Defending the Bounds of Cognition

Fred Adams
Department of Philosophy
University of Delaware
Newark, DE 19716

Ken Aizawa
Department of Philosophy
Centenary College of Louisiana
Shreveport, LA 71134

Question: Why did the pencil think that $2 + 2 = 4$?
Clark’s Answer: Because it was coupled to the mathematician.

That about sums up what is wrong with Clark’s extended mind hypothesis. Clark apparently thinks that the nature of the processes internal to a pencil, Rolodex, computer, cell phone, piece of string, or whatever, has nothing to do with whether that thing carries out cognitive processing. Rather, what matters is how the thing interacts with a cognitive agent; the thing has to be coupled to a cognitive agent in a particular kind of way. Clark (20??) gives three conditions that constitute a rough or partial specification of the kind of coupling required.

1. The resource has to be reliably available and typically invoked.
2. Any information retrieved from/with the resource must be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory.
3. Information contained in the resource should be easily accessible as and when

---

1 Clark does shy away from this from time to time, but more on this below.
required. (cf. Clark, (20??), pp. 6-7).

Granted condition three doesn’t fit the use of a pencil very well, since the mathematician is not really extracting information from the pencil, but blame Clark for that. After all, he likes the idea that the use of pencil and paper in computing sums constitutes part of an agent’s cognitive processing, hence it’s up to him to make his story work there.\textsuperscript{2}

When Clark makes an object cognitive when it is connected to a cognitive agent, he is committing an instance of a “coupling-constitution fallacy.” This is the most common mistake that extended mind theorists make.\textsuperscript{3} The fallacious pattern is to draw attention to cases, real or imagined, in which some object or process is coupled in some fashion to some cognitive agent. From this, slide to the conclusion that the object or process constitutes part of the agent’s cognitive apparatus or cognitive processing. If you are coupled to your pocket notebook in the sense of always having it readily available, use it a lot, trust it implicitly, and so forth, then Clark infers that the pocket notebook constitutes a part of your memory store. If you are coupled to a rock in the sense of always having it readily available, use it a lot, trust it implicitly, and so forth, Clark infers that the rock constitutes a part of your memory store. Yet coupling relations are distinct from constitutive relations, and the fact that object or process X is coupled to object or process Y does not entail that X is part of Y. The neurons leading into a neuromuscular junction are coupled to the muscles they innervate, but the neurons are not a part of the muscles they


innervate. The release of neurotransmitters at the neuromuscular junction is coupled to the
process of muscular contraction, but the process of releasing neurotransmitters at the
neuromuscular junction is not part of the process of muscular contraction. That’s a quick and
dirty run through what we might call the “coupling-constitution fallacy.” For a less quick and
dirty treatment, see Adams & Aizawa (in progress).

So, if the fact that an object or process X is coupled to a cognitive agent does not entail
that X is a part of the cognitive agent’s cognitive apparatus, what does? The nature of X, of
course. One needs a theory of what makes a process a cognitive process, rather than a non-
cognitive process. One needs a theory of the “mark of the cognitive.” It won’t do simply to say
that a cognitive process is one that is coupled to a cognitive agent, since this only pushes back
the question. One still needs a theory of what makes something a cognitive agent. This is
another weakness of extended mind theories. Yet, in all fairness to Clark and other extended
mind theorists, it must be admitted that one of the shortcomings of contemporary cognitive
psychology is that there is no well-established theory of just exactly what constitutes the
cognitive. Be this as it may, Adams & Aizawa (2001) set out a rather familiar proposal, namely,
that cognition is constituted by certain sorts of causal processes that involve non-derived content.
We motivated this proposal in two ways, by appeal to examples in other sciences, such as
chemistry and physics, and by appeal to what appear to be psychological laws. We mentioned in
particular psychophysical laws, such as Weber’s law, and psychological laws governing memory
formation and recall. We might well have extended our examples by appeal to further examples
to be found in cognitive psychology textbooks. What we, therefore, proposed is that the weight
of empirical evidence supports the view that, as a matter of contingent empirical fact, there are processes that are a) recognizably cognitive, b) take place in the brain, c) do not take place outside of the brain and d) do not cross from the brain into the external world.

We think that Clark has not yet come to grips with what we are getting at with the view that cognition is a species of causal processing involving non-derived content. Our paper did not provoke him to address what seems to us to be the two most widespread problems with extracranial and transcranial theories of tool use. That is to say, Clark provides no response to the coupling-constitution fallacy and he provides little more than a hint at what he thinks distinguishes the cognitive from the non-cognitive. Further, we are disappointed that we were unable to convey our objections clearly enough to forestall Clark’s criticisms.

**The Intrinsic Content Condition**

In Adams & Aizawa, (2001), we proposed that “A first essential condition on the cognitive is that cognitive states must involve intrinsic, non-derived content” (p. 48). This hypothesis has some calculated openness in it. Suppose that during the course of a cognitive process an agent entertains the thought that John loves Mary. This cognitive agent might thus pass through a cognitive state containing the representations JOHN LOVES MARY. Then, our proposed condition would be satisfied. But, suppose that instead the cognitive agent passed through a cognitive state that has JOHN LOVES MARY followed by a period or maybe some

---

4 In a conference presentation in which he responds, in part, to Adams & Aizawa, (2001), Clark alludes to Dennett, (1990), providing an argument against non-derived content. Clark does not refer to this argument in this volume, so we have produced an independent critique of Dennett’s paper in Adams & Aizawa, (under review).

parentheses thrown in. Still, our proposed condition on the cognitive would be satisfied. The hypothesis has this latitude, since we think that while we have good reasons to believe in the existence of intrinsic content, we have no good reasons to think that cognitive states must consist entirely of intrinsic representations or that cognitive states must be, in their entirety, content bearing. This is why we said that, “it is unclear to what extent each cognitive state of each cognitive process must involve non-derived content” (Adams & Aizawa, 2001, p. 50).

Despite our attempts to present the foregoing position clearly, Clark criticizes us both for being too demanding and too lenient on the role we think non-derived content plays in cognition. Early in his section on intrinsic content he writes, “The question is, must everything that is to count as part of an individual’s mental processing be composed solely and exclusively of states of affairs of this latter (intrinsically content-bearing) kind? I see no reason to think that they must” (Clark, 20??, p. 10) Here Clark tars us with the overly strong view which we explicitly rejected, then proceeds to critique the overly strong view. (We shall return to this critique which we find unconvincing.) Later, when Clark comes to our claim that about the extent to which each cognitive state of each cognitive process must involve non-derived content, he treats the qualification as rendering the condition vacuous. Yet, this is not a very serious attempt to understand what we are after. Clearly, we mean that if you have a process that involves no intrinsic content, then the condition rules that the process is non-cognitive. In fact, that is exactly

---

6 If you think that a cognitive state is a total computational state of a computer, such as a Turing machine, then you will have another reason to doubt the view that a cognitive state must be representational in its entirety. In such views of cognition, at least some of the program states are not representational. That is, for at least some Turing machines, the read-write head of a Turing machine being in state $S_0$, or whatever, is not representational.
what the condition is used to show in our 2001 paper. The images on the CRT of the TETRIS video game are not representations of blocks to be rotated, they are the blocks to be rotated.\textsuperscript{7}

Although Clark attributes to us a view we rejected, we find that his case against this misinterpretation is unconvincing. We want to review this simply to clarify, where we can, features of the distinction between derived and underived content. So, what is Clark’s case against thinking that not all of an individual’s cognitive states must be exhaustively constituted by non-derived representations? The following:

suppose we are busy (as part of some problem-solving routine) imaging a set of Venn Diagrams/ Euler Circles in our mind’s eye? Surely the set-theoretic meaning of the overlaps between say, two intersecting Euler circles is a matter of convention? Yet this image can clearly feature as part of a genuinely cognitive process (Clark, 20??, p. ??). Evidently the problem here is supposed to be that there are some mental states that have contents in virtue of a social convention. So, Clark implies that there are \textit{bona fide} cognitive processes that involve derived content. Clark explores a line of response he thinks we might try. That line, however, strikes us as very weak. We’ll bother with none of it. Our view is that Clark’s analysis of the Euler circles case is superficial and confused.

To begin, let us draw a rough and ready distinction between mental representations of natural objects and mental representations of objects with derived content.\textsuperscript{8} The idea is that there


\textsuperscript{8} We might run what follows using a different terminology. We might talk about states in which the contents are natural objects and states in which the contents are objects with derived content. We choose to write about mental representations simply for convenience.
are mental representations of things like trees, rocks, birds, and grass, on the one hand, and mental representations of words, stop signs, warning lights, and gas gauges, on the other. Perhaps a better terminology can be chosen, but the names are really inessential. Now, by our lights, words, stop signs, warning lights, and gas gauges mean what they do through some sort of social convention. By our lights, mental representations of natural objects, such as trees, rocks, birds, and grass mean what they do in virtue of satisfying some naturalistic conditions on meaning. Many of the papers in Stich and Warfield, (1994), present some of the options that philosophers have proposed in these latter cases. Clark’s example with the Euler circles draws attention to a muddier case, the case of mental representations of items with derived content. How do these get their meanings?

As noted above, Clark suggests that mental representations of items with derived content get their content by social convention. Now, it is common ground that social convention is in some sense involved in the meaning of the overlap of Euler circles. But, that is a logically separate matter from what makes an imagistic mental representation of intersecting Euler circles mean what they do. Intersecting Euler circles on paper getting their meaning is one thing; intersecting Euler circles in mental images getting their meaning is another. Clark apparently overlooks this difference, hence does not bother to provide a reason to think that Euler circles in mental images get their meaning via social convention. For all Clark says, mental items that have Euler circles as their content could mean what they do by some naturalistic theory of content, just as we suppose that mental representations of natural objects do. So, for all Clark says, a mental image of an intersection of two Euler circles means what it does in virtue of
satisfying the conditions of Fodor’s, (1994), asymmetric causal dependency theory of content. Moreover, what we have just said about Euler circles applies just as well to mental representations of words, stop signs, white flags, and warning lights. It can be a matter of convention that “dog” means dog, that a stop sign means that you should stop, that a person raising a white flag means to surrender, and that a red light flashing means that something is overheating. But, that does nothing to show that it is not the satisfaction of some set of naturalistic conditions on non-derived content that get something in the head to have the meanings of “dog”, a stop sign, a white flag, and a warning light.

But, suppose Clark acknowledges that there is a conceptual difference between how mental objects get their contents and how artifacts outside the mind get their contents. He might give the following argument for his view. He might still think that there cannot be mental images in which intersecting Euler circles mean set-theoretic overlap unless there were a social convention according to which intersecting Euler circles meant set-theoretic overlap. He might say that this is a kind of derivation of meaning. The meaning of the mental image derives in part from the prior existence of the meaning of physical pictures. The meaning of the mental image might be said to depend on the existence of a prior meaning.

At first blush this argument may seem compelling, but in reality the argument merely trades on an ambiguity in the notions of derivation and dependency. Insofar as there must be a social convention regarding the intersections of Euler circles in order to have a mental representation regarding the intersections of Euler circles, this is not a fact about the constitution

\[9\] See Dennett, (1990), and our discussion of it in Adams & Aizawa (under review), for another instance of this kind of problem.
of the content of a mental image of the intersections of Euler circles. It is, if anything, a kind of historical fact. One would not have a mental image involving the intersection of Euler circles meaning set-theoretic overlap without having had at some prior time the social convention involving the intersection of Euler circles meaning set-theoretic overlap. It is like this. The dependence of meaning of the mental image of intersecting Euler circles on the social contrivance regarding the intersection of Euler circles is just like the dependence of the meaning of a mental representation of a car on the contrivance of a car. Had the car not been invented, there would not have been mental images of cars. Had the usage of Euler circles not been invented, there would not have been mental images of Euler circles for set-theoretic purposes. This sort of historical truth, if it is a truth, does not show what Clark might want it to show, namely, that the content of certain mental items derives (in the relevant sense) from a social convention.

Suppose, now, that Clark concedes that there is a conceptual difference between how mental objects get their meaning and how physical objects outside the mind get their meaning and admits that he has no argument for the former having derived content, but then demands some reason to think that mental objects do not have derived content. Maybe he has no argument in support of his view, but what reason is there against his view? In the arrangement of social conventions, we have some access to the items bearing the content we want. A community might get together and decide that a yellow flag, rather than a white flag, means surrender, that

---

10 There is room to challenge the historical claim that had the use of Euler circles not been invented, there would not have been the use of the mental images of Euler circles. For present purposes, however, we will not pursue this.
“bad” or “cool” makes a positive commentary on a thing, or that “WC” is a symbol for the facilities. To do these things, there has to be some way to specify or access the would-be syntactic item that is to figure in the semantic convention. Yet, with the brain, we have no such access to the syntactic items we would like to have bear a particular content. We cannot make, say, the firing of a particular set of neurons mean what it does, simply by an agreement that it does. We cannot do this since we have no way to identify particular tokens of brain states qua syntactic items in order to affix contents to them. Given the state of current science, we only identify a person’s brain states via inferences to the content of those states. We think that Jones wants to go to that restaurant in Philly because she said she wants to go to that restaurant and is looking up the address in the phone book. Even when we know that Jones wants to go to that restaurant in Philly, we don’t know what specific syntactic item in the brain bears that content. This is not how conventional meanings work.

So, as far as we can tell, Clark gives no reason to doubt what we think is false, namely, that all cognitive states must be exhaustively constituted by content bearing items. Much less does he give any reason to doubt what we think is true, namely, that cognitive states must involve non-derived content. Further, there are reasons to believe that cognitive content is not normally derived via any sort of social convention. Perhaps there are futuristic science fiction scenarios in which humans have sufficient access to brain states that this situation could change, but then maybe it will be the case that cognitive content can at times be socially controlled. Maybe. After all, can a mental image of Abraham Lincoln really mean George Washington?

The Causal Processing Condition
Our appeal to scientific categorization via causal principles is meant to do two sorts of things for us. First, it is supposed to draw attention to what appears to be one of the principal differences between processes that occur in the brain and processes that occur outside of the brain. Second, it is supposed to draw attention to the unruly collection of processes that might fall under the rubric of a would-be “brain-tool science.” Although both of these contentions undermine transcranial theories of cognition, Clark directs most of his attention to the second use of the causal processing condition. He thinks that this argument is doubly flawed. We shall address each of these alleged flaws in turn.

The First Flaw. Clark begins his critique with the following:

The first thing to say in response to all this is that it is unwise to judge, from the armchair, the chances of finding ‘interesting scientific regularities’ in any domain, be it ever so superficially diverse. Consider, for example, the recent successes of complexity theory in unearthing unifying principles that apply across massive differences of scale, physical type, and temporality. There are power laws, it now seems, that compactly explain aspects of the emergent behavior of systems ranging from XX to YY. In a similar vein, it is quite possible that despite the bottom-level physical diversity of these processes that write to, and read from, Otto’s notebook, and those that write to, and read from Otto’s biological memory, there is a level of description of these systems that treats them in a single unified framework (for example, how about a framework of information storage, transformation, and retrieval!) (Clark, 20??, p. 14, cf, p. 15).

We find this passage indicative of a number of respects in which we have failed to make our argument sufficiently clear.

Let’s begin by clarifying what we take to be the epistemic status of our view. Clark claims that, “it is unwise to judge, from the armchair, the chances of finding ‘interesting scientific regularities’ in any domain, be it ever so superficially diverse.” This may be just a
generic rejection of anything like “armchair philosophy.” We don’t endorse armchair philosophy and we don’t see that we are guilty of it. We think that the available empirical evidence provides good reason to think that the chances of finding interesting cognitive regularities covering brains and tools is low. Bear in mind, we side with what is by all accounts scientific orthodoxy. Note as well that Clark does not respond to us by marching out an interesting scientific or cognitive regularity we didn’t see from our “armchairs.”\footnote{In truth, when Clark starts “pumping intuitions” (p. 3), talking about Martians (p. 11), and drawing attention to what could happen in science (p. 14), it begins to sound as if he is the one doing armchair philosophy.} Alternatively, Clark may be giving an argument for the conclusion that it is unwise to judge the chances of finding interesting scientific regularities that might constitute a “brain-tool science”. Clark’s argument may be that, just as we have found surprising new regularities through complexity theory, so we might find interesting new regularities in “brain-tool science”, perhaps they will be information processing regularities. This argument, however, is hardly compelling. Are we to think that a judgement is unwise simply, because it could be wrong? More compelling would be to argue that a particular judgement is unwise because it flies in the face of weighty empirical evidence. More compelling would be to show us an interesting cognitive brain-tool regularity that we have overlooked. Yet, Clark provides no such case.

Think of the foregoing this way. We maintain that the weight of empirical evidence supports the view that there are processes that a) are plausibly construed to be cognitive b) occur within the brain, c) do not occur outside of the brain and d) do not cross the bounds of the brain. One can challenge the evidence and the argumentation, but it is a bit much to suggest, as does
Clark, that there is none. We are, after all, siding with scientific orthodoxy. Since it is orthodoxy, there is at least some *prima facie* reason to think it is not scientifically groundless. Further, the fact that it is scientific suggests that the position is defeasible. So, it hardly helps Clark to point out that we could be wrong.

The observation that *it is possible that* there are higher level information processing regularities that cross the boundary of the brain does nothing to challenge our position, which is concerned with what the evidence shows. Now, however, let’s see what happens if we grant Clark a much stronger premise. Suppose we detach the modal operator. Suppose that there really are information processing regularities that cross the boundary of the brain.\(^\text{12}\) Perhaps processing information is what Clark thinks constitutes the mark of the cognitive, a condition other than being connected to a cognitive agent.\(^\text{13}\) Does this much stronger, non-modal premise suffice to establish that the mind extends beyond the bounds of skin and skull? No. The problem is that the empirical evidence we have indicates that the brain processes information according to different principles than do common brain-tool combinations. Just think of consumer electronics devices. We find that DVD players, CD players, MP3 players, tape recorders, caller ID systems, personal computers, televisions, FM radios, AM radios, cell phones, watches, walkie talkies, inkjet printers, digital cameras, and so forth, are information processors. The preponderance of scientific evidence, however, indicates that they process information differently than does the brain. That is why, for example, the brain is capable of linguistic

\(^\text{12}\) This is what Rowlands, (1999), clearly thinks constitutes a basis for a version of the extended mind hypothesis.

\(^\text{13}\) This harks back to our opening paragraph.
processing where these other devices are not. That is why, for example, the brain is capable of facial recognition over a range of environmental conditions, where these other devices are not. This is why the brain is crucial for humans to drive cars, where these other devices are not. The differences in information processing capacities between the brain and a DVD or CD player is part of the story why you can’t play a DVD or CD with just a human brain. These differences are part of the reason you need a radio to listen to FM or AM broadcasts. It is these differences that support the defeasible view that there is a kind of intracranial processing, plausibly construed as cognitive, that differs from any extracranial or transcranial processing. This is the first kind of work we take our appeal to causal processing to do.

We appeal to the nature of causal processing to do more work when we observe that consumer electronics devices and other tools differ amongst themselves in how they process information. DVD players process information differently than do digital cameras. Digital cameras and DVD players process information differently than do FM radios. This, after all, is what differentiates these tools from each other. What information processing principles do string, a rock, and DVD players have in common? When we press this point, we suppose that tools constitute an open-ended set of objects. Tools do not constitute a natural kind; tools are, after all, artifacts. It is for this reason that, a would-be brain-tool science would have to cover more than just a multiplicity of causal processes. It would have to cover a genuine motley. A brain-tool science would not have to cover a mere disjunction of things; it would have to cover an open disjunction. In our 2001 paper, we noted the existence of areas of scientific
investigation where there was an apparent fragmentation of a domain.\textsuperscript{14} The reason, we argued, that brain-tool science will not go the way of these other investigations is that a would-be brain-tool science would have to cover too broad a collection of processes. It would have to cover a motley of processes, not just a multiplicity of processes.

Clark has hinted that information processing constitutes the mark of the cognitive, but we have argued that this is implausible. What, then, of the possibility that Clark thinks that some other higher-level processes constitute the mark of the cognitive? Perhaps the higher-level processes that extend are of some other nature. Ok. But, what are these principles and what is the evidence for their existence? Clark gives us no clue. Note as well that it is not enough for Clark to show that “there is a level of description of these systems that treats [intracranial and extracranial processes] in a single unified framework.” Physics provides a reasonable approximation to such a thing. Biology and chemistry might also provide levels of description at which there are processes that are continuous across the boundary of the brain. What Clark needs is a \textit{cognitive} level of description of these systems that treat them in a single unified way. That is, he needs a plausible theory of what constitutes the cognitive. That is where our theory of non-derived content and causal processes supports intracranialism.

\textit{The Second Flaw}. What, now, of the second way in which Clark thinks our appeal to causal processing is doubly-flawed? Clark observes that cognition might fragment into a motley of causally distinct processes without even a family resemblance. Perhaps the folk notion of visual processing will break down into two subtypes, visual processing that eventuates in

perceptual experiences and visual processing that guides action independently of perceptual experiences. Extrapolating from what Clark writes, we might add that memory might break down into distinct kinds, short-term memory, long-term memory, visual memory, etc. A folk notion of auditory processing could fragment into auditory processing and linguistic processing. Olfaction could have a generic smell component alongside a system for processing pheromones. If cognition is a motley, then Adams and Aizawa’s standard will judge intracranial cognitive science just as much a bust as a would-be brain-tool science.

To address this objection, we can apply much of what we said above. To begin with, we do not suppose that the decomposition of the cognitive into a motley is in any sense impossible. We made this epistemic point above. We think that the weight of argumentation supports our view. So, insofar as Clark cares to address our position, he evidently needs at least the non-modal conclusion that cognition fragments into a motley collection of principles. This, however, we are not prepared to concede. In our earlier discussion we drew a distinction between a multiplicity of principles being at work in some domain and a genuinely motley, open-ended collection of principles being at work. We think that the available scientific evidence makes it plausible that there are distinct sorts of cognitive processing occurring in the brain, processing corresponding to many distinct forms of visual processing, memory processing, and so forth. Yet, we see no reason to extrapolate to the conclusion that there is an open-ended collection. The brain is at least in the running to be a natural kind, where brain-tool combinations are hybrids of natural kinds and artifacts. Outside the realm of science fiction, the brain is constrained to develop only a limited set of distinct structures with a bounded range of plasticity. An
organism’s genome and environmental interactions limit what can be done with neurons and glial cells. Clark appeals to the wide diversity of organisms that might be capable of cognitive processing, but this does not show that there is an open-ended range of things that can constitute cognitive processing. By contrast, tools can be made of anything and work according to any number of distinct principles. They are clearly artifacts and not natural kinds. That is good grounds for saying that intracranial processing is a collection of disparate mechanisms, while brain-tool combinations are an open-ended collection.

Finally, suppose that Clark is right about cognition breaking down into a genuinely open-ended collection of principles. Even that would not necessarily vindicate extracranialist or transcranialist theories of cognition. As long as the multiplicity or motley collection of plausibly cognitive intracranial causal processes is distinct from the set of extracranial and transcranial processes, there will be a basis upon which to say that cognition is intracranial. Even if we were to concede the idea that there could be a science of the motley, a science of the motley would not vindicate extracranialism. So, as far as we can tell, Clark has said nothing that challenges our original analysis of the role of causal processing and non-derived content in the demarcation of the cognitive.

**Conclusion**

In “The Bounds of Cognition” we thought that the principal weakness in extracranialist theories of tool use was inadequate attention to the mark of the cognitive. Since then, however, we have been impressed with the extent to which this inattention appears to have been involved
in so many process externalists succumbing to one or another version of the coupling-constitution fallacy. It would certainly do much to advance the transcranial theories of cognition were Clark not only to address our theory of the mark of the cognitive, but to address the pervasive coupling-constitution fallacy and set out a plausible theory of what distinguishes the cognitive from the non-cognitive.

References


