Cognitive Systems and the Extended Mind
Robert D. Rupert
New York: Oxford University Press, 2009
288 pages, ISBN: 0195379454 (hbk); $55.00

In an agenda setting paper, Andy Clark and David Chalmers (1998) argued that "*beliefs* can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes" (p. 12). Almost immediately, this claim provoked a spate of difficult questions about the relevant role that features of the environment must play to count as constituent parts of cognitive processes. In attempting to answer these questions, sympathetic voices have drawn support for this hypothesis of extended cognition (HEC) from a dizzying array of empirical and theoretical resources including research in robotics and artificial intelligence, cybernetics, informatics, theories of developmental and dynamical systems, and theories of distributed representation. Such appeals increase philosophy's "coolness quotient." However, they also make it difficult to evaluate the plausibility of HEC without adverting to mere dogma or intuition. Against this background, *Cognitive Systems and the Extended Mind* offers a well-written, clearly argued, and sophisticated guide to precisely what is at stake in the defense of HEC. Sadly, it also demonstrates that many arguments that have been taken to support HEC can be easily accommodated with a conservative amendment to traditional, individualist cognitive science.

After a brief introduction, Part I of *Cognitive systems* (Chapters 2-4) attempts to establish three claims. First, Rupert offers a compelling argument to show that cognitive science is (at least implicitly) committed to treating cognitive processes as the mechanistic constituents of integrated systems; hence, only those features of the environment that contribute "causally to the production of a wide range of cognitive phenomena, across a variety of conditions, working together in an overlapping way with a variety of other mechanisms of similar standing" (p. 41) should be treated as genuinely cognitive. Second, Rupert argues that this systems-based approach offers the only plausible "principle of demarcation" that can distinguish between those features of the environment that make *genuinely cognitive contributions* to cognitive processes and those features that *merely cause* cognitive processes. More troublingly, he contends that no principle of demarcation that licenses HEC can provide a plausible foundation for cognitive science, and that proponents of HEC can explain the success of cognitive science at the cost of descending into "mere relabeling or the superfluous construction of extraneous systems" (p. 46). Finally, Rupert argues that considerations of conservatism and simplicity militate against amending cognitive science to include the inflationary claims of HEC. In short, Part I purports to establish that any defense of HEC will require offering a plausible principle of demarcation as well as positive evidence to demonstrate that HEC yields explanatory advantage beyond what is available from the perspective of traditional, individualist cognitive science.

The arguments in Part II addresses the most promising attempts to demonstrate that HEC can provide a plausible framework for research in cognitive science. Rupert first addresses arguments that are grounded in functionalist considerations about the multiple realizability of natural kinds (ch. 5). His response to these sorts of arguments is familiar—but powerful. Appealing to the fine-grained functional decompositions of the capacities of organismically-bounded entities that are commonly deployed within cognitive science can offers no *positive* support for HEC; moreover, for a functionalist argument to demonstrate the necessary continuity between extended and non-extended processes, it will be necessary to adopt an implausibly weak interpretation of cognitive functions, yielding a scientifically inadequate account of the cognitive states and processes of bounded individuals. In an important sense, arguments grounded in developmental systems theory (DST) and the role of linguistic scaffolding fare slightly better (ch. 6). DST establishes both 1) the possibility of extended individuals in cases where distinct organisms have a "shared fate" (p. 115), and 2) important respects in which language, tools, and the structure of our environment play a critical role in human cognition. However, Rupert contends that in the case of cognitive systems, such observations demonstrate only the existence of *embedded* cognitive systems that use their environment in various ways. Similarly, he maintains that, while arguments from dynamical systems theory provide a further reason for taking on an embedded and embodied theory of cognition, no current form of this approach yields support for HEC (ch. 7). Finally, Rupert addresses claims grounded in the phenomenology of situated and skilful coping and argues that such phenomenological considerations provide no support for HEC so long as we recognize, as we should,

that the mechanisms of cognition are not consciously accessible (ch. 8). In short, Rupert is at pains to demonstrate that each kind of argument that has been offered in support of HEC fails to provide sufficient explanatory and predictive benefits to justify abandoning, or even revising the traditional categories employed by cognitive scientists.

The arguments in Part III address the revisionary claims that are advanced by proponents of embedded and embodied accounts of cognition. Building upon the claim that mental representations are action-oriented, narcissistic, and deeply context dependent, proponents of HEC often suggest that it is a mistake to claim that cognitive states must be implemented neurologically. On the assumption that cognitive science, thus, relies on a mistaken view of mental representations, it is often argued that a unified—even if revisionary—strategy must be employed to explain the embodied, embedded, and extended nature of human cognition. Against this line of argument, Rupert counters that although many cognitive processes critically depend on external resources, there is no compelling reason to abandon individualist cognitive science. Through a careful analysis of a wide range of empirical data, he demonstrates that there is reason for a slight nudging rather than a coup, "a push toward different representational formats, perhaps computational architectures that both extract information from distinctively modal representations and use such representations as computational atoms" (p. 242). Thus, although we must attend to the causal impact of environment features on psychological processes, cognition is—where we always thought it was—in the head.

While Rupert is right that cognitive science attempts to explain goal-directed behavior in terms of integrated computational systems, his arguments leave open plausible strategies for defending HEC. No doubt, biological organisms tend to be systems in which various processes "contribute causally to the production of a wide range of cognitive phenomena, across a variety of conditions, working together in an overlapping way with a variety of other mechanisms of similar standing" (p. 41). Yet, the persistence of psychological capacities is not always best explained by appeal to the durability of the biological organism. One strategy that is left open by Rupert's arguments turns on the claim that some collectivities (e.g., navigation crews [Hutchins 1995], CSI teams [Barber et al 2006], and political campaign committees) can be seen as cognitive systems with cognitive capacities that persist as a result of the *integration of information* (Huebner 2008; forthcoming; in prep). In these cases, a variety of mechanisms, including human individuals and their "equipment," causally contribute to the production of collective behavior across a variety of contexts, by working together to encode, store, and manipulate collectivity-relevant information.

On a smaller scale, the functional specialization that spontaneously emerges in romantic couples often allows for the distribution of memory in a way that allows the couple to function as a transactive system. In experimental situations, romantic couples remember more items than unfamiliar pairs; but, when couples are forced to encode memories in ways that diverge from their familiar strategies for allocating information, they *remember significantly fewer items* than do unfamiliar pairs (Wegner et al 1991). More familiarly, when a partnership is terminated, this often leaves retrieval information in place, but no access to the relevant memories. This yields involuntary "feelings of knowing" as well as individual mistakes about what the group remembers. With these phenomena in hand, Dan Wegner and his colleagues (Wegner 1986; Wegner et al, 1991; Wegner, 1995) have argued that the integration of informational structures that are distributed across the members of a couple causally contribute to the production of an interesting range of couple-relevant behavior.

Of course, such claims will not persuade Rupert. Indeed, Rupert is right to note that an explanation of any collective phenomena in terms of embedded individual capacities will always be available. So, to defend the claim there are genuinely distributed cognitive systems, it is also necessary to provide answers to hard theoretical questions about the extent to which cognitive science ought to take the behavior of organisms as its primary *explanandum*. Rupert has done an admirable job of making it clear that any plausible defense of HEC will require accepting the pragmatist's gambit and arguing that cognitive science must begin by examining stable patterns of interrelated, socially significant practices, wherever they arise, inferring the existence of cognitive systems to explain the persistence of these patterns. Assuming that cognitive explanation must begin with individuals as they are embedded in wider computational systems "presupposes that there is a distinction between the practices and the instrumental network that embeds them" (Rowlands, 2009, p. 58). Perhaps this distinction is untenable (cf., Huebner, in prep)—but establishing that this is the case requires a much longer argument! Regardless, the key virtue of *Cognitive systems* is that it makes it clear precisely how hard the proponent of HEC must *nudge* traditional cognitive science.

References

Baber, C., Smith, P., Cross, J., Hunter, J., & McMaster, R. (2006). Crime scene investigation as distributed cognition. *Pragmatics and Cognition*, 1, 357–385.

Clark, A., & Chalmers, A. (1998). The extended mind. *Analysis*, *58*(1), 7-19.

Huebner (2008). Do you see what we see? An investigation of an argument against collective representation. Philosophical psychology, 21(1), 91-112

Huebner (forthcoming). Genuinely collective emotions. European journal for the philosophy of science.

Huebner (in prep). *Distributed minds and collective mentality.*

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

Rowlands, M. (2009). Enactivism and the extended mind. *Topoi*, *28*, 53-62.

Wegner, D. (1986). Transactive memory. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior*. New York: Springer-Verlag, 185-208.

Wegner, D., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, 61, 923-929.

Wegner, D. (1995). A computer network model of human transactive memory. Social Cognition, 13, 1-21.

Bryce Huebner
Department of Philosophy
234 New North
Georgetown University
37th and O ST NW
Washington, DC 20057, USA
Email: lbh24@geogetown.edu