symbols suffices for verbalizing. CLARION suffers from the problems of all computational models with respect to the phenomenality of consciousness—it is very difficult to see how a straightforward computational model can capture this aspect. Sun makes a number of interesting suggestions but acknowledges that this is primarily a problem for the future.

Sun's attempt to stretch the CLARION model to accommodate social influences on cognition shows signs of definite strain. The duality at the social level that he focuses on is that between emergent social processes and structures, on the one hand, and the level of the individual cognition of the participants in such structures and processes, on the other. I agree that this is a crucial point of emergence, but CLARION is then used to model the distinction between absorbed involvement in a social role, via the lower level of implicit knowledge, and an explicit level of instruction or advice taking. The social level of emergent structures and processes has been lost. Furthermore, here we find again the assumption that the translation between internal and external symbols, as in advice

taking, is unproblematic. But, if that were so, language learning would not need to occur.

In general, I very much applaud Sun's concerns with foundational issues of representation, consciousness, and sociality. All too often, it is simply assumed that these can be safely set aside for later work. But, if the framework assumptions are in error, then current work can end up being irrelevant in the long run. No one reads classic associationistic psychology anymore (well, almost), and the premier journal, *The Journal of Verbal Learning and Verbal Behavior,* no longer exists. So Sun's interest in being able to extend the CLARION model is refreshing in its recognition that models of skill learning must ultimately be capable of integration with models of representation, consciousness, and sociality, to name three foundational constraints.

My concern, however, is that the framework assumptions of computer modeling in general, as well as those of information processing approaches in psychology, and, more broadly, the dominant approaches in cognitive science writ large, including its intersecting fields of psychology, philosophy, artificial intelligence, linguistics, anthropology, and neuroscience, are themselves in foundational error.

This is a complex set of issues—not to be solved or even surveyed here. I would like, however, to point out an apparent dilemma that contemporary sciences of mind face. On one hand, mind is the quintessential domain of normativity. Representational content is that which a representation *ought* to represent. Learning is to avoid various kinds of *error.* Rationality generates reasoning and problem-solving *norms.* Emotions sort our experiences and our worlds into *positive* and *negative* (not mutually exclusive). Language and sociality are saturated with interpersonal *expectations,* criteria of *appropriateness,* norms of *politeness,* grammatical *constraints,* and cultural *frameworks* for life and world. Personality is subject to the *dysfunctionalities* of psychopathology, and so on. There is no mind without normativity.

And yet, on the other hand, we are caught in a time in which attention to normativities is at best without philosophical and theoretical guidance, and at worst seems like giving credence to something mystical. Normativity does not seem to exist in the "natural" world, so how can normative phenomena be modeled without violating naturalism, without introducing some sort of dualism? Proscriptions against normativity range from the impossibility of determining the truth or falsity of our representations of the world that is argued for in radical skepticism—any such determination requires comparing what the representation is *supposed* to represent against what it is in the world that it is *being taken* to represent, but that "represented in the world" can only be accessed via the representation in question, so any such check is circular—to the ubiquitous constraint derived from Hume that norms cannot be derived from facts: no "ought" from "is." Kant recognized that if Hume's constraint holds, then knowledge cannot be accounted for within a naturalistic framework: Knowledge is inherently normative. But neither Kant's nor any successor's attempt to reconcile these points has been successful, and today we work in a framework most akin to Quine's austere ontology, barren of any normativity or necessity or intensionality. That may be a good self-discipline for avoiding dreaded dualism, but it does not suffice for addressing and modeling phenomena that are intrinsically normative, such as mind.

In sum, then, I find Sun's CLARION to be a strong and interesting exploration of the power of hybrid models to capture properties of human learning and problem solving. It is among the best and is to be highly recommended to anyone for whom computer models in this domain are in the least relevant. I find the attention to broader foundational concerns—representation, consciousness, sociality—to be refreshing and stimulating in a book with a focus on a computational model. Again, Sun is to be strongly applauded for recognizing that such issues must be addressed and cannot simply be postponed. These broader issues, however, also illustrate, in my judgment, some strong limitations of CLARION in particular, and, more generally, a serious dilemma in contemporary studies of mind across all relevant domains. I have some thoughts of my own about what is required to transcend this dilemma, but my intent at this point is simply to suggest that the dilemma exists and demands attention. ❑

References

Bickhard, M. H. (2002). The biological emergence of representation. In T. Brown & L. Smith (Eds.). *Emergence and reduction: Proceedings of the 29th Annual Symposium of the Jean Piaget Society* (pp. 105–131). Mahwah, NJ: Erlbaum.

Christensen, W. D., & Hooker, C. A. (2000). An interactivist-constructivist approach to intelligence: Self-directed anticipative learning. *Philosophical Psychology, 13*(1), 5–45.

Morrison, C. (1997). Are we wrong about representation? *Journal of Experimental and Theoretical Artificial Intelligence, 9,* 441–469.