“Memory, natural kinds, and cognitive extension; or, Martians don’t remember, and cognitive science is not about cognition”

Robert D. Rupert, U. of Colorado, Boulder

Forthcoming in a special issue of *Review of Philosophy and Psychology*

I. The natural-kinds argument for the extended mind

The last quarter of a century has seen an explosion of experimental and theoretical work in what might generally be called ‘situated’ cognitive science (Robbins and Aydede 2008). Such work emphasizes the role of the organism’s (or artificial agent’s) interaction with its environment during cognitive processing (that is, processing that produces intelligent behavior). Early contributions to the field can be found in Rodney Brooks’s work in robotics (Brooks 1986), Esther Thelen and Linda Smith’s research in developmental psychology (Thelen and Smith 1994), David Kirsh’s exploration of epistemic actions (Kirsh and Maglio 1994, Kirsh 1995), Dana Ballard’s research on animate vision (Ballard 1991), and Tim van Gelder’s dynamicist philosophy of cognition (van Gelder 1995), among many other sources. Inspired partly by this burgeoning research program, Andy Clark and David Chalmers (1998) wrote a highly influential paper, titled simply “The Extended Mind.” In it, Clark and Chalmers advanced the bold hypothesis that the realizers of human mental states and cognitive processes are (frequently?) at least partly located beyond the boundary of the organism. They claimed that such a view has sweeping implications for the “methodology of research in cognitive science” as well as for our thinking about “moral and social domains” (1998, 18). The paper closes in dramatic pronouncement: “once the hegemony of skin and skull is usurped, we may be able to see ourselves more truly as creatures of the world” (ibid).

What might justify such a radical vision of human thought? Clark and Chalmers present various supporting arguments, but one holds special interest for those oriented toward cognitive science: their argument from natural (or causal-explanatory) kinds. After making their case for extended beliefs, via the now-famous example of Otto the Alzheimer’s patient who stores information in his notebook, Clark and Chalmers remark:

We do not intend to debate what is standard usage [of ‘belief’]; our broader point is that the notion of belief ought to be used so that Otto qualifies as having the belief in question....By using the ‘belief’ notion in a wider way, it picks out something more akin to a natural kind. The notion becomes deeper and more unified, and is more useful in explanation. (1998, 14; also see their comments about explanation at 9–10)

How does the discussion of Otto’s notebook bear on situated cognitive science? Here are three possibilities:

*A.* It provides a merely schematic model for genuinely naturalistic reasoning. The particular case may be somewhat far-fetched. And, yes, cognitive scientists may not seem to focus much on belief or belief-desire explanations.[[1]](#footnote-1) The suggested strategy is clear, however: argue for extended cognition by trying to show that our best cognitive science deals in natural kinds – either fine-grained or coarse-grained – that have both internally located and externally located instances, regardless of whether these kinds bear the names of, or are identical to, kinds discussed by the folk.

*B.* The discussion of Otto provides a model of how relatively pretheoretical conceptions of mental states interface with cognitive science. *A prioristic* reasoning about the coarse-grained causal roles of folk mental states, together with everyday empirical facts about the causal roles played by certain things in the environment, strongly suggests that some human mental states are externally located. If successful research programs in cognitive science reinforce the judgments yielded by suchargumentation, this helps to show that the folk have latched onto genuine natural kinds with their use of ‘belief’, ‘desire’, etc. Folk descriptions of these kinds may be more coarse-grained than descriptions of typical cognitive kinds, so the folk kinds may not be cognitive kinds, but perhaps they occupy a higher-level domain that bears a systematic relation (say, of supervenience) to the cognitive, and this explains why consilience at the cognitive level helps to confirm folk psychology. One might here add a stronger claim on behalf of the folk categories: they should direct cognitive scientific enquiry, shaping its investigations, and, in some sense, constraining how it is allowed to develop or which theories are to be taken seriously by cognitive scientists.

*C.* The thought-experiment involving Otto provides an argument for extended mind that has nothing to do with cognitive science or a naturalistic outlook. On this view, any use of the language of natural kinds – at least as I understand them (see note 3, below) – is misplaced. On this approach, the nature of mental states is revealed by introspection and conceptual analysis. Cognitive science – with its contrived data and fancy tools for modeling – investigates a wholly different domain, one independent of folk psychology, although, for rhetorical purposes, it might be useful to offer arguments for the extended mind and for extended cognition side by side.

In what follows, I’ll focus on A. I find this version of the natural-kinds argument to be of special interest, not only because of its potential connection to the philosophical foundations of cognitive science, but because I’m inclined to think that the methods of the natural sciences, broadly construed, are more likely than are other methods to yield knowledge about mind, self, and cognition (if these things exist!).[[2]](#footnote-2) If we would like to know, for example, whether mental or cognitive states extend beyond the boundary of the human organism, we should ask where things stand with regard to our best or most promising scientific theories of mind and cognition. Thus, in my view, some version of the argument from natural (or causal-explanatory)[[3]](#footnote-3) kinds offers more promise than any other argument that’s been given in support of the extended view.

As a final preliminary point, it might be worth remarking on the connection between Clark and Chalmers’s thought-experiment involving Otto and the topic of memory. Otto’s case involves (dispositional) belief, and thus my use of it to frame questions about memory and cognition might raise eyebrows. Bear in mind, though, that Clark and Chalmers describe Otto’s notebook as playing the role of “biological memory” (1998, 12) and throughout their paper repeatedly treat Otto’s nonoccurrent belief as a kind of memory or as analogous to it (13, 15–16; and see Clark 2008, 76). This places the discussion squarely in the realm of cognitive science, so long as we take Clark and Chalmers to be talking about memory as it – or what we’re inclined to categorize as memory-related behavior – has long been an object of scientific study. So, it does not stretch connections too far to move from natural-kinds reasoning, as applied to the hypothetical case of Otto, to a discussion of memory, then to cognitive science’s search for natural kinds.

*II. Natural-Kinds reasoning meets the Parity Principle*

In previous work, I pursued this memory-based tack, evaluating the natural-kinds argument by asking whether the science of memory supports a specific instance of the argument; I concluded that, at least so far as one can generalize from this test case, the natural-kinds argument for cognitive extension fails, and does so in an instructive way (Rupert 2004, 405–424).[[4]](#footnote-4) The argument was widely criticized by defenders of extended cognition – partly, I think, because it was widely misunderstood. Below I offer what I hope is useful diagnosis and elaboration but, first, a more careful presentation of the natural-kinds argument and my response to it:

*The Natural-Kinds Argument for Cognitive Extension*

*Premise 1*. If the most explanatorily powerful (known) framework for theorizing in a given domain presupposes a given taxonomy of kinds of states, we should at least tentatively accept the existence of states of the kinds in question.

*Premise 2.* The most explanatorily powerful (known) framework for theorizing about intelligent behavior presupposes kinds that, in fact, have a significant number of instances external to the organism.

*Conclusion*. Therefore, we should at least tentatively accept the extended view of human cognition.

In response, I argued that premise 2 of the natural-kinds argument falls to a dilemma: either the proponent of cognitive extension individuates the relevant causal-explanatory kinds in a fine-grained way or in a coarse-grained, or generic, way. With respect to the first horn, I argued that we shouldn’t expect repositories for external memories to exhibit the fine-grained properties and dynamics (e.g., conversational dynamics) of interest to working cognitive scientists. *Ergo*, our most powerful framework for explaining intelligent behavior in humans – at least if we limit our focus to fine-grained properties and explananda – does not seem to presuppose states that have both biologically internal and biologically external instances. My discussion of the second horn was less extensive, but the essential worry was this: characterizations of generic kinds of the germane sort (ones likely to have external instances) would be so thin as to rob them of causal-explanatory power (one side-effect here being problems of cognitive bloat) (Rupert 2004, sections V-VIII, and for the two horns explicitly presented side-by-side, see *ibid*. 407, 418-19, 424).

The argument drew many responses; some misunderstood the dialectical purpose of the argument, taking it to be a direct attack on the extended view, while others missed the dilemma-structure of the argument, thinking I was somehow on about fine-grained differences only – that I took fine-grained similarity of inner and outer to be necessary either to the extended view or to the natural-kinds argument – while others thought I was on about the Parity Principle (rather than the natural-kinds argument) (Clark 2008, 112–115; Menary 2006, 331, 333, 334, 339–340; Rowlands 2009,3; Bartlett 2008**,** 171; Sprevak 2009, 506; Levy 2007, 58–59; Adams and Aizawa 2008, 13). Much of the remainder of the present essay is an attempt to straighten out the relation between the natural-kinds argument and the status of memory and cognition as genuine scientific kinds – whether fine-grained or coarse-grained – and how all of this relates to the supposedly extension-supporting role of the Parity Principle.

Early in their paper, Clark and Chalmers assert, “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” (1998, 8). This came to be known as the Parity Principle, and on one way of interpreting it, it simply asserts that if something’s cognitive, then it’s cognitive, regardless of where it sits; as such, it is a truism. If it has any value, it is as a warning against bio-chauvinistic prejudice.

I agree with Wheeler (2011, 419–420) that this bare, anti-prejudice reading sheds no light on the nature of cognition or on the boundary between what is cognitive and what is not (Rupert 2009, 30–35; Walter and Kästner 2012); instead, it invites the formulation of a location-independent theoretical account of cognition (for an account in this vicinity, see Rupert 2009, chapter 3; 2010). The Parity Principle itself is of no direct use in this process. Beyond the truism that something is what it is regardless of where it’s located (provided the change in location doesn’t change what it is – Rupert 2009, 33), the Parity Principle seems to ask us to consult pretheoretical intuitions in order to identify a natural, or scientific, kind (*ibid.* 32; 2010, 345–346). Such intuitions may be indispensable when as a means of identifying the *explananda* of cognitive science, but they have little place in the construction and interpretation of theoretical accounts of the *explananda*, and it is here where the natural kind *cognition* (be there such a kind) has its home; cognition is supposed to be the process that produces intelligent behavior – not the behavior to be explained – and the nature of that process should be discovered in the standard way, by the causal-explanatory theorizing and experimentation distinctive of the empirical sciences.

There is another way to gloss the Parity Principle, which might be called a ‘benchmarked’ interpretation (Wheeler 2010, 254; 2011, 418).[[5]](#footnote-5) On this way of understanding and applying the Parity Principle, one locates an uncontroversial case of a cognitive state or process, then commits oneself to fairness: any place where *that* kind of state or process appears, it is cognitive, even if it’s external to an organism. This jibes well with Clark and Chalmers’s discussion of Otto (as Clark acknowledges – 2011, 451 – even though he in fact prefers the nonbenchmarked reading of the Parity Principle).[[6]](#footnote-6) They compare our man Otto to a counterpart, Inga, who has intact biological memories; Otto’s notebook-based memories play the same coarse-grained roles in the production of behavior as those played by Inga’s bio-memory, and on this basis – together with the fairness enjoined by the Parity Principle – Clark and Chalmers conclude that Otto has an extended belief.

Thus, the Parity Principle can be applied in a nonbenchmarked or a benchmarked way.[[7]](#footnote-7) The Natural-Kinds Argument can be understood in two parallel ways. Consider, on the one hand, a benchmarked reading of the argument. On this view, natural-kinds reasoning requires that the kinds in question have both a significant number of known internal instances, in addition to a significant number of external instances; this is suggested by Clark and Chalmers’s talk of explanatory unification and more deeply unified kinds and by their comparison of the causal role of Otto’s notebook to the causal role of the relevant part of Inga’s brain. (This is how I interpreted the natural-kinds reasoning in Rupert [2004], which suggested, to those focused on the Parity Principle, that I was gunning for a benchmarked version of the Parity Principle.) Perhaps the benchmarked presentation of natural-kinds reasoning is an artifact of Clark and Chalmers’s concern about folk mental states; if the point is to convince readers that mental states, as the folk understand them, sometimes are externally realized, they must convince readers that the external is on par with the internal – given that, for the folk, internal states provide the paradigmatic mental states.

On the other hand, the Natural-Kinds Argument might be given a nonbenchmarked reading. The nonbenchmarked version of the Natural-Kinds Argument’s Premise #1 would say something along the following lines: “If our best account of the distinctive drivers of intelligent behavior invokes natural kinds that have instances beyond the boundary of the organism, then so be it.” On this reading, the argument makes no commitment to the existence of natural kinds that have both internal and external instances; it presupposes no internal benchmark at all, not even a human one. In some sense, there’s nothing controversial here, but when we turn to Premise #2, we might begin to lose our bearings. Regardless of whether we’re talking about fine-grained or coarse-grained kinds, we might expect our kinds to be based in the first instance on what we know about humans. How would we identify cognitive kinds that have no internal instances in known cases, and what kind of cognition do we know much about, except human cognition?

Here’s where things stand, then. In the search for cognitive kinds, including cognition itself, we ask whether or not we should rely on a benchmark. If we appeal to a benchmark, we then ask whether, in the benchmarking category (e.g., among humans), fine-grained or coarse-grained/generic kinds do causal-explanatory work sufficient to secure their status as natural kinds. If we forgo a benchmark species or type of agent, we must decide how to frame our project, that is, decide what measure of success in a causal-explanatory project might secure nonbenchmarked natural kinds such as *memory* or *cognition*.

I take it that, benchmarked to humans, fine-grained *memory* fails to satisfy both extension-friendly desiderata: (a) does sufficient causal-explanatory work to qualify as a natural kind and (b) is likely to have enough extended instances so as to alter the way we do cognitive science or our fundamental conceptions of ourselves as thinkers; this is the upshot of the discussion in Rupert (2004). I take it that *generic memory*, benchmarked to humans, is in the same boat; although, I supported this judgment with some reasons – including concerns about what has come to be known as cognitive bloat (*ibid*., 421) – this is more controversial, and I return to this issue below, in section IV. There, I also ask how we might identify natural kinds with only (or primarily) extended instances. In particular, I examine the nonbenchmarked approach and to related questions about the standards for identifying natural kinds.

*III. Methodological interlude: Operationalism, natural kinds, and folk kinds*

How do the sciences home in on nature’s kinds and properties or select the language to be used in the formulation of their most successful theories?[[8]](#footnote-8) The naturalist’s way (Quine 1969b, Quine and Ullian 1970) to answer this question would be to examine what appear to be our most successful scientific theories, checking to see how they were formulated and chosen over competitors. In this section, I focus on the circumstances surrounding the emergence of a new science, when we are faced with what we think is a distinct domain of phenomena the investigation of which demands introduction of new vocabulary, concepts, principles, and laws.

In the present context, the most pressing questions are “what data is the introduction of ‘memory’ and ‘cognition’ *as theoretical terms* meant to help us to systematize or explain?” and “why think there are scientific kinds *memory* and *cognition*?” Answering these questions is no simple matter. Complications arise at every turn – from the identification of data, to the attempt to describe what seems to unite them in one domain of enquiry, to the characterization of the natural kinds supposed to play a substantive role in our best theories of those data.

So far as I can tell, successful scientific (sub)disciplines take shape largely in response to some range of observable phenomena that (a) are unexplained (not currently subject to prediction/retrodiction or manipulation and not currently subsumed under a general theory the principles of which have garnered scientific support from circumstances of successful prediction/retrodiction and manipulation) and (b) seem similar enough such that when they are accounted for, we suspect it will be by some common set of basic principles, theories, or models. What makes us think that given collection of phenomena are of a theoretical piece? It may be partly because we deploy the same everyday terms (e.g., ‘remembering’ or ‘intelligence’) in our descriptions of them or, in some cases, because we have had limited success predicting and explaining those phenomena using the same proto-theory or folk theory (for discussion of one relevant case, that of a folk theory of mind, see Gopnik and Wellman 1992). Early thinkers in the field then propose models of the mechanisms that produce these phenomena or hypothesize properties the interactions among instances of which account for the observed phenomena. The target of such modeling is likely, especially in the early days, to be a small subset of the phenomena in the domain. Depending on the degree of success such modeling meets with, a given model, theory, or collection of models might be generalized to account for a wide range of the initial phenomena and, perhaps, in a flash of insight, to account for phenomena that weren’t thought to be of the same type as the phenomena to which the model in question was, initially, successfully applied; additionally, if an otherwise powerful model or family of models does not account for one of the phenomena originally thought to be in the relevant domain, we may expel that phenomenon from the domain. Thus, begins a reciprocal dance: we recategorize phenomena into new similarity-groups based on the similarity of the models (the properties invoked by the models and actual relations claimed to hold between elements in the models) that can successfully account for at least some of them, while the constitution of what seem to us to be current similarity groups (including second-order similarity groups) guides our search for more general models that unify a broader range of phenomena. Along the way, relations of overlap between domains are discovered, single domains split into multiple ones, and the place in the enterprise of everything from individual bits of data to entire phenomena can be reconceived.

The preceding description of the birth and development of science goes all too quickly, running roughshod over the wealth of nuance historians of science have unearthed in recent decades. Nevertheless, this brief description is, I think, accurate as far as it goes, and it makes certain of my commitments clear. Two are of special importance.

The first concerns operationalism: nothing in my sketch of the birth and development of a science entails operationalism of the sort reviled in philosophy of science; in fact, quite the contrary. According to the operationalist view, a theoretical term expresses nothing more than a claim about the results of measurement-operations that guide the application of that term: if we measure temperature by use of a mercury thermometer, then ‘temperature of system x’ simply means ‘the readout on a mercury thermometer after it’s been placed in system x in such-and-such way for n units of time’. Thus, operationalism yields structured sets of sense-data or of observations sentences as the meanings of such terms as ‘temperature’, ‘charge’, ‘spin’, as well as ‘remembers that’. In order to incorporate distinctively theoretical statements (which undoubtedly play a role in science), these operational meanings might be thought of as inference tickets to move from certain observation sentences (“the thermometer reads n at t”) to certain theoretical sentences (“the temperature of the system in question is n at t”), the latter of which can be plugged into formulae that license further claims about the results of measurements (“the thermometer will, at t+10, read n+5”).

Although philosophers of science have rejected such orthodox operationalism, its spirit lives on. Many philosophers and scientists remain suspicious of theoretical posits and continue to suspect that terms for theoretical kinds or properties amount to nothing more than a way of categorizing observations or observable phenomena, for instance, that the property of having a memory that P is nothing over and above the subject’s exhibiting certain behavior (that is, the behavior we would normally categorize – for practical purposes – as expressing one’s memory or acting on one’s memory that P). For some, this persisting intuition takes a more liberal form, according to which a term for a given theoretical property or kind applies to a given state if and only if that state is now producing or has before produced the phenomenon of interest.

My sketch of the birth and development of a scientific (sub)domain supports none of these operationalist intuitions, neither the strictly operationalist, nor the more liberal vaguely related, ones. I hope to discover the nature of memory and, more broadly, the nature of cognition, by identifying the privileged theoretical constructs that emerge from the ongoing interplay described above; such an enterprise, even when embellished with all of the nuance of real-life practicing science, typically (perhaps always) deals in theoretical constructs that are not understood simply as whatever, in actual cases in which observable phenomena of interest occur, produced those phenomena – and which are *surely* not to be equated in any way with the results of specific measurements. Subjects can have thoughts and memories never voiced or acted upon and the ways of collecting data to try to characterize what it is to have thoughts and memories need not require the production of the primary phenomena to be explained. For example, neuroscientific evidence might help to verify the occurrence of a memory-forming process (because we’ve detected the kind of neural activity hypothesized by our best theory of the production of memory-related phenomena to occur when memories are formed) in a case in which the subject never exhibits any memory-related behavior. (The preceding example is not meant to privilege neuroscience; if the proponents of an extended view of memory are correct, then we might someday have evidence that a given subject has a behaviorally inert memory because we have evidence of the occurrence in her of an at least partly external process that normally occurs when subjects form extended memories.)

The second point is this: the picture of science sketched above leaves plenty of room for a contribution from the folk, even iffolk psychological terms do not refer to natural kinds (that is, kinds that will be the subject matter of our mature scientific theories). Such folk terms as ‘remembers’ and ‘memory’ (and ‘learned’, and ‘meaning’, and ‘concept’ – cf. Machery 2009) may well be of use in framing the phenomena originally thought to be of a piece, and thus worthy of investigation as part of the same organized enterprise, even when they ultimately turn out to be bankrupt as kind terms. Moreover, beyond the early grouping together of phenomena, these terms might continue to be used as a matter of convenience – for example, to provide a convenient way to refer to multiple, distinct kinds or to refer to the various products of historically related research projects; in this way, the terms serve something more like a sociological purpose than a purpose internal to the relevant scientific enterprise itself. (Analogy: One might wish to refer to all of the innovations that came out of Bell Labs – perhaps as a way of commenting on the culture at Bell – and thus talk conveniently of the “Bell Labs Research,” without thinking the research thus issued exhibits any deep natural unity.)

*IV. Memory, generic memory, and Martian memory*

Let us return, then, to the question of benchmarking. If (what we judge relatively pretheoretically to be) intelligent human behavior serves as our *explanandum* and we benchmark *memory* to the products of cognitive-scientific theorizing about this *explanandum*, then, I maintain, *memory* simpliciter is not a natural kind. Memory in humans is a widely disparate phenomenon (Tulving 2000, 41; Michaelian 2010), and different memory systems are likely to be at least as different from each other, in ways that matter to cognitive-scientific explanation, as they are from other systems that are not associated with memory (compare Machery’s [2009] eliminativism about concepts). Thus, benchmarked versions of the Natural-Kinds Argument (at least those that appeal to memory research) seem to fail. Rupert (2004) shows that benchmarking to fine-grained states undermines the argument’s second premise, and the widespread acceptance of the disunity of memory shows that the generic strategy, benchmarked to humans, fails as well.In the remainder of this section, I consider attempts to show that generic kinds *memory* and *cognition* help to explain human behavior scientifically, even though these kinds aren’t necessarily benchmarked to humans. Given that natural kinds can, in principle, be either fine-grained or coarse-grained (Rupert 2004, Sprevak 2009, 512), it seems reasonable to try to find a role for coarse-grained, or generic, kinds in the most developed cognitive science we have – the study of human cognition – and given the liberal attitude normally associated with the extended view, it’s also natural to try to do so without a commitment to benchmarking.

Wheeler offers the following argument for the scientific legitimacy of generic kinds. He asks us to consider the discovery of someone whose inner mechanisms related to “context-sensitive information storage and retrieval” do not exhibit the standard fine-grained human causal profiles. He claims that cognitive psychologists would treat this as “one possible form of the psychological phenomenon of memory” (2010, 258), and that this establishes a genuine role for the kind *generic memory* in cognitive science. This approach is benchmarked in a way, because it’s built into the hypothetical case – as one aspect of it that generates Wheeler’s intuitive response – that the subject behaves in a way that is sufficiently similar to the standard human profile (or, at the very least that the subject is a member of the same species and so should be treated, by default presumption, as having the same internal kinds as other humans). That, however, is not the central issue. Ultimately, the question is whether generic memory plays a role explaining intelligent behavior. Wheeler claims that his hypothetical case establishes the explanatory credentials of *generic memory*, by showing that *generic memory* does “important work in organizing and shaping the project of cognitive-scientific explanation” (2010, 258). It is, however, one of the primary theses of this essay that doing such work does not suffice to establish something’s credentials as a genuine scientific kind (particularly if one takes a liberal attitude toward explanation, but even if one does not). Whether Wheeler’s prediction is correct, that scientists would call this subject’s processing ‘memory’, is itself neutral with regard to the question of kinds.

Let us now come at the issue from a different angle, from the perspective of someone in the grips of the so-called Martian intuition. Here is Mark Sprevak’s presentation of the Martian intuition as it pertains to the current debate. “[Rupert’s and Adams and Aizawa’s] objection to HEC [the hypothesis of extended cognition] is that fine-grained features of human cognition are necessary for mentality. But this seems wrong. Martians could differ from us in all kinds of fine-grained psychological ways and still have mental states. Therefore, such features are not necessary for mentality” (2009, 509). Although this doesn’t quite get my objection right (see above), it does take us quickly to the heart of the matter. If one attends only to human-benchmarked, fine-grained natural kinds, then the Martians in question simply don’t have memories! And that’s preposterous! So, either one should not focus on fine-grained states or one should not benchmark, or both. (But this is not Sprevak’s ultimate conclusion; see below.)

So far as I can tell, though, it is not preposterous to deny memories to our hypothetical Martians. In fact, it seems quite plausible. Assume that Martians exhibit the same kind of behavior we do but produced by mechanisms quite unlike ours. What theoretical grounds might there be for claiming that Martians have memories? In principle, there seem to be only three avenues of approach: *(1)* establish that Martians have states that are, from a fine-grained perspective, similar enough to human benchmarks; *(2)* argue that generic *memory* has a causal-explanatory role to play – either *(2a)* in the human case alone (in which case, Martians might have memories simply because they have coarse-grained states similar enough to human generic memories) or *(2b)* in the combined case, by showing that *generic memory* plays a causal-explanatory role when the human and Martian cases are pooled together, even though it doesn’t play such a role in the case of humans alone; or *(3)* argue for a nonbenchmarked identification of *generic memory* as a kind that plays a role in Martian psychology but not in the human case.

Possibility *(1)* is ruled out by hypothesis. Possibility *(2a)* can’t be ruled out entirely, but it seems to me to that the non-kind-supporting, practical explanation of the continued use of ‘memory’ is far superior to the claim bruited in *(2a)*.[[9]](#footnote-9) Moreover, if one were to pursue *(2a)*, one would have to take a different tack from the appeal to Martian intuitions. Whether a given property plays an actual role in our most successful models of human behavior is determined by the models themselves, not by intuitions about hypothetical beings. Maybe such intuitions support *(2b)*, however. By showing that there must be *something* common to the Martian case and the human case, one establishes that they share some natural property, and perhaps doing so will help us to interpret properly our existing models. But, here’s where the buck stops. The intuition that the Martians would have something in common with humans should be treated in the same way I suggested that we treat Wheeler’s intuitions. If humans and Martians exhibit the same forms of behavior (but note how incredibly implausible this is, particularly if one thinks of this as benchmarked to the human behavior normally of interest to memory scientists), there may well be organizational or pragmatic reasons for continued use of the word ‘memory’ (see the “Bell Labs” example above), but this in no way implies the existence of anything with significant ontological (or even methodological) import. We have the intuition that the two species must share something, but most plausibly this is driven by behavioral benchmarking only, and that does not a science make. Lastly, I propose to ignore possibility *(3)*. So far as we know, there are no Martians of the sort at issue. It seems ridiculous that we might propose a radical reinterpretation of, and methodological change in, *our* cognitive science of humans by considering whether a nonexistent science of nonexistent creatures might include something we might be inclined to call ‘memory’.

Are all the possibilities exhausted, then, with only negative results? Some of Clark’s remarks, as well as some of John Sutton’s, suggest a different way of arguing for *(2a)*: to accept that human memories make up a motley at the fine-grained level, but to insist that there is a proper science of memory nevertheless. The remainder of this section explores this strategy (although note that Clark has more recently changed his mind about the underlying assumption of a motley: 2011, 452).

When discussing the motley of causal processes involved in the production of intelligent behavior, Clark expresses hope for a science of cognition regardless of such disunity: “The study of mind might...need to embrace a variety of different explanatory paradigms whose point of convergence lies in the production of intelligent behavior” (2008, 95; see also, 2010a, 64).[[10]](#footnote-10) And in advocating for an overarching, interdisciplinary approach to memory, Sutton suggests that we should pursue all manner of memory-related phenomena, looking for “higher-level accounts which do find commonalities” (2010, 214) in spite of disunity at the nitty-gritty levels.

I agree with Sutton that, generally speaking, we must “wait and see” (*ibid*., 215) what sorts of fruit interdisciplinary study of memory will yield. But, I take the Natural-Kinds Argument to rest on stronger claims about where we have already arrived. If Premise #2 of the Natural-Kinds Argument worth pursuing asserts an empirical claim, we must ask whether our current sciences of memory support that claim.

Nevertheless, we speculate a bit about where the empirical pursuit of generic kinds might lead, if only to clarify what would count as success. In this regard, we should ask what distinguishes a legitimately scientific generic kind from an illegitimate one or a merely nominal one. First, their instances must share more than that they are “the kinds of things that produce the phenomena of interest.” In other words, it will not do simply to formulate a general description of the coarse-grained phenomenon of interest, for example, in the case of memory, “states or processes producing behavior that matches, along dimensions of content or structure, external material with which the subject has causally interacted” or, in the case of cognition, “states or processes that produce intelligent behavior.” These descriptions can, and should, be deployed at the initial stage, as hooks to try to get hold of the natural kinds of *memory* and *cognition*. But, whether there are such kinds depends on whether the hooks do, in fact, get attached to something.

What, then, distinguishes genuine generic kinds from motley collections that do not constitute natural kinds (even though the terms referring to these motleys may serve some useful purpose in scientific discussions)? The general idea I would advance is that various instances of natural, generic kinds bear some kind of family resemblance to each other (cf. Wheeler 2011), but not just any old family resemblance: it must be a family resemblance determined by the causal-explanatory roles of the components of the generic kind’s instances; there must be a unity to various instances that is legitimated by the theoretical interests of the relevant disciplines themselves. Each instance (or kind of instance) of a generic kind is, I maintain, constituted by a cluster of mechanisms; variations in this components of this cluster from one species of the generic kind to another (and variations in the relations between them) determine the fine-grained differences in the causal profiles of various species of the generic kind. Think in terms of partially overlapping models. The models of the way in which various species produce instances of the relevant *explananda* must have significantly overlapping elements and relations among them. This would seem to be the order of the day in most sciences; an initial (typically simple) model of some paradigmatic phenomenon succeeds (well enough), then related phenomena are modeled by the “tweaking” of the initial models – terms are added, parameter values adjusted, etc. If a phenomenon that might have been thought to be of a piece with the others turns out not to be amenable to this “tweak and extend” treatment, we treat this as a different kind of phenomenon after all; and this is when we say that original, full range of phenomena weren’t all of the same kind – that is, there is no generic kind that subsumes them all.

This approach to generic kinds places a genuine constraint on the collections of instances (or kinds of instances) that can be of the same generic natural kind. If, for example, a Martian exhibits “memory-related behavior,” but that behavior is produced by a collection of *very* different mechanisms from the ones that produce memory-related behavior in humans, then the Martian behavior is not produced by memories, at least not if we want to use ‘memory’ as a natural-kind term (rather than, say, as a merely organizational term).

V. *The systems-based account*

What about cognition? Is it a natural, or scientific, kind? And if so, how broadly is it instantiated?

As a first step, we should try to understand the nature of cognition in the most salient case, the human one. In other work (2004, 2009, 2010), I have argued (a) that, across a wide range of cases, experimental work and research programs that treat organisms as containing cognitive systems has been very successful, and (b) that a striking fact about cognitive modeling provides both the best explanation of (a) and grounds a theoretically based location-neutral account of cognition. The striking fact is that virtually all forms of cognitive modeling distinguish between, on the one hand, the persisting architecture, which is taken to have a relatively fixed number of elements (e.g., connectionist units) and stable relations among them (e.g., degrees of inhibition or ways in which the degrees of inhibition change over time), and on the other, more transient causal contributors that, together with aspects of the persisting architecture, produce intelligent behavior (cf. Wilson 2002). Think of this as an inference to the best (available) explanation, twice over. First, the fact that the distinctive and central aspect of cognitive modeling – the persisting architecture – is typically instantiated within the organism explains why there’s been as much success as there has been in doing “organism-oriented” cognitive science. Second, that the persisting architecture is the distinctively cognitive thing best explains why it appears in all different forms of modeling.

I will not defend this line of reasoning here but confine myself to some relevant observations about the approach and result it yields:

First, I take it that, for most subjects, at most times, the persisting system is housed within the organism. Note that Otto and his notebook can be described so that, by hypothesis, the notebook becomes part of his cognitive architecture, but that’s irrelevant to the claim that we’re undergoing, or have undergone, or should undergo a revolution in cognitive science; Otto’s case is make-believe, after all.

Second, my strategy is to find a common thread among successful model-types, issuing from the entire range of orientations in cognitive science (including modeling that has sometimes been interpreted as extension-friendly): connection, computationalist, dynamicist, brute biological, robotics-based, and artificial-agent-based.[[11]](#footnote-11) If this “supervaluation” strategy is right, it provides powerful evidence in favor of the generally nonextended view (so long as the preceding observation is correct as well). Moreover, this evidence is of precisely the kind that Ross and Ladyman (2010) rightly demand. The evidence has nothing to do with intuitions about causality and constitution or everyday examples involving air conditioning or stereo systems. My claim rests on the real, scientific modeling, on the ground. Note, too, that contrary to what Ross and Ladyman suggest, the supervaluation strategy works: when one considers the full range of successful models, one does not find such a widespread, context-sensitive shift, where sometimes the privileged elements (the architectural elements) are inside the organism and sometimes out.[[12]](#footnote-12)

Third, my approach is locationally neutral. I attempt to find a central and pervasive distinction present in wide range of successful cognitive models, then check, afterwards, to see where the elements on either side of the divide appear. The division is not sought with malice aforethought (my initial reaction to the extended view was actually sympathetic, even though I distinguished it from the embedded view – see Rupert 2001, 505, n7); in fact, the distinction that seems to be most central to cognitive modeling leaves wide open the possibility that elements on either side of the division fall on either side of the organismic boundary.[[13]](#footnote-13)

Fourth, the present exercise does not presuppose that a mark of the cognitive is needed to do cognitive science. Successful work in cognitive science helps us to see what the mark of the cognitive might be (or what might be necessary conditions for something’s being cognitive), but we absolutely do not need a mark of the cognitive to do cognitive science. The present issue is how best to interpret the cognitive science we have (and correspondingly, how to interpret claims that cognition extends beyond the boundary of the organism). We begin with a revisable agreement about the data to be explained by our budding science (see section III), and perhaps some general idea how the data might be approached (via, say, some provisional metaphor), but that’s quite different from having in hand an account of the central theoretical construct(s) that will, in the end, be used to account for the data (or an appropriately revised and augmented set of data). How in the world could someone have *that* in hand at the outset?

Fifth, my proposal fills Mark Rowlands’s demand for an owner of cognitive states (Rowlands 2009). Rowlands realizes that his various conditions for extended cognition lead to unacceptable results (cognitive bloat, in particular) absent a further constraint on cognitive extension; to be part of a cognitive process it has to be owned by a subject. Naturally, philosophers of cognitive science will want a theory of the self and of ownership. I have offered an empirically motivated one: the self is the cognitive architecture, and it owns a state just in case that state is a state of one of the architecture’s component mechanisms.

Sixth, contrary to some of Mark Sprevak’s claims, we do not need to settle this issue by appealing to intuitions about hypothetical cases. Sprevak (2010, 361) assigns a substantive role to the Martian intuition, and to intuitions about thought experiments more generally; we must appeal to them if, for example, we are to decide between an extended interpretation of successful empirical work and alternative, less radical interpretations of that work (such as the hypothesis of embedded cognition – Rupert 2004). I disagree. We need a theoretically motivated, location-independent account of cognition to do the job, and we can get that from successful cognitive scientific practice itself. Cognition is whatever distinctive property produces intelligent behavior. I maintain that the property in question is *being the activity of mechanisms that are part of the persisting architecture*.[[14]](#footnote-14)

What, then, of the kind *cognition*? According to the systems-based view, a state (or process) is cognitive (if and?) only if it is the state of a (non-background) mechanism (or is a process made up wholly of causally connected states of various mechanisms) that is a component of a persisting architecture – that is, a member of the relatively persisting set of mechanisms that co-contribute, in various intersecting subsets, to the production of the relevant set of instances of intelligent behavior (those that are part of a single biography). Might this view ground a generic conception of cognition, of the sort suggested in section II? There are, on this view, component mechanisms (a) individual ones of which make systematic contributions to the instantiation of a variety of kinds of *cognition* and (b) manipulation of which can cause systematic variation in what we might recognize as the degree of intelligence of the associated behavior (because what we’re *really* out to explain is not just intelligent behavior, but also variations in the degree of intelligence that a given form of behavior exhibits).

Nevertheless, *this is only one kind of cognition*. We simply don’t know about other species or other kinds of cognition. Moreover, variations in kinds of cognition will likely involve second-order variation – in the kinds of systemic integration, for example – and are a bit difficult to get our heads around. If there is to be a generic kind, *cognition*, different species (or different sub-groups within species) must produce intelligent behavior via different sets of mechanism that share some highly abstract properties, but not others (properties of dynamics, architectural organization, etc.) that meet the conditions given in section II for various processes’, states’, or systems’ being different sorts of a generic kind; they must all be produced by “tweak and extend” accounts of alien intelligent-behavior-producing systems – that is, by accounts that are systematic extensions and variations on my systems-based view. This is a tall order, and the tallness of it provides some reason for skepticism about the existence of generic cognition. At the very least, we can be fairly certain that not much cognitive science is engaged in the exploration of generic cognition at present.

*VI. Afterword*

Clark and Chalmers’s central claim concerns human cognition and mental states; that’s what’s so striking about it (we’re not where we thought we were!). Whether or not there exists a natural kind, *generic cognition*, that covers all processes that produce intelligent behavior depends on empirical facts beyond our ken: we must know whether, of all the creatures that exhibit intelligent behavior (if this is a well-defined category), the various structures of mechanisms producing that behavior have the right kind of unity.

Consider: Clark describes a case of Metamorpho (2011, 456–457), a creature that (hypothetically) travels the world assembling and disassembling “himself” in certain respects during the completion of cognitive tasks. Clark asks rhetorically whether Metamorpho would cognize? I answer that it depends. Metamorpho doesn’t engage in human-benchmarked cognition. Thus, the alternative is that he might engage in generic cognition, that is, that Metamorpho’s processes and those in humans manifest a single generic kind, *cognition*. The details would have to be worked out. If there is no unity, even at the generic level, then surely Metamorpho’s way of producing intelligent behavior is irrelevant with regard to the application of the Parity Principle. The fact that a human uses something that would be of an altogether different natural kind from human cognition were it to be used by Metamorpho has no bearing on whether it is an instance of the nongeneric kind *human cognition* when humans use it. Even if hypothetical Metamorpho would, if he were to exist, produce intelligent behavior via processes that are, at a very abstract level, of a piece with human cognition, we should still wonder at the import of that fact. Until we discover what sort of properties Metamorpho’s system has, and why some of them correspond to interesting aspects of the way we model intelligent behavior in humans, there’s no reason to reinterpret existing models of the production of intelligent behavior. At the very least, the ball is in the court of those who say it should. Find those creatures, show that their behavior is produced by processes that share enough with humans’ that the two kinds of processes legitimately fall under the same generic kind, and show that the reasons for thinking they do so fall bear in some important way on the way we model or interpret our models. At this point, one might reasonably ask that philosophers of the special sciences focus more carefully on actual cases.

Think of matters in this way. Sprevak (2009) takes his argument to be a *reductio* of the sort of functionalism that gives rise to the extended view (partly via the Parity Principle). Instead, I take his argument to be an injunction against only commonsense analytic functionalism. A functionalism driven scientific enquiry itself will uncover natural kinds only where they actually are and help us to correct folk intuitions that are off base. In contrast, analytic functionalism will probably endorse empirically useless generic kinds, and we shouldn’t tie ourselves in knots trying to accommodate intuitions concerning the application of the corresponding terms and concepts. I say we walk through the open door of naturalism: admit that either human-benchmarked cognition is the only kind of cognition there is, and so cognitive science is about that actual kind (even though the term ‘cognition’ might have broader non-kind-revealing pragmatically driven application), or hold that there may be other natural kinds of cognition, and so *cognition* itself is a legitimate generic kind, but accept that we have no access to them at present and should not attempt to interpret our cognitive science as if it were a science that covers or invokes *those* kinds of processes.

**References**

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14,

43–64.

Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Oxford: Blackwell.

Adams, F., & Aizawa, K. (2010). Defending the bounds of cognition. In Menary (2010),

67–80.

Anderson, Michael L., (2010) Neural reuse: A fundamental organizational principle of

the brain. *Behavioral and Brain Sciences* 33, 245–313.

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence* 48, 57–86.

Bartlett, G. (2008). Whither internalism? How internalists should respond to the

extended mind hypothesis. *Metaphilosophy* 39, 2, 163–84.

Brooks, R. (1986). Arobust layered control system for amobile robot. *IEEE Journal of*

*Robotics and Automation*, Vol. RA-2, 14–23.

Chalmers, D. (2008). Foreword to Andy Clark, *Supersizing the Mind*. Oxford: Oxford

University Press, *ix–xvi*.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*.

New York: Oxford University Press.

Clark, A. (2010a). Memento’s revenge: The extended mind, extended. In Menary (2010),

43–66.

Clark, A. (2010b). Coupling, constitution, and the cognitive kind: A reply to Adams and

Aizawa. In Menary (2010), 81–99.

Clark, A. (2011). Finding the mind. *Philosophical Studies* 152, 447–461.

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

Coleman, S. (2011). There is no argument that the mind extends. *Journal of Philosophy*

108, 2, 100–108.

Fodor, J. (1974). Special sciences. *Synthese* 28, 77–115.

Kitcher, P. (1984). Species. *Philosophy of Science* 51, 308–33.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence* 73, 31–68.

Kirsh, D., and Maglio, P. (1994). On distinguishing epistemic from pragmatic action.

*Cognitive Science* 18, 513–49.

Levy, N. (2007). *Neuroethics: Challenges for the 21st century*. Cambridge: Cambridge

University Press.

Menary, R. (2006). Attacking the bounds of cognition. *Philosophical Psychology* 19,

329–44.

Menary, R. (Ed.) (2010). *The extended mind.* Cambridge, MA: MIT Press.

Michaelian, K. (2010). Is memory a natural kind? *Memory Studies* 4, 2, 170­–189.

Quine, W. (1969a). Natural kinds. In *Ontological Relativity and Other Essays* (New

York: Columbia), 114–38.

Quine, W. (1969b). Epistemology naturalized. In *Ontological Relativity and Other*

*Essays* (New York: Columbia).

Quine, W., and Ullian, J. (1970) *The web of belief.* New York: Random House.

Robbins, P., & Aydede, M. (Eds.) (2008). *Cambridge handbook of situated cognition*.

Cambridge: Cambridge University Press.

Ross, D., and Ladyman, J. (2010). The alleged coupling-constitution fallacy and the

mature sciences. In Menary (2010), 155–166.

Rowlands, M. (1999). *The body in mind: Understanding cognitive processes*.

Cambridge, MA: Cambridge University Press.

Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philosophical*

*Psychology* 22, 1, 1–19.

Rowlands, M. (2010). *The new science of the mind: From extended mind to embodied*

*phenomenology*. Cambridge: MIT Press.

Rupert, R. (2001). Coining terms in the language of thought: Innateness, emergence, and

the lot of Cummins’s argument against the causal theory of mental content.

*Journal of Philosophy* 98, 499–530.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of*

*Philosophy* 101, 389–428.

Rupert, R. (2009). *Cognitive systems and the extended mind*. Oxford: Oxford University

Press.

Sprevak, M. (2009). Extended cognition and functionalism. *Journal of Philosophy* 106,

503–27.

Sprevak, M. (2010). Inference to the hypothesis of extended cognition. *Studies in History*

*and Philosophy of Science* 41, 353–362.

Stich, S. (1996). *Deconstructing the mind*. Oxford: Oxford University Press.

Sutton, J. (2004). Representation, reduction, and interdisciplinarity in the sciences of

memory. In H. Clapin, P. Staines, and P. Slezak (Eds.), *Representation in mind:*

*New approaches to mental representation* (Elsevier), 187–216.

Sutton, J. (2010). Exograms and interdisciplinarity: History, the extended mind, and the

civilizing process. In Menary (2010), 189–225.

Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of*

*cognition and action*. Cambridge, MA: MIT Press.

Tulving, E. (2000). Concepts of memory. In E. Tulving and F. Craik (Eds.), *The Oxford*

*Handbook of Memory* (Oxford: Oxford University Press), 33–43.

van Gelder, T. (1995). What might cognition be, if not computation? *Journal of*

*Philosophy* 92, 7, 345–381.

Walter, S., and Kästner, L. (2012) The where and what of cognition: The untenability

of cognitive agnosticism and the limits of the Motley Crew Argument. *Cognitive*

*Systems Research* 13, 12–23.

Wheeler, M. (2010). In defense of extended functionalism. In Menary (2010), 245–270.

Wheeler, M. (2011). In search of clarity about parity. *Philosophical Studies* 152, 417–425.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*

9, 625–36.

1. This should give pause to Mark Sprevak. In his discussions of inference to the best scientific hypothesis, Sprevak (2009, 524; 2010, 358) claims that Clark and Chalmers’s extended explanation of Otto’s behavior has an advantage over an embedded one (Rupert 2004), by dint of its coarse-grained approach to action-explanation, but so far as I can tell from my day-to-day involvement in the cognitive science unit at my institution, hardly a research program in cognitive science focuses on such coarse-grained belief-desire explanations as are deployed in Otto’s case.

   Readers not convinced of the marginalized position of coarse-grained, commonsense belief-desire explanation should consider the following argument (which might called ‘the argument from nested modeling’): When (a) a given *explanandum* can be successfully accounted for by both a less-articulated model as well as a more articulated one, (b) the less articulated model is ontologically opaque, in the way extended models are (that is, they’re subject to parsing in equally reasonable ways – Rupert 2004), and (c) the more articulated model accounts for more of the variance in the relevant behavioral data, then we should derive our ontological conclusions (or the procedurally oriented equivalent) from the more articulated model. Our current situation satisfies the compound antecedent of the preceding conditional; ergo, we should, in the current situation, place more weight on our fine-grained models. Additionally, our more articulated (i.e., fine-grained) models virtually all make the distinction I describe below, the distinction between activities of the components of the persisting architecture and the other causal contributors to the production of intelligent behavior, which tends to favor a nonextended view (at least for most subjects, most of the time – see Rupert 2009, 2010). [↑](#footnote-ref-1)
2. In contrast, Clark and Chalmers seem fairly clearly to have B. in mind. Clark explicitly endorses the commonsense functionalist interpretation of the original discussion of Otto and, more generally, a commonsense functionalism that delivers coarse-grained (i.e., very broad and nondetailed) functional roles that might then be fleshed out in their particular instances by cognitive science (2008, 88, 96; 2010a; see also Chalmers 2008). Two comments: First, I agree (Rupert 2004, 422–423) with Chalmers’s observation (2008, xii) that certain aspects of folk conceptions of mental states are likely to be at odds with the extended view, and so Clark’s methodology should be clarified and defended more pointedly. Second, I think Clark should be dissatisfied with the reason he gives for preferring commonsense to empirical functionalism. He claims that empirical functionalism (or psychofunctionalism) robs us of the possibility of multiple realizations of the mental states (2008, 88 n6), but I can’t see how (cf. Wheeler 2010, 260). A system’s use of, for instance, the distance between two receptors as a way of computing the location of the source of a sound is a fairly fine-grained functional property (a) that is not part of folk theory, (b) the importance of which in cognition is empirically discovered, and (c) that can be multiply realized – in say, the distance between human ears or the distance between a robot’s sound receptors. This example represents the tip of the iceberg: almost every role in any going computational model of actual human cognition is multiply realizable and is the subject matter of an empirical enterprise not constrained in any significant way by folk psychological commitments. And although most of the properties in which cognitive scientists are interested are fine-grained, we should leave open the possibility that empirical science will find coarse-grained states, such as belief and desire, to be useful, and of course such states, being coarse-grained, are particularly susceptible to multiple realizations. Thus, Clark should remain open to the possibility that such coarse-grained mental states as belief and desire will turn out to be empirically useful kinds (the natures of which are revealed by science) and multiply realizable. [↑](#footnote-ref-2)
3. Henceforth, I omit the qualification ‘causal-explanatory’. I include it here partly to indicate how thin a notion of natural kinds can be appealed to in getting Clark and Chalmers’s natural-kinds argument off the ground. I assume that for the purpose of understanding the natural-kinds argument, being a natural kind has nothing necessarily to do with average persons’ categorization of items they encounter in their natural environment and that it needn’t imply the existence of microstructural essences (how could it require the latter? if it were to, the kinds and properties of fundamental physics wouldn’t be natural kinds!) or be associated with homeostatic property clusters. Natural kinds are simply the causal-explanatory properties and kinds of the successful sciences, or to be a bit more careful, the properties and kinds that our sciences attempt to identify. As such, they are the kinds or properties that ground successful induction (Quine 1969a), appear as relata in laws of nature (Fodor 1974), or play causal-explanatory roles (Kitcher 1984). A positivist might insist on a linguistic formulation, claiming that talk about natural kinds is merely a way of talking about terms that play certain roles in scientific discourse, for example, general terms appearing in covering-law-based scientific explanations. I will not wade further into any of this. Interpreted as philosophers of cognitive science, I take Clark and Chalmers to be suggesting that our best cognitive science will deal in such terms as ‘belief’ and ‘memory’ and apply them to states that are at least partly constituted by physical matter beyond the boundary of the human organism, independent of any particular theoretical orientation in philosophy of science. (Because I see natural kinds as something one might roughly label ‘scientific kinds’, I group what Walter and Kästner [2012] call ‘natural kinds’ with all other kinds they treat as scientifically legitimate – including certain cluster-based kinds or family-resemblance kinds.) [↑](#footnote-ref-3)
4. Many others have written about memory as it relates to the hypotheses of extended mind and extended cognition: Rowlands 1999, Adams and Aizawa 2001, Sutton 2004, 2010. [↑](#footnote-ref-4)
5. Wheeler uses the terminology of ‘benchmarking’ more broadly than I do. He counts a theoretical specification of the nature of cognition as a benchmark (2011, 425), whereas I apply ‘benchmark’ only when our measure of the cognitive rests explicitly on reference to the actual states or processes of some specified population. There might be borderline cases, however. If we observe the actual states of only one kind of creature or artificial agent, and that inspires us to build a theory of cognition, then the distance between Wheeler’s use and mine is reduced. (It complicates matters to include multiple, but not all possible, species in one’s benchmarking category, but the essential points remain the same.) [↑](#footnote-ref-5)
6. Although Clark prefers the nonbenchmarked interpretation, he appreciates the difficulty of doing without a benchmark (2010a, 54). In the end, though, he thinks we can do without one, simply by adopting a commonsense functionalism and thinking in terms of the coarse-grained causal roles of everyday mental states (*ibid.*, 55). I’m skeptical, though; philosophical intuitions driven by commonsense functionalism are liable to at least implicit benchmarking. There are infinitely more possible coarse-grained roles than there are mental states of the sort we folk associate with our own everyday mental lives. So, when folk generate descriptions of coarse-grained roles, they are winnowing the possibilities somehow, and, most plausibly, they’re doing so by working from their own case (why else care about *these* coarse-grained roles, among the many, many possibilities, except that they’re the ones we take to play a role in our own psychologies?). Thus, so far as I can tell, Clark’s recommendation that we use such roles to identify the cognitive or the mental does not, in the end, provide a recipe for a nonbenchmarked approach to cognition or mentality. [↑](#footnote-ref-6)
7. Note a further complication in the application of the benchmarked Parity Principle. How do we handle negative cases, in which the benchmark is something we have decided is noncognitive – because, say, if it were in our heads we would count it as noncognitive (Coleman 2011)? [↑](#footnote-ref-7)
8. Hereafter, I omit qualifications meant to emphasize this paper’s neutrality in respect of the issue of scientific realism and competing, more procedurally oriented interpretations of scientific enquiry. I ask instead that the reader take this neutrality as read, even when the particular choice of words suggests a thoroughgoing realism. [↑](#footnote-ref-8)
9. The present paper grows out of a presentation made at the 5th International Conference on Memory, an enormous conference held at the University of York in the summer of 2011. So far as I can tell, although ‘memory’ appeared in the name of the conference, none of the hundreds of cognitive scientists in attendance reported research on just plain memory, that is, generic memory; this is at least anecdotal evidence that, while ‘memory’ might play an organizational role, it is not treated as a natural kind of interest in cognitive science. Keep in mind, too, the objections of cognitive bloat to which the notion of *generic memory* is liable to give rise (Rupert 2004, 421). [↑](#footnote-ref-9)
10. Clark is reacting partly to the specter of eliminativism about the mind or the category of the mental (cf. Sprevak 2009, 522–523). [↑](#footnote-ref-10)
11. See the last argument in footnote 1 for a complication and an argument that the complication does nothing to weaken the case for my systems-based view. [↑](#footnote-ref-11)
12. In this context, it might be worth revisiting a worry about such entities as the sun; on the measure I’ve proposed to diagnose the scope of the persisting integrated architecture (2009, 42–43), something that consistently contributes, along with other mechanisms, to the production of a wide range of forms of intelligent behavior is almost certain to qualify as part of the persisting cognitive system. Objection: doesn’t the sun fit that category? Doesn’t the big bang? Yes, and although I’ve tried to avoid these consequences in various plausible ways, it may be best to rely on the “common element” strategy. All forms of models leave, for instance, the sun out of the architecture, and this gives us reason to toss out (by reflective equilibrium, one might say) some mechanisms that might otherwise seem to contribute causally in the way deemed adequate by my formal measure. The models all treat the sun as a background condition and that alone justifies treating it as such.

    It might also be worth pointing out that my measure of the clustering of mechanisms (that is, of the scope of the integrated architecture) is consistent with a modular architecture (*contra* the suggestion made by Clark – 2011, 456 – and others). The measure is sensitive to way in which factors co-contribute to the causal production of intelligent behavior, not to whether the factors causally interact with each other when producing that behavior. (One way to think about the mechanisms in question is as local neural mechanisms that perform simple computational functions and contribute to the production of various forms of intelligent behavior partly by being assembled in different combinations for different purposes (see Anderson 2010).) [↑](#footnote-ref-12)
13. Perhaps the standard *explananda* of cognitive science somehow bias the development of models, so that no matter what orientation one works from, the models developed are more likely than they should objectively be to contain (a) elements of an architectural sort and (b) architectural structures that are more likely than they should objectively be to appear within the boundaries of organisms. Perhaps, but I’d like to see this sort of concern worked out in some detail and accompanied by a plausible alternative suggestion about the *explananda* of cognitive science that would not have this biasing effect (or any other). [↑](#footnote-ref-13)
14. In a pair of recent papers, Adams and Aizawa (2010) and Clark (2010b) debate the status of a pencil used by a mathematician to solve problems. Adams and Aizawa don’t take very seriously the idea that the pencil itself is cognitive, and, in response, Clark argues that they’re asking the wrong question; they shouldn’t ask whether the stand-alone pencil is cognitive, as if it’s a property the pencil might have in isolation. I think Clark is right, but his being right about this significantly constrains the application of the Parity Principle. The fact that a state or object might be part of a Martian’s cognition has no bearing on whether it’s cognitive in the human case. Whether or not it’s part of human cognition depends on how the human is using it. The state doesn’t acquire the property of being cognitive, neat, simply because it’s cognitive in a different context, when a Martian interacts with it. In order that the state be cognitive in the case when it’s part of a human-centered system, it must satisfy a location-independent criterion for something’s being cognitive (or of human cognition, if cognition is not a natural kind) that is sensitive to the state’s status on the particular occasion in question. And, on my view, that involves reference to the persisting set of mechanisms that co-contribute, in various overlapping subsets, to the production of a wide range of forms of intelligent behavior.

    Although limitations of space prevent a full discussion of complementarity-based (or so called second-wave – Sutton 2010) arguments for cognitive extension, my response invokes the systems-based criterion. It is true that inner and outer contributors to the production of intelligent behavior complement each other in deep ways, but this doesn’t change matters with regard to the arguments for the systems-based view. The inner and outer play significantly different roles in cognitive modeling, and that difference is *the* central distinction between different kinds of causes to the production of intelligent behavior. Thus, if there is an interesting distinction between some causes and others, it is to be found on one side but not the other side of the line I have identified. [↑](#footnote-ref-14)