Keeping HEC in CHEC: On the Priority of Cognitive Systems

Robert D. Rupert, University of Colorado, Boulder

According to the hypothesis of extended cognition (HEC, hereafter), human cognitive processing extends beyond the boundary of the human organism.[1] As I understand HEC, it is a claim in the

[1] For arguments in support of HEC, see Andy Clark and David Chalmers, "The Extended Mind," *Analysis*, 58, 1 (January 1998): 7-19; Clark, *Being There: Putting Brain, Body, and World Together Again* (Cambridge: MIT Press, 1997); "Reasons, Robots, and the Extended Mind," *Mind & Language*, 16, 2 (April 2001): 121-45; *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence* (Oxford: Oxford University Press, 2003); Mark Rowlands, *The Body in Mind: Understanding Cognitive Processes* (Cambridge: Cambridge University Press, 1999); John Haugeland, "Mind Embodied and Embedded," in Y. Houng and J. Ho (eds.), *Mind and Cognition* (Taipei, Taiwan: Institute of European and American Studies, Academia Sinica, 1995), pp. 3-37; Tim van Gelder, "What Might Cognition Be, If Not Computation?" *Journal of Philosophy*, 92, 7 (July 1995): 345-81; Daniel C. Dennett, *Kinds of Minds: Toward an Understanding of Consciousness* (New York: Basic Books, 1996); Ruth Garrett Millikan, *White Queen Psychology and Other Essays for Alice* (Cambridge: MIT Press, 1993), Essays 7-9; Robert Wilson, *Boundaries of the Mind: The Individual in the Fragile Sciences* (Cambridge: Cambridge University Press, 2004); Susan Hurley, "Vehicles, Contents, Conceptual Structure, and Externalism," *Analysis*, 58, 1 (January 1998): 1-6; Michael Wheeler, *Reconstructing the Cognitive World: The Next Step* (Cambridge: MIT Press, 2005). A wealth of work outside of academic philosophy has inspired HEC: Edwin Hutchins, *Cognition in the Wild* (Cambridge: MIT Press, 1995); Richard Dawkins, *The Extended Phenotype* (Oxford: Oxford University Press,

philosophy of cognitive science; HEC is supposed to reveal something important about human beings' core cognitive faculties and, in doing so, reshape the study of human cognition. Although many authors have reacted to HEC with reasoned skepticism,[2] HEC has able champions, Andy Clark among them. In a recent paper, "Curing Cognitive Hiccups: A Defense of the Extended Mind,"[3] Clark responds to various criticisms of HEC, targeting my concerns in particular.[4] In this essay, I clarify and expand upon my objections, arguing that Clark's defenses come up short.

1982); Esther Thelen and Linda Smith, *A Dynamic Systems Approach to the Development of Cognition and Action* (Cambridge: MIT Press, 1994); Merlin Donald, *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition* (Cambridge: Harvard University Press, 1991); Rodney Brooks, "A Robust Layered Control System For A Mobile Robot," *IEEE Journal of Robotics and Automation*, Vol. RA-2 (March 1986): 14-23.

[2] See Fred Adams and Ken Aizawa, "The Bounds of Cognition," *Philosophical Psychology*, 14, 1 (2001): 43-64; Gabriel Segal, "Review of Robert A. Wilson, *Cartesian Psychology and Physical Minds: Individualism and the Sciences of Mind*," *British Journal for the Philosophy of Science*, 48, 1 (March 1997): 151-56; Margaret Wilson, "Six Views of Embodied Cognition," *Psychonomic Bulletin & Review* 9, 4 (2002): 625-36; Rick Grush, "In Defense of Some 'Cartesian' Assumptions Concerning the Brain and Its Operation," *Biology and Philosophy*, 18, 1 (2003): 53-93; Keith Butler, *Internal Affairs: Making Room for Psychosemantic Internalism* (Boston: Kluwer, 1998), chapter 6; Daniel Weiskopf, "Patrolling the Mind's Boundaries," *Erkenntnis*, 68 (2008), 265–76.

[3] *Journal of Philosophy*, 104, 4 (April 2007): 163-92; all page references are to this paper unless otherwise indicated.

I. The natural-kinds argument and the Parity Principle

In "The Extended Mind," Clark and David Chalmers (C&C, hereafter) make the following claim:

> If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process ("The Extended Mind," p. 8).

This has come to be known as the Parity Principle (PP, hereafter). In "Hiccups," Clark accuses HEC's critics of founding one of their foremost concerns on a misinterpretation of PP (pp. 165-66).[5] The criticism at issue is focused on significant dissimilarities between internally realized cognitive states and those external states purported to be cognitive. As Clark sees things, we critics wrongly take PP to entail a fine-grained similarity between internal and external cognitive states, if any of the latter there be, which explains why we are put off by the many dissimilarities.

---

[4] As expressed in "Challenges to the Hypothesis of Extended Cognition," *Journal of Philosophy* 101 (August 2004): 389-428 ('CHEC', hereafter); "Innateness and the Situated Mind," in P. Robbins and M. Aydede (Eds.), *Cambridge Handbook of Situated Cognition* (Cambridge University Press, 2008), pp. 96-116; "Representation in Extended Cognitive Systems: Does the Scaffolding of Language Extend the Mind?" forthcoming in R. Menary (ed.), *The Extended Mind*.

[5] For a similar charge, see Richard Menary, "Attacking the Bounds of Cognition," *Philosophical Psychology*, 19, 3 (June 2006): 329-44, at pp. 339-40.

This diagnosis misses the mark badly. In the paper in question, CHEC, I explicitly set out to criticize C&C's (1998, pp. 13-14) argument from natural—or causal-explanatory—kinds. According to the distinctive premise of C&C's argument, cognitive science benefits from construing its causal-explanatory kinds in such a way that many external and internal states are of the same natural kind or instantiate the same theoretically relevant cognitive property. In response, I offered C&C a choice: characterize the relevant causal-explanatory kinds in terms of fine-grained functional properties (in terms of the reaction times they support, for instance) or instead opt for a coarse-grained conception of cognitive kinds ('generic' kinds, I called them). I argued that neither alternative in fact offers both *(a)* extended kinds (i.e., kinds that singly subsume both internally and externally realized states in a significant number of cases) and *(b)* a resulting causal-explanatory advantage; the first alternative is not likely to yield extended natural kinds at all (given the sorts of fine-grained properties of interest to cognitive psychologists), and the second alternative yields extended kinds unlikely to do substantive causal-explanatory work. Thus, the natural-kinds argument fails, by dint of a false premise, the one claiming that a HEC-friendly taxonomy provides a more powerful framework for cognitive psychology.

This debate about natural kinds (or properties) has nothing much to do with PP and, *a fortiori*, nothing much to do with a misreading of it. The discussion in CHEC does not presuppose that HEC entails fine-grained similarity of internal and external states (*cf.* Clark, p. 166-67)—the mistake that is alleged to follow if one construes PP in the erroneous manner in question. To interpret CHEC in the suggested manner would make a hash of the structure of its sections V-VIII, most obviously the discussion of generic kinds. Why would I have considered the possibility that C&C have in mind generic kinds if I were interpreting HEC to entail fine-grained similarity of inner and outer kinds? Moreover, if I had thought HEC entails such fine-grained similarities, my treatment of generic kinds would have been much different: I would simply have pointed out the "incoherence" of arguing for HEC by appeal to generic kinds. I did, however, no such thing. Rather, I argued that generic cognitive kinds fail to play a substantive causal-explanatory role in cognitive science, exactly in keeping with a consideration of the natural-kinds argument. It is most uncharitable, then, to interpret CHEC as offering a PP-related criticism of HEC.

Of course, to criticize an argument for position *Q* is not to criticize *Q*.[6] My attack on the natural-kinds argument simply returns the ball to the HEC-theorist's court. In response, the proponent of HEC might proceed in various ways. One approach would address my argument head on, by trying to show that cognitive science does, or will, benefit substantially from the adoption of an extension-friendly taxonomy of natural kinds. Below I return to some of Clark's suggestions of this sort. Alternatively, the HEC-theorist might instead set aside considerations of natural kinds and properties and develop a different style of argument for HEC—perhaps even an argument from PP. How promising is this latter tack?

In the remainder of this section, I argue that PP provides no sound basis for HEC, and for reasons that frame much of the discussion to come. The Parity Principle rests on the correct intuition that cognition is cognition, *wherever* it occurs; it *is* unjustified to demand arbitrarily that cognition appear only in certain, pre-specified locations. Yet, PP takes no account of independent reasons for locating the cognitive system in one place rather than another. For comparison, consider a lone neuron in preparation. It may be that, were the neuron in my head, it would have

[6] A point that seems to be misunderstood in Gary Bartlett, "Whither Internalism? How Internalists Should Respond to the Extended Mind Hypothesis," *Metaphilosophy*, 39, 2 (April 2008): 163-84, at p. 171.

the property *being part of my brain* (and such related properties as *being a contributor to large-scale neural processing*), and we would recognize it as such. Nevertheless, in the dish, it is not part of my brain. As an organ, the brain is a physically and functionally integrated system, and something can have the capacity to be part of that system without actually being part of it. The antecedent of this instance of a neural Parity Principle is plausibly true: if the neuron were in my skull, we would regard it as part of my brain. (Absent further information concerning how the neuron might have gotten into my skull, the most plausible chain of events in which the neuron lands in my skull is one in which the neuron has been physically and functionally integrated into my brain. Why else would it be there?) At the same time, the consequent is false: the lone neuron in preparation is not part of my brain.

This phenomenon is not limited to neural or cognitive contexts. For example, the same act of shooting a gun—performed by the same person in the same geographical location—can instantiate military properties or not, depending on the social and political context. Whether the shooting has military properties depends on whether it occurs as part of certain kind of social-political system: a war.

Thus, a generalized Parity Principle should be rejected outright. It is simply not true that, in general, a thing recognized as having P were it in one location therefore has P regardless of its location. Location does not matter *only if P is the sort of property that survives a change in location*. Lots of properties, however, do not, as a general rule, survive such changes. The Parity Principle is thus deeply uninformative. It warns us not to be biased by unexpected location, but this is sound advice only if location makes no difference; the location makes no difference, though, only if change in location is not correlated with change in the relevant status of the item or process in question. Thus, when applying a generalized PP, we can have confidence in PP's verdict only if we already know whether the property in question is the sort of property that survives a change in location; this requires knowing something about the nature of the property in question, and here PP falls silent, telling us nothing about the nature of cognition.

Consider how this concern applies in the cognitive domain. I argue presently that something is cognitive if and only if it is part of a persisting, integrated cognitive system; I suspect, too, that for humans this integrated system typically appears inside the boundaries of the body. Thus, it may be that some external states or processes would be recognized as cognitive were they in the head, but only because if they were in the head, they would most likely be integrated into the cognitive system. We might think, "Were that process located in the head, it would be part of the cognitive system," precisely because we have independent grounds for thinking the cognitive system is in the head and for thinking that, were the external process in the head, it would be incorporated into that integrated, internal system. This is all consistent with the process's not being part of the cognitive system when external, and thus not cognitive when located outside the organism. This reasoning does not demonstrate the falsity of PP, so much as it shows that we can have little confidence in its naive application. Depending on what features actually make something cognitive and how much we know about these features, our reactions to PP's counterfactuals might or might not track genuine cognitive status. Thus, in the absence of a theory of cognition, PP is untrustworthy; in the presence of a theory of cognition, PP is superfluous.

II. Supervenience bases and selection

Below I return to the question of cognitive systems. In this section, I consider two lines of argument "Hiccups" offers in defense of HEC. The first consists in an appeal to local mechanistic supervenience bases (p. 171). One of the goals of cognitive scientific explanation, Clark claims, "[I]s to display the machinery that underpins an agent's current mental state or that explains some specific cognitive performance" (p. 171). Given that some physical processes beyond the bodily boundary sometimes serve as supervenience bases of cognitive *explananda* and that these processes play a causal-mechanistic role in bringing about those *explananda* (or their realizers), such external processes should be considered genuinely cognitive; or so Clark concludes.

There is certainly something correct about this approach, but we should wonder why *all* such machinery constitutes the cognitive process in any new or surprising sense. Part of the physical machinery of perception is the object seen; and if the object is seen in natural light, the sun contributes causally as well. I suspect, though, that these are not the sorts of factors Clark wants to count as *local*. But why not? Clark needs some way to distinguish the *cognitive* contributors from the merely causal ones, such as the sun and the object seen, the causal roles of which are widely recognized by nonextended theories of vision.[7] Thus, although it makes sense

[7] *Cf.* Adams and Aizawa's observation that "the mere causal coupling of some process with a broader environment does not, in general, thereby, extend that process into the broader environment" (*op. cit.*, p. 56). Adams and Aizawa have since dubbed this, variably, the 'causal-constitution' or the 'coupling-constitution' fallacy, depending on the context of discussion (see Adams and Aizawa, "Why the Mind Is Still in the Head," in Robbins and Aydede, *op. cit.*, pp. 78-95, and *The Bounds of Cognition* (Malden, MA: Blackwell, 2008); *cf.* Ned Block, "Review of Alva Noë, *Perception in Action*," *Journal of Philosophy*, 102, 5 [May 2005]: 259-272).

Clark approvingly cites(p. 184) Hurley's explanation-based response to Adams and Aizawa. If an external factor explains cognitive phenomena, the idea runs, that external factor counts as part of the cognitive system. This tack is unsatisfactory, however. Simply because A to locate a cognitive process where its local mechanistic supervenience base is, the proposal is decidedly unhelpful in present context. The location of the cognitive *system* separates—from the top-down, one might say—the local mechanistic supervenience base of cognition from the other things that causally contribute in some important way to cognitive *explananda*. Independent cognitive facts determine which mechanistic contributors *are* the local supervenience base of a cognitive process, as opposed to factors, such as the sun, that merely causally interact with the local supervenience base of that cognitive process.

Putting the point epistemically, then, knowing what constitutes the local supervenience base requires *first knowing where the cognitive system is*. How best to think of cognitive systems individuation is, however, the question at hand. Thus, although I accept Clark's general point, it begs the question. It is not enough for Clark to observe that, with regard to cognitive phenomena, "the target performance depends upon a far wider variety of factors and forces than we initially imagined" (p. 170, footnote deleted). Dependences are simply too cheap; they are everywhere. Clark needs a principled way to separate the dependences that create cognitive systems from

---

explains B, even nontrivially, does not determine A to be part of B or part of a single A-B system of importance in its own right; nor does it confer on A properties that would normally be associated with B. At least in the right context (e.g., someone asks, "how could it have come about that people exist and write philosophical essays?"), the occurrence of the Big Bang contributes importantly to an explanation of my now writing a philosophical essay. Clearly, though, this does not give the Big Bang authorial or philosophical properties; nor does it entail the existence of a single system of interest to cognitive science: me plus the Big Bang. These concerns echo my earlier criticism of epistemic arguments for HEC (CHEC, pp. 395-96).

those that do not. Only when this is done will we know whether the local, mechanistic supervenience base of cognitive processing is extended.

Let us turn now to the second argument, which rests on a kind of interdependence that results from processes of selection, broadly understood (pp. 180-84, 190). Such processes, evolutionary or developmental, can cause the functioning of a human cognitive system to depend causally on the contribution of some resource typically present in the human's environment or at least typically present when the human undertakes the relevant task. In such cases, the current shape of the internal (e.g., neural) resources is to be explained partly by (a) the way in which the internal resources (or their "ancestors") causally interacted with external resources and (b) the success that such interaction engendered. Moreover, in many of these cases, selectional processes explain cognitively relevant properties of the external resources: these resources have changed, or the relative distribution of their variants has changed, so that members of the current "population" of external resources are, on average, more easily employed by humans than were their "ancestors." Language provides one of the clearest examples. Certain human cognitive abilities may develop against the backdrop of spoken guidance, e.g., the guidance provided when one talks oneself through a problem. A given human might solve a particular class of problem—say, certain kinds of mathematical problems—by providing her own verbal guidance, and this might be the result of the way in which she learned to solve that kind of problem. Furthermore, it might be that in learning to solve the problem in that way, the subject's neural resources became specially configured to handle the guiding verbal input normally used to solve these problems. What is more, the linguistic structures and practices employed (e.g., "first, do... [pause] then [with emphasis], do...") may persist largely because the human brain handles these structures and practices more efficiently than it handles alternatives. An interlocking system of internal and external resources results, according to Clark.

This relation of interdependence, deep though it may be, is subject to a straightforwardly nonextended interpretation: the organismically bounded cognitive system develops the capacity to use or interact with a certain kind of resource in order to achieve its goals. Quite naturally, the cognitive system "chooses" to use resources that are easier for it to use, and in some circumstances, this increases the likelihood that resources of that kind will persist or be produced at a higher rate than competing not-as-easy-to-use resources. This seems to me a Scotch verdict: two ways to interpret selection-created interdependence present themselves, and each seems conceptually adequate to a reasonable range of the relevant data. Why prefer one interpretation over the other? This is the stand-off that leads to Clark's hiccups, the standoff I hope to resolve in the remaining sections.[8]

III. Cognitive systems: HEC, HEMC, and the problem of demarcation

In this section, I present two arguments against HEC. These arguments rest on the availability of a nonextended approach, which, in CHEC, I dubbed 'the hypothesis of embedded cognition', or 'HEMC'. According to HEMC, the human cognitive system does not extend beyond the boundary of the organism, although during cognitive processing, the human exploits environmental objects and structures in surprising and extensive ways.

a. The conservatism-or-simplicity argument

[8] For critical discussion of selection-based and language-based arguments for HEC, see M. Wilson, *op. cit.*, and CHEC's (pp. 403-4) discussion of Clark's 'tailoring'-criterion, as well as "Innateness and the Situated Mind" and "Representation in Extended Cognitive Systems."

In "Hiccups" (pp. 188-90), Clark distinguishes between two of my criticisms of HEC.[9] First is the concern that organismically bounded cognitive systems play an important role in cognitive science, a role to which the typical extended system is not well suited. Clark tends to express this challenge as a worry about *selves* or *subjects* of study, but I would rather put it in terms of persisting, integrated systems (*cf.* CHEC, section IX). Human organisms are the persisting locus of an integrated set of cognitive capacities; organisms exhibit cognitive behavior across a variety of circumstances, and their persisting traits explain these regularities (including developmental regularities). In contrast, the extended systems of interest vary greatly in their constitution and, by some obvious measures (e.g., constitution by functionally relevant subparts), do not persist beyond the time of the interaction between organism and external resources. As a result, with regard to the phenomena of focal interest in cognitive science, patterns of similarity and difference in behavior across contexts cannot be explained by the persisting capacities of extended systems.[10] Second is the challenge posed by HEMC. If HEMC accounts for the results that impress advocates of HEC, the more conservative, simpler HEMC wins the day.

These two concerns operate most effectively in tandem.[11] Begin with the first point. Cognitive scientists have reason to be specially interested in the human organism; in many cases, regularities in organisms' behavior are the initial *explananda* of cognitive science and, moreover, appear to be relatively independent of materials with which the organism interacts. For instance,

---

[9] Beyond the criticism discussed above in connection with PP.

[10] M. Wilson (*op. cit.*, pp. 630-31) expresses similar concerns.

[11] In other words, the argument presented in the present subsection combines the two criticisms discussed by Clark (at pp. 189-90). The argument discussed in section III, b, below, constitutes an additional line of criticism.

a literate English speaker can read virtually any material written in English: books, pamphlets, fliers, and so on. Furthermore, it is the behavior of the organism itself that evidences such a capacity: the organism produces verbal or written output indicating comprehension of the material read. To the extent that different readers exhibit different capacities, this variation is best explained either by the past experience of the organism—for example, its experience with external linguistic products that are no longer present in the environment—or by the organism's genetically inherited characteristics. In cases such as this, both *explanandum* and *explanans* place the organismically bounded system in a privileged position: the organism (not the package organism-plus-environment) exhibits a pattern of behavior best explained by the properties of the organism (its persisting capacities, abilities, mechanisms, etc.).

The success of the organism-based approach suggests that the human organism, or some proper part of the human organism, houses a set of integrated cognitive capacities the regular operation of which explains regularities in the organism's behavior across cases and across variation in external materials. Notice that I do not say *the* system of integrated cognitive capacities; that would beg the question. At this point, I have argued only that there is *some* system of integrated capacities that is bodily bounded and the operation of which accounts for many important cognitive-scientific explananda. Whether this organismically bounded system is merely one part of a larger cognitive system remains to be seen.

Here my second concern comes into play. Having posited a set of integrated capacities, we seem to have in hand materials sufficient to explain the phenomena that motivate HEC. The properties of the very system posited to account for reading and the like—properties of the internally instantiated cognitive architecture—also account for the ways in which, in Clark's terms, the neurally centered cognitive system recruits external resources (see Clark's discussion of the hypothesis of organism-centered cognition, or HOC, at p. 192).

A competition results. On the one hand, there is HEMC, which takes much of cognitive processing to involve interaction between the cognitive system and external resources. On the other hand, there is HEC, which posits neural systems responsible for recruiting external resources and making them part of the cognitive system. Given the first point made above, however, it seems likely that the integrated, internal system instantiates something much like a standard cognitive architecture: if HEC does not include an image of this structure in its models, HEC will lose out on grounds of accuracy and explanatory power. Once, however, the HEC-theorist's model has been appropriately articulated—so that it includes the integrated internal architecture necessary to explain such organismically local behavior as reading—it seems likely that both HEC- and HEMC-theorists will offer structurally similar explanations of interactive cognitive processing, i.e., of the cases that motivate HEC. The difference, then, will be mostly in the labeling. According to HEMC, there exist an internal cognitive system, some external materials, and interaction between the two. According to HEC, there exist an internal system (which is cognitive), some external materials (also cognitive), and interaction between the two. Looked at in this way, HEC loses on grounds of conservatism: it is an uninteresting position that merely adds the label 'cognitive' to the external resources, the contribution of which is already taken into account. Alternatively, if HEC's distinctive contribution is to posit a unified system, a hybrid system, then HEC seems to lose out on grounds of simplicity. The HEMC-based approach explains all that need be explained without positing an additional system. HEC explains the phenomena by positing the same number of elements, the internal architecture, the interactive process, etc., then lumps these parts together under the label 'cognitive system'. This addition is gratuitous.[12]

b. A direct argument for the nonextended view

The preceding argument works in two steps: identify some system or other that contributes distinctively to the production of cognitive explananda, and then try to show that this system's activity suffices to explain all that needs explaining in cognitive psychology and cognate fields. A distinct tack insists that this construct is the *only* principled and plausible theoretical system on the table[13] and thus that inclusion in it determines what is cognitive and what is not. On this view, all it is to be cognitive is to be a state of a part of the persisting, obligate system (Wilson 2002) that produces the behavior to be explained by cognitive science. Here, then, is the Argument from Demarcation:

[12] If one's interpretation of HEMC recognizes the existence of a larger system only part of which is cognitive, then we return to the first case: HEC is an exercise in relabeling. If so, considerations of simplicity do not decide the issue, but considerations of conservatism do: mere relabeling does not constitute scientific progress.

[13] This claim to exclusivity is difficult to support absent an exhaustive survey of leading alternatives; for a more complete approach, see my forthcoming *Cognitive Systems and the Extended Mind* (Oxford: Oxford University Press).

Premise #1. The boundaries of the cognitive system mark the principled boundary between what merely causally contributes to cognitive phenomena and what is distinctively cognitive.[14]

Premise #2. The best (perhaps the only plausible) candidate for such a system is the persisting, integrated architecture.[15]

Premise #3. In the human case, the persisting, integrated architecture is, as a matter of contingent fact, internally instantiated.

Conclusion: Thus, contrary to HEC, the human cognitive system is internally instantiated.

As emphasized above, many things contribute causally to cognition. Inclusion in an integrated system, a cognitive architecture, marks the distinction between the sun as it contributes to visual processing in natural light and, for example, the set of feature-detecting mechanisms that work in concert to produce a serviceable image—fleeting, partial, and inaccurate though it may be.

c. Commentary on the arguments. Many fruitful research programs have produced robust results by placing human organisms in a variety of situations and recording their responses. A research program produces robust results when those results exhibit regular patterns observed across a

[14] Compare Adams and Aizawa's demand for a "mark of the cognitive" ("The Bounds of Cognition," p. 46). Adams and Aizawa do not, however, endorse the mark I suggest in the text.

[15] Compare a point made by Gabriel Segal: "Whole subjects plus embedding environments do not make up integrated, computational systems...the whole subject is the largest acceptable candidate for the supervenience base because it is the largest integrated system available" ("Defence of a Reasonable Individualism," *Mind*, 100, 4 [October, 1991]: 485-94, p. 492; see also Segal, "Review of Wilson," pp. 152-53).

variety of circumstances: such regularity can amount to (1) consistency in results when the concrete materials used vary; (2) consistent correlation between variations in results and variations in abstract properties of the materials used, a correlation insensitive to cross-cutting variations in the concrete properties of the materials used; or (3) regularity in the changes of patterns of response (where concrete materials used can vary at each stage) as a function of past experience or age of the organism. In cases where robust results are achieved (and, note, I am *not* claiming that they always are), the most straightforward explanation appeals to the reappearance of some organismically bounded system with persisting properties; these include higher-order properties of being likely to change first-order properties in certain regular ways as the result of aging or experience.

Visual perception provides a straightforward illustration. The organismically bounded subject can be placed in a wide variety of perceptual circumstances and will, highly reliably, have a sensory experience of the kind of object with which she interacts, regardless of its constituent materials or her particular view of it. If placed in front of a medium-sized chair in reasonably good light, she will report seeing a chair, regardless of whether it is made of metal, rattan, or tinker-toys. Why? Presumably, it is because the organism, or some proper part of it, instantiates certain properties persisting properties the effects of which explain why her reported perceptions are a regular function of what she casually interacts with in the environment.

Now consider the role of the persisting system in the investigation of human capacities for memory and language use. Such work normally presupposes that the subjects of investigation are persisting organismically bounded systems. Some such studies are explicitly longitudinal.[16]

[16] Harry Bahrick, "Maintenance of Knowledge: Questions about Memory We Forgot to Ask," *Journal of Experimental Psychology: General*, 108, 3 (1979): 296-308; "Semantic Memory

In other cases, the work is not longitudinal but is concerned specifically with the way in which an organismically bounded capacity leads to predictable results as environmental circumstances vary. Marsha Lovett, Larry Daily, and Lynne Reder[17] provide an elegant example of this sort of research. Working within the ACT-R framework,[18] Lovett et al. first run an experiment to estimate individual differences in short-term memory capacity—amount of source-activation, in particular. Then, using the source-activation value estimated for each subject, Lovett et al. make highly accurate, zero-parameter predictions of the performance of individual subjects on a significantly different experiment (in the first case, the modified digit-span, or MODS, task; in the second case, the *n*-back task). In a similar vein, Morton Gernsbacher and David Robertson[19] have found a significant correlation between capacities to understand narrative delivered in various forms: pictorial, verbal, or written. This demonstrates the persistence of a more general capacity to understand narratives, a capacity that operates in various circumstances on various kinds of material.

Content in Permastore: Fifty Years of Memory for Spanish Learned in School," *Journal of Experimental Psychology: General*, 113, 1 (1984): 1-29.

[17] "A Source Activation Theory of Working Memory: Cross-Task Prediction of Performance in ACT-R," *Journal of Cognitive Systems Research*, 1, 2 (June 2000): 99-118.

[18] John R. Anderson, Daniel Bothell, Michael Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin, "An Integrated Theory of Mind," *Psychological Review*, 111, 4 (2004): 1036-60.

[19] "Watching the Brain Comprehend Discourse," in A. Healy (ed.), *Experimental Cognitive Psychology and Its Applications* (Washington, D.C.: American Psychological Association, 2005), pp.157-67.

To take a more complex kind of example, consider the developmental trajectory of children's performance on the false-belief task, which appears robustly, across materials. Whether experimenters use puppets, stories, cut-outs, or live action, success rates on the task are consistent by age, as is the developmental trajectory.[20] Some factors, such as the inclusion of deception in the story, can enhance performance on the false-belief task. Nevertheless, the effects are of the depiction of deception, not of the particular materials used. Thus, the best explanation of this consistent performance across environmental changes adverts to certain persisting (but also maturing) capacities of the organism: its way of representing, and processing the representation of, deception.

Consider a rejoinder: The defender of HEC might respond that although the extended system does contain different *token* parts, an extended system persists across experiments because the external materials, even though they vary tremendously, instantiate the same abstract properties. Proponents of HEC should be unhappy with this tack, for at least two reasons. First, consider a contrast. In the case of organismically located capacities, there is a persisting basis, e.g., the brain, for the continued presence of the system's capacity. In the case of the proposed extended systems, however, there is no such basis of persistence. The various external materials might, on various occasions, instantiate the same abstract properties, but these various property-instantiations have no shared physical basis. The puppet show does not share any persisting physical basis with the story read, so the various instantiations of the abstract property *deception* do not appear to be part of a single persisting system. Thus, if the defender of HEC intends to

[20] For overview of the issues and meta-analysis of hundreds of experiments, see Henry Wellman, David Cross, and Julanne Watson, "Meta-analysis of Theory-of-Mind Development: The Truth about False Belief," *Child Development*, 72, 3 (May/June 2001): 655-84.

introduce a single system that persists and does the explanatory work across contexts that an organismically bounded system does, she can offer only a shrinking-and-growing cognitive system. This new metaphysics comes with no particular benefit; the explanation of growing and shrinking, i.e., of recruitment and discharge, can be expected to mirror the HEMC-based explanation of the relevant interactions.

Second, this defense of HEC runs strongly counter to most extant arguments in support of HEC; those arguments emphasize the specific material contexts of cognition: the specific form of the external resources and of the physical body, and the way those physical forms interact.[21] If the defender of HEC appeals to abstract properties instantiated by various stories, puppet shows, etc., she shifts our attention away from the specific material conditions of cognitive processing; and in doing so, she raises difficult questions concerning the organism's ability to track such abstract properties. One obvious proposal holds that the organism becomes sensitive to these abstract properties by representing them; this, however, invites standard internalist, or at least HEMC-based, explanation of the processing involved. If representations of the abstract properties appear in the internal system, why bother including as part of the cognitive system the abstract properties represented or their instantiations?

Note, too, that there is a natural explanation for the persistence of the capacities I have been emphasizing (in contrast to the situation for the sorts of reappearing, abstract properties discussed immediately above): they are physically realized, and the persisting organism provides their integrated, physical substrate; the organism as an integrated physical entity appears in the

[21] See, for example, Clark, "Language, Embodiment, and the Cognitive Niche," *Trends in Cognitive Sciences*, 10, 8 (August 2006): 370-74; "Material symbols. *Philosophical Psychology*, 19, 3 (June 2006): 291-307.

various circumstances of interest, and *its* persistence explains the persistent appearance of the integrated set of cognitive capacities realized by the organism. That there is a set of such integrated cognitive capacities explains why it is fruitful to subject the same physical organism to varied stimuli. If there were not an integrated, organismically bounded cognitive system, the robustness of results and fruitfulness of such research programs would be a mystery. If the operative system in these cases were an extended one, whose parts change as the organism interacts with various stimuli, we should expect our contrary assumption—of a system that persists across cases—to produce a hodge-podge of perplexing results. This is not, however, what we find in a wide range of research programs in various areas of cognitive science.[22]

Typically, the cognizer's integrated architecture—a set of primitive states (or representations) and primitive operations—undergirds such consistent behavior. Absent a

---

[22] In the text I have focused on cognitive psychology, but a similar pattern appears throughout the cognitive sciences. For example, Jon Driver and Jason Mattingly ("Parietal Neglect and Visual Awareness," *Nature Neuroscience*, 1, 1 [May 1998]: 17-22) describe a range of findings on parietal neglect that are robust and theoretically fruitful. Patients with right-parietal damage fail to register much of what comes into their left visual hemifield (particularly when they attend concurrently to stimuli in the right visual hemifield). Interestingly, these patients also neglect the left sides of objects that appear entirely in their right visual hemifield. As Driver and Mattingly describe the research, "This form of 'object-based' neglect has now been found in numerous cases, by separate research groups using various tasks...It also fits with findings of object-segmentation effects on visual attention in normal human subjects" (*ibid.*, p. 19, references excluded). This point appears to generalize to many aspects of the neuroscientific investigation of color vision, motion vision, and the what- and where-systems.

completed cognitive science, it is premature to pronounce on the precise form(s) of the human cognitive architecture and how exactly its elements are unified. It does seem, though, that the operation of such capacities as short-term memory, linguistic processing, and visual shape recognition are integrated with each other in a way that these skills are not integrated with external resources. When one reads, one draws on all three of these skills; when one listens to someone else talk, one draws on only the first two, unless listening partly involves processing the lip-movements of the speaker, in which case all three capacities are implicated. Typing involves all three skills, plus motor planning. Motor planning, short-term memory, and linguistic capacities are all active when writing on the chalkboard. And so on. In contrast, even where a particular kind of cognitive performance depends heavily on (i.e., virtually always involves the use of) a specific kind of external resource, that resource is relatively task-bound.

Consider, then, a more precise characterization of what it is for a set of resources to be integrated. Each token instance of cognitive behavior (alternatively, each act of completing a cognitive task) in a given subject involves the causal contribution of certain mechanisms, abilities, or capacities,[23] factors that make a causal contribution distinctively to the production of the cognitive *explananda* in question. Thus, for a given subject at a given time, there exists a set of mechanisms, capacities, external resources, etc. each of which has contributed distinctively to that subject's cognitive processing on at least one occasion. For each such type of mechanism, relative to each kind of cognitive phenomenon that it helps to produce, there is a conditional probability (determined by relative frequency, if by nothing else) of its use relative to each of the

[23] It is a difficult problem to identify precisely those conditions that metaphysically individuate kinds of mechanism, ability, capacity, etc. This is, however, a general problem; it matters not whether the mechanisms in question are internal, external, or hybrid.

other mechanisms, abilities, etc. in the set, as well as a conditional probability of its use relative to each subset thereof. These conditional probabilities can be rank ordered. Assume that mechanisms A, B, and C have contributed distinctively to cognitive processing in the subject in question at least once (but not necessarily to the same token process). Now assume that, given the history of the co-contribution of A, B, and C, $P(C \mid A\&B) = 0.7$; then the set {A, B, C} goes on the list as a 0.7 (relative to some particular kind of cognitive outcome) and will likely come in ahead of many other sets (for a particular subject at a particular time). The same set might also appear at a different place in the ordering because, for instance, $P(A \mid B\&C)$ does not equal 0.7 relative to the same, or to some other, kind of cognitive outcome. Moreover, given the variety of kinds of cognitive outcome, the same sets appear on the list many more times, most likely with many different associated probabilities. Next, consider the likelihood of a natural cut-off between the higher probabilities and lower ones, a gap that separates highly interdependent mechanisms from those that are less so. (In the absence of a significant gap in the ordering, 0.5 would seem to mark the relevant cut-off point.) Now count the number of times each type of mechanism appears on the list of sets with higher conditional probabilities (i.e., those sets above the significant gap on the list); these frequencies themselves exhibit a rank ordering, and a natural cut-off (another significant gap) separates those mechanisms that appear frequently on the list--i.e., are highly interdependent and heavily co-employed--from those that appear rarely. This indicates which mechanisms are parts of the integrated set to be identified with the cognitive system and which are, in contrast, resources used by the cognitive system.

This diagnostic measure does not isolate precisely what is intended. Certain consistent contributors to perceptual processing—the sun, for example—arguably earn a high score with respect to the typical human's cognitive behavior. Such resources work in a cooperative way with a variety of other mechanisms to produce a variety of cognitive outcomes. The concern about the sun, in particular, may be overblown; the sun is likely to be disqualified because of its nondistinctive role in producing cognitive outcomes. This seems largely accidental, however. The general pattern appears plausible enough: a particular gas in the atmosphere—unnecessary for life—may facilitate all chemical bonding in olfactory processing.

In response, one might reasonably appeal to an independent account of perception to screen off perceptual inputs from the integrated cognitive system: if a resource *R* achieves a high integration score simply on account of its effect of perceptual mechanisms, then *R* is struck from the list of integrated elements. The sun contributes to a wide range of cognitive outcomes but does so only by stimulating retinal cells, and so—assuming that our best independent theory of perception counts such cells as the peripheral perceptual mechanisms—the perceptual screening-off condition excludes the sun from the integrated cognitive system; and likewise were my hypothetical gas real. Presumably, the proponent of the extended view does not—no matter how enlarged the cognitive system is—wish to do away with distinctively perceptual mechanisms, that is, mechanisms that contribute to cognitive processing via their sensitivity to proprietary signals originating in the environment beyond the boundary of the cognitive system. Thus, the appeal to perceptual screening-off does not beg the question against the extended view.

Alternatively, an appeal to the role of representations may best explain the sense in which such resources as the sun are not part of the human cognitive system. Adams and Aizawa (2001, 2008) have emphasized—rightly I think—the contribution of nonderived representations to cognitive processing. As an absolute condition on cognitive processing, however, their view has proven controversial; even if nonderived representations play a central role in cognitive processing, the mechanisms operating on these representations may operate on nonrepresentational states as well, and do so as a practically indispensable part of the cognitive process. Thus, it will not do simply to say that, because the sun is not a nonderived representation, it is not part of the cognitive system. Nevertheless, if mechanisms can be individuated independently of their use of representations, we might consider the following combination of my position and Adams and Aizawa's: a mechanism is a candidate for inclusion in the integrated cognitive system only if it sometimes operates on representations with nonderived content. The sun—individuated physically as a star—does not operate on any nonderived representations (so as to contribute causally to the production of the explananda of cognitive psychology); thus, the sun is not a candidate component of a human cognitive system.

I suggest the appeal to representations, though, only to express a reservation about it. I suspect that a necessary condition on something's being a mental representation with nonderived content is that the structure in question be part of a system of representations and, moreover, that some collection of structures can be a system of mental representations only if it is being used by a cognitive system to produce cognitive phenomena (language-use, the behavior associated with inference, etc.). Thus, the appeal to the use of representations to individuate the cognitive system threatens circularity. It may be better, then, to focus on perceptual screening-off.

Niceties aside, I should emphasize that a mechanism's measuring up in the way I have described does not *constitute* its status as part of an integrated cognitive system. Integration is a theoretical kind, not a notion introduced by definition, and thus may not be subject to exhaustive reduction. Nevertheless, so long as a subject has a fair amount of experience in the world, this measure is, I submit, highly correlated with integration. Moreover, this measure allows, as it should, for developmental variation in a given subject's cognitive system as well as for variation among cognitive systems from one subject to the next and, on a larger scale, from typical members of one culture or society to typical members of another.

This approach to integration does not beg the question against HEC; it does not rule out, by principle, extended human cognitive systems. I suspect, however, that for the typical human, external resources either fail to appear in sets determined by higher conditional probabilities or, if they do appear, they fail to appear in very many of these, their being dedicated to the solution of specific problems. If this is correct, the typical human's integrated cognitive architecture is instantiated internally and the study of what little extended cognition there may be constitutes no revolution in cognitive science.

In effect, then, the direct argument identifies systemic integration as the mark of the cognitive, or at least a necessary condition of the cognitive: *a state or process is cognitive (if and?) only if it is part of an integrated system that produces cognitive explananda*. Proponents of HEC must explain, by appeal to principled grounds, why the sun is not part of the cognitive system, in contrast to other resources that contribute causally. The direct argument for the nonextended view offers such a principled story, and when this principle is applied, the nonextended view results. Think of this point a little differently. Computation and the flow of information (and, for that matter, dynamical interactions)—phenomena deeply of a piece with thought and representation—permeate the universe. The integrated architecture and its properties provide for the clearest and most plausible distinction between, on the one hand, cognition as a natural kind and, on the other hand, pandemic, sometimes cognition-related computation, information-flow, and dynamical interaction.

d. The dialectic. As Clark sees things, the HEMC-based approach places an "intuitive ban" on the inclusion of inputs as cognitive (p. 185); it rests on a "simplistic armchair vision" (p. 189); and it implies that cognition *must* stop at the boundary of the organism (p. 179; here Clark emphasizes the contrast, that it *need not* stop there). Clearly, though, the preceding arguments do not rest on brute armchair intuitions about where cognition must appear and do not place any absolute, armchair, or *a priori* ban on extended cognition. The arguments are driven primarily by contingent, empirical premises regarding what explanatory strategies have been fruitful in cognitive science, what an integrated cognitive architecture is, and where it is instantiated in the human case. Intuitions play some role: I take for granted that conservatism is a legitimate methodological principle. There is, however, nothing in the arguments or the related intuitions that elevates "anatomic and metabolic boundaries into make-or-break cognitive ones" (p. 192), at least not if "make-or-break" implies that the barrier is absolute or that some interest *in the barrier itself* drives the argument in favor of HEMC. Clark harbors suspicions of arbitrariness. He claims that "there seems no principled reason to suddenly stop the spread [of cognition] the moment skin meets air" (p. 180). Clark is right. There is no principled reason, if what that amounts to is the invocation of a metaphysically necessary boundary revealed by *a priori* philosophical intuition. If, however, by 'no principled reason' Clark means something much more general—something to the effect that no reasonable argument can be given for preferring HEMC over HEC—then I take the preceding discussion to show otherwise.

Notice, too, that my arguments depend in no way on there being a Cartesian Theatre or anything like it, in contrast to Clark's suggestion that HEMC depicts "bodily or outer resources as doing their work only by parading structure and information in front of some thoughtful neural overseer" (p. 191). In "The Extended Mind" (p. 17), C&C tentatively suggest that internal consciousness must validate the cognitive status of external states. In CHEC (pp. 404-5), I argued that such a view runs toward HEMC more than it does toward HEC. We must, however, keep the logic straight here. It is one thing to assume, as I did in CHEC, that if there is a privileged internal consciousness before which structure and information must be paraded in order that they be cognitive, then HEMC (most likely) wins the day. It is quite another to assume the converse conditional: if HEMC is true, there is a privileged internal consciousness before which structure and information must be paraded in order that they be cognitive. My criticisms of HEC in no way presuppose the second conditional, which I take to be false.

Lastly, consider a pragmatic point. Clark sometimes suggests that the adoption of anything short of HEC obscures the importance of the environment from cognitive-scientific view (p. 191). There is, though, no reason to think HEMC occludes the environment's contribution to human cognition. Quite the contrary: HEMC's expressed agenda is that cognitive science focus on ways in which the human cognitive system interacts with and exploits external resources; it will be an odd HEMC-theorist who ignores the role of the environment. Clark's concern would be more compelling were there actual cases in which the HEC-based perspective led to cognitive-scientific advances and where HEMC, had it been adopted in place of HEC, would have prevented these advances. So far as I can tell, though, the empirical research taken to support HEC was motivated not by a specific commitment to HEC *or* to HEMC, but rather by a general sense that interaction with the environment plays an important role in cognitive processing. Consider the way Dana Ballard and colleagues describe their project in one of the central empirical papers in the situated tradition: "Our central thesis is that intelligence has to relate to interactions with the physical world, meaning that the particular form of the human body is a vital constraint in delimiting many aspects of intelligent behavior."[24] This thesis entails neither HEC nor HEMC, yet it captures the approach of some of the most influential empirical work supposed to support HEC. Thus, I see no reason to think that, if, on the basis of independent arguments, we adopt HEMC instead of HEC, the context of discovery in cognitive science will be impoverished; HEMC leaves in place the emphasis on interactive processing—the primary theoretical vision driving the research claimed to support HEC directly.

IV. Clark's empirical challenge

The evaluation of HEC depends greatly on empirical work. Accordingly, Clark devotes much of "Hiccups" to the review of empirical results that he takes to bolster HEC. He emphasizes three kinds of result: the demonstration of cognitive impartiality by Wayne Gray and his colleagues, the work on gesture done (separately) by Susan Goldin-Meadow and David McNeill, and Chandana Paul's research in robotics. Notice that most of this work falls under the rubric of *embodied* cognition, i.e., the study of ways in which the fine-grained structure of the physical body, or the conscious experience of it, contributes to cognitive processing. As such, the research seems to be of the wrong sort to support HEC over HEMC. In fact, one might well construe it as bolstering HEMC, by showing what kinds of capacities, mechanisms, and skills the organismically bounded human cognitive system instantiates in virtue of which the human interacts effectively with the environment during cognitive processing. There may, though, be arguments for HEC here, so let us take a closer look.

[24] Dana Ballard, Mary Hayhoe, Polly Pook, and Rajesh Rao, "Deictic Codes for the Embodiment of Cognition," *Behavioral and Brain Sciences*, 20, 4 (1997): 723-43, p. 723; *cf.* CHEC's (pp. 393-94, footnote 9) discussion of the ambiguity of message in Kevin O'Regan's work on vision.

At one point (p. 180, n41), Clark attempts to place my position on a slippery slope from embodied cognition to extended cognition. Given that I do not advocate for a metaphysically necessary bodily boundary to human cognition, it is consistent with my basic view of cognition that it be extended. Nevertheless, I see no basis for a slippery-slope argument. As Clark runs the argument, since Goldin-Meadow and others have provided evidence of embodied cognition, it would be arbitrary to draw a line at the skin, excluding external cognition. The arguments of section IV, however, provide reasons to think human cognition does not extend beyond the boundary of the organism. Nature contains some boundaries, the edge of a lake, for instance; and sometimes we have reason to think the boundary is theoretically important: given the nature of their gills, fish cannot live beyond the boundary of the lake. In such cases, it is not arbitrary to claim that there is a boundary. If we understand the theoretical and empirical bases for drawing the boundary where we have, we should not think the boundary will shift; i.e., we have no reason to think a slippery slope awaits. It is no use asserting that there could, in principle, be fish that breathe outside of water; that does not a slippery slope make. Similarly, it does no good showing that cognition is embodied and that, in theory, it could be extended. Clark and I agree on both claims. Rather, to work a slippery slope argument, Clark must offer reason to think that my conditions for the existence of an integrated, embodied cognitive system are, given the empirical facts and the nature of these conditions, likely to be satisfied by extended systems. Clark's slippery slope argument lacks this essential element.

A closely related argument for HEC holds that the empirical work discussed by Clark at least shows us *how* cognition *might* be extended. The research does empirical spade-work; it opens our minds to the possibility of extended cognition. Perhaps, but this is a very weak point, given that all parties to the present debate accept the possibility of extended human cognition.[25] Thus, Clark's discussion of empirical work seems germane only to the extent that it illustrates genuinely extended cognition.[26]

In the remainder, then, I focus primarily on the work of Gray and his collaborators; for of the three examples Clark discusses, only Gray's work[27] bears directly on the evaluation of HEC. In a series of experiments, Gray and associates measure subjects' tendency to use internal memory, as opposed to accessing information encoded in external structures, by manipulating the relative time-cost of the use of internally and externally encoded information. The results manifest a regular relation: increase the cost of access to environmentally represented

---

[25] *Cf.* Adams and Aizawa, "Bounds of Cognition," p. 46; CHEC, pp. 393-93.

[26] Clark's report of Feynman's remarks (pp. 179-80) might constitute a further argument. If so, it seems best characterized either as inappropriate appeal to authority or an appeal to Feynman's verbal reports on his own cognitive processing. Regarding the latter possibility, I grant that first-person reports on cognitive processing are not entirely beside the point in cognitive science: it can be useful to collect problem-solving protocols or to ask subjects which memory strategy they experienced using when performing an experimental task. Nevertheless, self-generated reports are generally unreliable as descriptions of subjects' cognitive processing (as opposed to *being data* for which our theories must account); see Richard E. Nisbett and Timothy DeCamp Wilson, "Telling More than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, 84, 3 (May 1977): 231-59.

[27] See Wayne Gray, Chris Sims, Wai-Tat Fu, and Michael Schoelles, "The Soft Constraints Hypothesis: A Rational Analysis Approach to Resource Allocation for Interactive Behavior," *Psychological Review*, 113, 3 (2006): 461-82.

information, and subjects are more inclined to use internally encoded information. The cognitive system seems to "care" about only the cost of access to information, not about its location *per se*. Clark takes this to show that the external locations are part of the cognitive system.

Why, though, should we not take Gray's results to show that, when there is no great cost in terms of time, the cognitive system uses resources beyond its boundary? Clark's argument rests on the following premise: a system that uses resources beyond its boundary must (or at least is very likely to) treat the external nature of the location of those resources as intrinsically relevant to the decision whether to use those resources. I take this premise to be exceptionally implausible, a matter to which I return after making a pair of preliminary points.

First, we should not be misled by Gray and Veksler's characterization of the model: "The central controller makes no functional distinction between knowledge in-the-head versus in-the-world" (quoted at p. 172). This is an overstatement. Gray et al.'s full model *must* draw a functional distinction between the use of internal stores and external stores. The use of the latter requires the application of perceptuo-motor routines not required for the use of the former. For the two kinds of location to be treated differently—one accessed via perceptuo-motor routines, the other not—there must, by definition, be a functional difference between the role of the two locations. More accurately, then, the central controller does not treat the external store's being external as *in itself relevant* to the calculation concerning which resource to use.

The preceding observation suggests a second. When we consider what *is* relevant in the calculation, our attention is directed to the mechanisms by which the cognitive system gains access to internal and external information; for which mechanisms are used helps to determine the time-cost of such access. Use of the internal store need not involve the running of on-line perceptuo-motor routines, whereas external stores are accessed only via such routines. Perhaps the process of accessing external stores draws on the system's implicit knowledge of reliable sensori-motor contingencies: motor-command *p* is followed by a sensory experience that should be treated as the answer to a given query. Represented this way, though, there is no question of a competition between internal and external memory stores. Rather, there is a competition between the use of one internal store and the use of a distinct internal store. The computational process "chooses" between the retrieval of information from various internal registers: memory register A—part of, say, short-term, declarative memory—and register B—a visual buffer. In cases where B is chosen, the central controller "cares" only that the information shows up reliably in the sensory register, not where the information is in the external world. Thus, it seems plausible that both of the locations from which information is accessed are inside the organism, and the process of choosing between them has no bearing on HEC, except perhaps a negative one: by showing how the internal system chooses between two internal stores (standard memory versus information held in sensory buffers), the work of Gray and his colleagues shows how the organismically bounded cognitive system manages its interaction with, and exploitation of, environmental structures.

Perhaps the preceding point rests too heavily on a certain view of sensory input, one that proponents of HEC are likely to challenge. Set aside, then, questions about the internal sensory register, and let us return to my primary objection, to do with the ambiguity of Gray et al.'s results, for this worry persists even when we take at face value the externalist characterization of the location of information accessed via sensori-motor routines. Assume that the abstract representation of the computational process fails to assign, in terms of absolute privilege or specially strong weighting, a unique status to the internal resources, and thus is impartial in Clark's sense. Nevertheless, this kind of impartiality is also to be expected in cases where we contrast the use of internal and genuinely external storage. Take a cognitive system that sometimes uses resources external to it, where 'external' means beyond the system's boundaries, wherever they happen to be (no matter how far the system might extend beyond the boundary of the organism). Now assume the system uses some decision-procedure for selecting between the use of internal resources and genuinely external ones. Why should we expect the system to mark the difference between external and internal resources, assigning internal resources privilege simply in virtue of their status as internal? If there is no reason to expect this, then Gray's results alone provide no reason to prefer one of the two following descriptions of the human decision-making process: as involving (1) a choice between two systemically internal locations or (2) a choice between the system's internal resources and some external resources to which it can also gain access.

Clark has given us no reason to think that if there are two resources, one internal and another genuinely external to a given system, then if that system chooses between the use of these two resources, the system must (or is at least very likely to) treat the differential status of the locations as intrinsically relevant. Without this premise, however, Gray's results provide no support for HEC. Gray et al.'s demonstration is neutral in respect of two kinds of system—an extended system and one that is internal but sometimes chooses to use external resources—because both systems would be modeled in the same way: with a range of locations that are, in the abstract, treated on a par, as two possible locations the choice between which is determined by further considerations. In both cases, the representation of the relevant algorithm will, if it contains a representation of the external location at all (as opposed to a sensory register), not mark it as intrinsically second-rate: the various locations of information will simply be marked with subscripts or the like, so that they can be treated as distinct options when the decision-making computation is carried out. Thus, the sort of impartiality discovered by Gray et al.—even if it is understood in terms of a choice between internal and external locations—does not increase the likelihood of the extended, relative to the nonextended, hypothesis. Each hypothesis predicts that aspects of human cognition can be effectively modeled using an impartial algorithm. There is nothing in Gray et al.'s work to decide the issue, and so independent considerations must be brought into play. Such considerations, in the form of the arguments of section III, support the nonextended conclusion.

Above, I argued that Clark's slippery-slope argument misses the mark; the "slippage" premise goes unsupported. Nevertheless, given that Clark appeals specifically to research on gesture to support his slippery slope claim (p. 180), I close this section with two points about that research. One point concerns the *other* premise in the slippery-slope argument: the claim that research on gesture shows human cognition to extend in substantial ways to the extraneural body. The second point revisits my initial response to the slippery-slope argument, applying it in more detail, although with a caveat.

It is not clear to me that gestures themselves do the work Clark attributes to them. Any of the supposedly cognitive effects of gesture might just as well be effects of efferent copies of commands to gesture.[28] Clark emphasizes the following kind of result (p. 179). Give subjects a

---

[28] For discussion of efferent copy and the role it might play in cognitive processing, see Rick Grush, "The Emulation Theory of Representation: Motor Control, Imagery, and Perception," *Behavioral and Brain Sciences*, 27, 3 (2004): 377-396. Given the richly interconnected network by which neural systems maintain the body schema and make use of the information so maintained, it seems particularly plausible that efferent copies of motor commands affect other, ongoing cognitive processes, including speech, via neural routes; see, for example, Shaun

verbal or a spatial memory task. Then give them an intervening math problem to solve. While solving the math problem, notice which subjects use gesture and which do not. Then test all subjects on the original memory task. You will find that the subjects who used gesture while solving the intervening math problem perform significantly better on the memory task, whether spatial or verbal. According to Clark, this shows that the use of gesture is not merely a way of transferring the cognitive load to a neural subsystem specialized for physical-spatial cognitive processing; for, presumably, the use of that system would be required for the maintenance of the information that is required for subjects to perform well on the *spatial* version of the memory task. Instead, Clark concludes, the use of gesture plays a substantial role in cognitive processing itself.

Clark does not, however, explain the positive role of gesture. The gestures used during the intervening math problem do not seem to bear any particular relation to the content of either of the original memory tasks, and so we should wonder why gesturing improves performance on those tasks. Here is a possibility: the efferent copies, or other forms of collateral discharge, of the outgoing motor commands help to maintain high activity levels in the neural system, levels conducive to the stabilization of memories. This is purely speculative, but this is as plausible as any explanation Clark offers for the enhanced performance, and my explanation does not accord a cognitive role to the gestures themselves. On the view I have suggested, gestures themselves are epiphenomenal, although not in the way Clark considers and rejects: they are not mere expressions of fully formed thought. Rather, they are expressions of motor commands that

---

Gallagher, *How the Body Shapes the Mind* (Oxford: Oxford University Press, 2005), pp. 96, 124ff.

themselves contribute to problem-solving; the gestures and the salutary effects on problem-solving result from a common cause: outgoing motor signals.

A similar proposal covers cases in which what we might call the 'content' of gesture seems to contribute actively to problem-solving; for the specific content of efferent copies (at least insofar as their neural properties stand proxy for such content) can affect further cognitive processing. This gibes well with Goldin-Meadow's claim that verbal and gestural systems are driven by the same set of representations.[29] Given the ambiguity of the data, then, I do not take the research on gesture to demonstrate any especially striking kind of embodied cognition.

Nevertheless, it remains plausible that gesture itself makes a distinctively cognitive contribution. If so, does this ground a slippery-slope argument against HEMC, as Clark claims? Here I advocate the straightforward application of the systems-based approach, which appears to reinforce, not obscure, the line drawn by HEMC. Perhaps gesture itself contributes to the solution of a wide range of problems and so on, and thus qualifies as a component of the integrated cognitive system. This gives us no reason to think that extra-bodily resources achieve the same status. The possibilities must be dealt with case by case.

V. Conclusion

In closing, Clark says, "Both challenges have now been met" (p. 190). I think Clark's conclusion is based partly on a misconception of my concerns. Clark sees two distinct challenges where there is a single complex one. The persisting set of integrated cognitive capacities *is* the subject we are after; the existence of a persisting set of integrated cognitive capacities explains—partly via the positing of an architecture—why it has been empirically fruitful to proceed on the

[29] Susan Goldin-Meadow, *Hearing Gesture: How Our Hands Help Us Think* (Cambridge: Harvard University Press, 2003), chapter 2.

assumption that organismic subjects exercise their cognitive capacities across contexts. Recall that this has nothing to do with a Cartesian theatre or the like; the point of discussing the subject is to highlight the integrated set of capacities *as a causal-explanatory construct*, nothing more. Because this integrated set of capacities suffices to explain cases of interactive processing, the HEMC-reinforcing application of simplicity and conservatism stands; for we have been given no reason to think that HEC's explanation of the recruitment and discharge of external resources differs, in its essential structure, from HEMC's competing explanation, put in terms of cognitively relevant causal interactions between the organismically bounded cognitive system and its environment.

I have also articulated a further challenge to HEC. If there is a genuine phenomenon, cognition, we should want a plausible and principled way to circumscribe it. What strikes me as the only plausible proposal on the table adverts to the persisting, integrated set of mechanisms, capacities, abilities, etc. This is instantiated within the organism, in the human case. Thus, if there is cognition—over and above the mere flow of information, computation, and dynamical interaction that permeate the universe—it is, in the human case, contingently and mostly in the organism.