

Extended Circularity: A New Puzzle for Extended Cognition^{*}

J. Adam Carter & Jesper Kallestrup

University of Edinburgh

Abstract

Mainstream epistemology has typically taken for granted a traditional picture of the metaphysics of mind, according to which *cognitive* processes (e.g. memory storage and retrieval) play out entirely within the bounds of the skull and skin. But this simple ‘intracranial’ picture is falling increasingly out of step with contemporary thinking in the philosophy of mind and cognitive science. Likewise, though, proponents of *active* externalist approaches to the mind—e.g. the *hypothesis of extended cognition* (HEC)—have proceeded by and large without asking what *epistemological* ramifications should arise once cognition is understood as criss-crossing the bounds of brain and world. This paper aims to motivate a puzzle that arises only once these two strands of thinking are brought in contact with one another. In particular, we want to first highlight a kind of condition of *epistemological* adequacy that should be accepted by proponents of extended cognition; once this condition is motivated, the remainder of the paper demonstrates how attempts to satisfy this condition seem to inevitably devolve into a novel kind of epistemic circularity. At the end of the day, proponents of extended cognition have a novel epistemological puzzle on their hands.

What kind of things can feature in *cognitive processes*? The traditional answer is a very simple one: cognitive processes take place in the brain; therefore, a process that is cognitive is realised *intracranially*—*viz.*, within the bounds of skull and skin. In short, cognitive processes supervene on *intracranial biological processes* and nothing else¹. That’s the standard story at any rate, and the inherited approach to mainstream *epistemological* theory has always taken

^{*}To appear in *Extended Epistemology* (eds.) J.A. Carter, A. Clark, J. Kallestrup, S. O. Palermos & D. Pritchard, Oxford University Press.

¹For a sustained recent defence of this traditional picture of the bounds of cognition, see Adams & Aizawa, (2008).

something like this traditional picture of cognition for granted² in the course of making epistemic evaluations.

But mainstream epistemology shouldn't get too comfortable with this background picture. The past several decades of work in the philosophy of mind and cognitive science have been increasingly receptive to a more radical picture of cognition—one that diverges sharply from the traditional view—in the form of *active externalism*³. Unlike 'passive externalist' approaches such as *content externalism* (e.g. Putnam 1975; Burge 1986), according to which mental *content* is determined by one's physical or social environment, *active externalist* views (e.g. the distributed cognition thesis⁴, the extended mind thesis⁵ and the hypothesis of extended cognition (HEC)) insist that the vehicles that run the content are constituted in part by elements of one's physical environment, for instance, notebooks, smart phones and (in the case of distributed cognition) even other individuals.

The particular version of active externalism that has gained the most traction recently—and especially over the past five years or so⁶—is the *hypothesis of extended cognition* (HEC), according to which what is claimed to 'extend' to include part of the world is cognitive *processes*⁷. As Clark (2007, §2) puts it,

²One simple way to make this point is in terms of competing ways of thinking about epistemic justification, on both sides of the internalism/externalism divide. On this score, epistemic *internalists* such as Conee & Feldman (2004) tell us that epistemic justification supervenes on mental states that are themselves regarded as underwritten by processes internal to the cognitive architecture of the biological agent. Epistemic externalists, despite offering a contrasting picture of epistemic justifiedness, tell a similar story as the epistemic internalist does, *vis-à-vis* the supervenience base of cognition; on the reliabilist program, for instance, epistemic justification is understood as entirely a matter of the reliability of belief-forming processes that occur squarely inside the agent's bodily boundaries. For illustrative passages see Goldman (1979, 346-7; 1986, 51). Thus both sides of this perennial first-order epistemological dispute persist with the traditional intracranial picture of cognition in the background.

³This terminology owes to Clark & Chalmers (1998), who use the term 'active' to distinguish this variety of externalism from the comparatively less radical *content externalism*. See Clark & Chalmers (1998); Clark (2007; 2008); Hutchins (1995); Menary (2006; 2007); Wheeler (2005) and Wilson (2000; 2004) for some notable defenses of active externalist positions.

⁴See, for example, Hutchins (1995); Huebner (2014) and Knorr-Cetnia (1999) for some representative examples...

⁵The seminal case for this view, according to which mental states can supervene on parts of one's physical environment, is Clark & Chalmers (1998); though, note that Clark & Chalmers also in their classic paper defending the extended mind motivated the comparatively weaker extended cognition thesis.

⁶See Carter et al (2014).

⁷Granted, many of the same reasons that would incline one to endorse extended cognition would lead one, *mutatis mutandis*, to endorse the claim distinctive of the extended *mind* thesis—*viz.*, that mental *states* (e.g. beliefs) can supervene on elements of the world external to the agent. But the move from HEC to the extended mind thesis is, to stress, not one of entailment, and is in fact often resisted. This point is easily overlooked given that Clark (2007; 2010) himself opts to move interchangeably between discussions of extended mind and extended cognition in his published work. But this is only because Clark's own preferred view of the nature of *beliefs* idiosyncratically identifies beliefs with processes as opposed to states. Against the background of a *non-static* view about beliefs, it is unsurprising that Clark is not interested in the differences between the views. Obviously, though, the received thinking about belief is as state—*viz.*, as a propositional attitude—and this leaves open an endorsement of

from the perspective of HEC, ‘the actual local operations that realize certain forms of human cognizing include inextricable tangles of feedback, feed-forward and feed-around loops: loops that promiscuously criss-cross the boundaries of brain, body and world’ (Clark 2007, §2).

To make this idea more concrete, consider Clark & Chalmers’ (1998) case of Otto:

Otto: Otto suffers from Alzheimer’s disease, and like many Alzheimer’s patients, he relies on information in the environment to help structure his life. Otto carries a notebook around with him everywhere he goes. When he learns new information, he writes it down. When he needs some old information, he looks it up. For Otto, his notebook plays the role usually played by a biological memory (Op. cit, 8).

Is Otto’s notebook a *part* of his cognitive process? The proponent of HEC typically answers—by reference to common-sense functionalism—as follows: if the *biological* memory of an ordinary agent (in the default case) is part of her memorial process, then so does Otto’s notebook count as part of his. Of course, one might object: ‘but the notebook is neither biologically constituted, nor is it located intra-cranially!’ But, as the line of thinking goes, this objection is not principled by the lights of common-sense functionalism, but just a revelation of ‘bioprejudice’⁸.

Clark & Chalmers (1998) accordingly propose, in order to cut-off the stock rejoinder, what they call the *parity principle*—a principle that provides a very useful way of thinking, free from any constitutional or locational prejudice, of what should be included in a description of cognition:

Parity Principle: If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is part of the cognitive process. (Clark & Chalmers 1998, 8)

With reference to the parity principle, Otto’s notebook gets ‘ruled-in’ as part of his *extended* memorial process—a process through which his notebook is playing a role that is functionally isomorphic to the role played by biological memory (*vis-à-vis* storage and retrieval). And so, by HEC, the notebook is claimed to be a part of Otto’s *extended* memory.

HEC without embracing the extended mind. Thanks to Andy Clark for clarificatory discussion on this point.

⁸See Clark (2008 Ch. 1, for this terminology).

1

We'll situate HEC within an epistemological setting shortly. But first, it's important to take note of the most famous contemporary line of resistance to HEC—the argument from ‘cognitive bloat’⁹. This is because, interestingly, the familiar patterns of the cognitive bloat objection to HEC share, as we'll soon see, some striking commonalities with a strand of thinking found in the classic reliabilist literature in epistemology during the later part of the 20th century. And it will be with attention to some of these commonalities that the puzzle we'll pose to the proponent of HEC gets off the ground.

First things first, though—critics of HEC are keen to compare the case of Otto (originally used to motivate HEC) with cases like the following¹⁰:

Telo: Telo has a normally functioning biological brain. Like Otto, but *also* like other individuals with properly functioning biological memory, he relies on information in the environment to help structure his life, and this includes the information in his Verizon phone book, which sits in his desk drawer by his telephone. When Telo needs to call someone, he looks up the number. For Telo, the phone book prevents him from having to memorise everyone's phone number, just in order to make a call.

The worry looks straightforward enough: HEC is going to be implausibly *inclusive* (with respect to what counts as part of a cognitive process) if it turns out that Telo's phonebook—which he consults only occasionally—gets ruled-in as part of Telo's memorial process: surely Telo's memory can not include the entire phonebook. To appreciate the force of the cognitive bloat, bear in mind (as further explained in Section III) that Telo's memorial processes serve to underpin much of his *knowledge*. Surely, to ascribe knowledge to Telo of the entire phone book strikes one as implausible. And, *mutatis mutandis*, a range of similar examples can be posed to generate comparable consequences¹¹. ‘Telo-style’ cases represent a familiar form of argument from slippery slope, and HEC thus needs a clear explanation for why we should treat Otto's case differently from such ubiquitous cases.

Proponents of extended cognition are of course keenly aware of this strand of objection, and wish to endorse HEC in a way that respects plausible boundaries between what is internal and what is external to cognition in some constitutive

⁹This general worry (expressed in terms of bloat by Rupert (2004)), has been expressed alternatively as the *coupling constitution fallacy* (e.g. Adams & Aizawa, 2008, 91). The common line is that: HEC threatens to include too much of the world into cognition.

¹⁰See for instance Rupert (2004).

¹¹See, for instance, Farkas (*Forthcoming*) for a case involving audiotapes of the complete history of Europe.

sense¹². More precisely, the pressing challenge is to draw a principled distinction between *genuine* cognitive processes and background processes or auxiliary resources that causally influence cognitive processing without constituting parts of the cognitive system.

Clark has attempted to meet this worry about demarcation by offering ‘integration conditions’ that have become known as his ‘trust and glue’ conditions for a non-biological entity to count as included within a cognitive system. According to Clark (2010, 46) the central integrations conditions that must be met are threefold:

Clark’s “Trust and Glue” Integration Conditions

- (1) “That the resource be reliably available and typically invoked.”
- (2) “That any information thus retrieved be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny. [...] It should be deemed about as trustworthy as something retrieved clearly from biological memory.”
- (3) “That information contained in the resource should be easily accessible as and when required.”

While these three integration conditions do well enough to get easy cases right (e.g. cases where one consults, say, a map, but just on one occasion¹³), they are as Rupert (2004) has noted, perhaps still too inclusive: notice here that, *even* when we have moved beyond the parity of reasoning at play in the parity principle and spelled out integration conditions (1) – (3), it *still* seems like a case could be made for thinking that Telo’s phonebook might be ruled-in.¹⁴

Effectively, then, what cases such as Telo highlight is the need to articulate some integration condition *beyond* (1) – (3), which would effectively (and in some principled way) rule Otto’s notebook in and rule Telo’s phonebook out. Unsurprisingly, one live research program in the literature on extended cognition is just how to spell out an additional condition to (1) - (3) that is neither too inclusive nor too exclusive¹⁵.

¹²For a helpful discussion of this point, see also Spaulding (2012). See also Palermos (2011; 2014) for a response strategy that appeals to dynamical systems theory. Whereas Palermos (2011) viewed a kind of ‘continual reciprocal causation’ condition as a suitable fourth condition, his more recent view (2014) is that the satisfaction of the CRC condition *de facto* satisfies Clark’s conditions (1-3).

¹³Such a case would plausibly fail at least (1) and (3).

¹⁴The sticking point is whether (3) is met in the case of Telo: given that Telo doesn’t carry his external device around with him at all times, perhaps that means it isn’t easily enough available as and when required. Be that as it may. The Telo case can surely be tweaked to satisfy condition (3). Thus Rob Rupert (2004) has shown that cases like Telo can be made to fit Clark’s (1) - (3).

¹⁵See here Pritchard (2010), Palermos (2011, 754-756) and Carter & Kallestrup (2014).

2

Let's now transpose the foregoing into an epistemological setting. Memory is characteristically regarded as a *basic* epistemic source—one that is widely thought to, in the absence of defeaters, support memory-knowledge¹⁶. To the extent that we are to take seriously what the proponent of HEC tells us—*viz.* that the extended process Otto enjoys is in fact one of *extended memory*—on a par in the relevant respects with biological memory—proponents of HEC must insist that Otto's extended process supports *memory* knowledge in a way that is relevantly analogous (and certainly, not obviously disanalogous) to the way that *biological* memory supports memory knowledge in the default case. To the extent that HEC fails to preserve such an 'epistemic parity' insight, we have reason to reject that what Otto exhibits is, in fact, extended *memory*.

A helpful way to think about this 'epistemological parity' condition of epistemological adequacy for a proponent of HEC is in terms of the more basic *parity principle* we've already considered in §1—that is, a proponent of HEC should be able to preserve that, *ceteris paribus*¹⁷, for a subject *S*, and proposition *p*, if *S* comes to believe that *p* by a process which, were it to go on in the head, we would have no hesitation in ascribing knowledge that *p* to *S*, then *S* knows that *p*. In short, the Otto case had better be a case featuring memorial knowledge. And even more weakly: there had better not be any obvious impediments that constitute a barrier to Otto's attainment of memorial knowledge. (And if there is, then HEC has a problem on its hands).

At this point, some parallels between the HEC and epistemology will be revealing. In particular, it is interesting to note that a proponent of HEC who wants to vindicate knowledge through (what is, by HEC) an *extended* process—while at the same time steering clear of the cognitive bloat objection—runs up against a pattern of problems that share some obvious commonalities with a famous strand of thinking found in the classic literature on *reliabilism* in epistemology in the 1980s and 1990s¹⁸.

To appreciate this connection, take as a starting point Lehrer's (1990) notorious 'TrueTemp' case, which Lehrer aimed at exploiting what are widely taken

¹⁶There is some dispute about *how* this works. According to *generativism* (e.g. Robert Audi (1995) and John Pollock (1974; 1986) memorial knowledge is like perceptual knowledge in the sense that, just as perception generates positive epistemic status for one's perceptual beliefs, so the phenomenology of recalling generates positive epistemic status for one's memorial beliefs. By contrast, *preservantists* (e.g. Burge) draw a closer analogy between memory and testimony than between memory and perception. The preservantist line is that memorial serves the function of *preserving* the epistemic status of past beliefs, and so on this model, one's knowledge is preserved through memory and remains known when recalled. See Carter & Pritchard (2015) for further discussion.

¹⁷This is a simple statement of the epistemic parity principle. See (REFERENCES SUPPRESSED) for a refinement, whereby which the principle is framed in terms of defeasible warrant as opposed to knowledge.

¹⁸One of the first cases of this sort was Bonjour's (1980) famous case of 'Norman the Clairvoyant'.

to be implausible implications of a flat-footed process-reliabilist line on which knowledge is true belief formed *via* a reliable belief-forming process.

Here's a condensed version of Lehrer's counterexample to simple reliabilism:

TrueTemp: TrueTemp has (though entirely unbeknownst to him) a temperature-detecting device implanted in his head that regularly produces accurate beliefs about the ambient temperature¹⁹.

As many commentators have accepted, the intuition is strong here that TrueTemp doesn't attain *knowledge* in the above scenario, even though (thanks to the implanted thermometer) he reliably generates true temperature beliefs, which by the reliabilist's lights is supposed to be all besides truth that matters. What accounts for this intuition to deny knowledge to TrueTemp, who we may grant reports with 100% accuracy?

A natural reaction will be to point out that the thermometer seems to function like an 'external' device outwith Temp's own cognitive architecture, which generates reliable information for which TrueTemp himself seems nothing more than a receiving parrot. Plausibly, and more carefully: for TrueTemp to attain knowledge, it seems the correctness of his beliefs must be down to *his own efforts*, as opposed to the production of some technological implant. So, to be precise, the problem here is not that the generation of temperature-beliefs by TrueTemp's thermometer cannot count as an 'extended' cognitive process, but rather that these beliefs fall short of knowledge on account of not being creditable to *his own faculties*. Hence, the problem is how this device can play a role in an 'extended' process that yields knowledge for TrueTemp.

The foregoing line of thought has motivated some philosophers to embrace (along with a reliability condition on knowledge) something like an 'ability' condition on knowledge²⁰:

Ability condition: *S* knows that *p* only if *S*'s believing correctly that *p* is significantly creditable to *S*'s exercise of cognitive ability²¹.

¹⁹This paraphrasing of the case is taken from Goldman (2011). See Pritchard (2010) for a discussion of integration conditions.

²⁰Of course, the point here is compatible with accepting that there can be other necessary conditions on knowledge that are not satisfied in cases like the TrueTemp case. The key point is that it's plausible to think that a sufficient condition for denying knowledge in the TrueTemp case is that he fails an ability condition.

²¹It is important to bear in mind that an ability condition on knowledge is independently motivated by considerations about certain Gettier cases. For instance, in Chisholm's case of the sheep in the field, the truth of your (justified) belief that there's a sheep in the field is not primarily down to the exercise of your perceptual abilities. Rather, it is true because of some environmental happenstance, *viz.* that a sheep is hidden behind the disguised dog at which you are looking. True, a modal condition on knowledge to do with safety or sensitivity might equally well explain your lack of knowledge in this particular instance. But now imagine instead a case where you form beliefs about the ambient temperature in a room on the basis

With reference to the ability condition, there appears to be a straightforward explanation for what is lacking in the TrueTemp case—*viz.*, although TrueTemp forms his belief *via* a reliable cognitive process, the process (in his case, where his belief is entirely down to the implanted thermometer of which he is unaware) is not part of his exercise of any cognitive *ability* he has. And, furthermore, the explanation for *why* it is not is, as Greco (2010) has noted, plausibly because the thermometer is not *appropriately integrated* into TrueTemp’s cognitive agency.

3

At this point, it should be clear how the notion of ‘cognitive integration’ of interest to proponents of HEC who wish to avoid the cognitive bloat objection is one with some precedent in *epistemological* theory²²—and just as in the case of extended cognition, the epistemological issue of accounting for just what it is in virtue of which some external device should be regarded as incorporated into an agent’s own *cognitive ability* requires the specification of some *integration conditions*.

Question: *why*, exactly, is the temperature-detecting device not appropriately *integrated* into TrueTemp’s cognitive agency? (Notice that the TrueTemp case seems to be one such that Clark’s integration conditions (1) – (3) are all satisfied.) In particular, we can certainly imagine that TrueTemp automatically endorses any *information* that his implanted device yields, such as that *the temperature is now 28°C*.

One obvious explanation for why the thermometer should not count as appropriately integrated in TrueTemp’s cognitive agency (in such a way that the correctness of his success could be attributable to any ability *of TrueTemp’s*) is plausibly that the reliability of the implanted device *itself* has never been positively acknowledged by TrueTemp²³. After all, TrueTemp has no inkling as to what even produces his temperature beliefs (!), let alone his track record of success²⁴.

of reading a faulty thermometer which randomly fluctuates between 15 °C and 25 °C. As it turns out, someone next door unwittingly changes the room temperature so as to perfectly match your actual and counterfactual beliefs. Only your failure to satisfy an ability condition can explain why your true belief falls short of knowledge.

²²See here also Beebe (2004) for a discussion of cognitive integration conditions for new perceptual faculties.

²³Perhaps of course there are other explanations for TrueTemp’s lacking knowledge, and we are happy to grant this. The point is just that his failure to ever take a reflective stance on the reliability of this process suffices to prevent his cognitive success from being creditable to an ability of his in a way that would be required for knowledge.

²⁴Erik Olsson (2007), a foremost contemporary reliabilist, thinks that such a tracking requirement is, as he puts it, ‘part of the cognitive environment in which reliabilist knowledge promotes stability of belief and thereby attains its full practical value. *Hence, even if knowledge is best defined in an externalist manner, the full realization of its value requires the satisfaction of a modest internalist condition*’ (Olsson 2007, 352). We raise this point not to defend any particular form of reliabilism, per se. Rather, we want to be clear that a tracking

By analogous reasoning, as Pritchard (2010, 144) observes, in order for—in the Clark/Chalmers extended case—Otto’s process of consulting his notebook to be sufficient for generating *knowledge*, Otto’s notebook is going to have to be—and much as TrueTemp’s thermometer was *not*—*recognised* and not merely (as with TrueTemp) *relied* upon. Here’s Pritchard:

Imagine for example, that Otto is simply fitted with a device which provides him with reliable information about his environment and he unquestioningly consults it when necessary while never questioning the source of this information or its epistemic pedigree...if Otto had no awareness at all of the source of the reliability of the belief forming process, nor that it was reliable, then it is hard to see why we would now regard the true beliefs that he forms as a consequence as knowledge (Pritchard, 2010, 145)²⁵.

The foregoing suggests that the second of Clark’s three ‘glue and trust’ conditions is simply too weak in merely requiring that ‘any information thus retrieved be more-or-less automatically endorsed’. A proviso must be included to the effect that the part of the world, for it to be included as part of a cognitive process, must not only be endorsed in the *weak* sense where the *deliverances* of the process are automatically endorsed (e.g. a condition clearly satisfied by TrueTemp, *vis-à-vis* the deliverances of the thermometer), but also in a stronger sense, where the resource *itself* is (at some point) affirmed and endorsed as reliable.

Take a resource to be reliable just in case the process of utilizing that resource in forming or sustaining belief is reliable. Here’s then a more demanding fourth integration condition which delivers the desired result of ruling the thermometer out from TrueTemp’s cognitive architecture:

(4) That the reliability of the resource be endorsed.

Obviously, then, (and crucially): if Otto has *failed* to endorse his notebook as reliable, then (in all relevant respects) he is no better a candidate for memorial knowledge than the hopeless TrueTemp *vis-à-vis* knowledge of the temperature. (And that’s of course bad news in so far as HEC is to preserve *epistemic parity*—that Otto is on an epistemic par with his knowing counterpart in the default biological case.)

Now, a *prima facie* worry that might be advanced at this point is that imposing (4) on any putative cognitive process is unduly strong. More precisely, the concern would be that (4), when added to the mix, would cleave a striking disanalogy—one that should be unpalatable for proponents of extended cognition—between biological memory cases and extended memory cases.

requirement hardly runs against the spirit of reliabilism.

²⁵Compare here with Beebe (2004), §7.

After all, in the case of innate biological memory we typically do not positively affirm its reliability before we rely on its deliverances for knowledge-preservation.²⁶ To put in place such a requirement, as this line of thinking insists, would most certainly divest young children and other sufficiently unsophisticated individuals of vast amounts of stored knowledge. Hence, we should impose no such requirement in the case of extended cognition—in so far as we are to maintain functional similarity; or so the thought might be.

One reply to this objection on behalf of the friend of extended cognition is to concede that one or more of (1) – (4) can be dropped in the case of non-extended cognition. Consider, after all, that these “rough-and-ready” criteria are explicitly designed by Clark (2010, 46; our italics) to be ‘met by *non*-biological candidates for inclusion into an individual’s cognitive system’. The intention is thus not for (1) – (4) to constitute necessary and sufficient conditions for *any* resource, internal or external, to constitute a part of a cognitive system. We could thus conceive of them as forming a cluster where a resource is cognitive just in case it meets enough of these criteria. What would count as enough can then vary from case to case depending on the empirical details, as brought to light by the relevant parts of cognitive science²⁷.

An alternative approach is to simply drop (4) in those cases where the external resource replaces an innate feature of an individual which is characteristic of members of the population to which that individual belongs²⁸. The thought would be that when it comes to cognitive integration of some external resource, the more the resource functionally resembles an individual’s *natural* endowment in just that sense, the less need there is for endorsement of its presence or reliability. If for some natural features, (4) plays no (significant) role in cognitive integration then the same should hold for external resources that merely serve as hardware substitutes for those features.

We shall not here probe deeper into the problem that (4) raises for ordinary biological memory. For instance, one might think the replies we sketched out represent too much of a departure from the common-sense functionalist spirit of

²⁶For instance, if Otto’s process is the functional analogue of a normally functioning biological memory, Otto must consciously endorse the reliability of his notebook, where such a requirement does not seem needed for an individual—call her ‘Inga’—whose biological memory is normally functioning. Inga isn’t prevented from knowing that the museum of Modern Art is on 53rd street when relying on her biological memory, *just because* she has never positively endorsed the reliability of her own biological memory.

²⁷One way to flesh out this idea more precisely is to adopt Lewis’ (1972) notion of a cluster of criteria (or “properties”) as a disjunction of conjunctions of most of the criteria. But one could also imagine cases where less than a majority of the criteria would suffice, say if perfect resemblance in one respect could outweigh dissimilarities in other respects.

²⁸It is important not to drop (4) unless the external resource replaces an organic feature that is both innate and typical of the population in question. Imagine a TrueTemp case where a genetic engineer had implanted a reliable temperature-detecting device from birth. In such a case we would arguably invoke (4) once the individual was sufficiently mature before the thermometer could be considered part of TrueTemp’s cognitive machinery. The case fails to satisfy a phylogenetic constraint to the effect that the relevant implant roughly corresponds in function to what results from evolution in humans, e.g. organic memory as we all know it.

the case for the extended cognition hypothesis. For the record, we do not share that concern. In the remaining part of this paper we shall pursue a different tack. In our view, the chief problem raised by (4) is its affliction of a malignant type of epistemic circularity that arises before any issues about classifying external resources as cognitive; or so we shall argue in the next section.

4

Suppose, as we recommend, that the proponent of extended cognition avail herself of the additional integration condition (4). To require that Otto also endorse the reliability of his notebook before it can be counted as part of a cognitive process that leads to memory-supported knowledge is not implausible. Such explicit affirmation of reliability may go beyond Clark & Chalmers' original set-up of the Otto case, but nothing in the way the Otto case is described prevents us from refining the example so as to include this additional feature. Consequently, supplementing integration conditions (1) - (3) with (4) presents no insuperable difficulty *vis-à-vis* accounting for the kind of extended cognition that the Otto case exemplifies. Or so it would seem at first blush.

We shall now argue that once the friend of extended cognition grants that Otto's notebook counts as integrated within his cognitive process only if he endorses the reliability of the notebook, a kind of 'catch-22' materialises. To bring the catch-22 into focus, consider first a variation on the original TrueTemp case:

*TrueTemp**: TrueTemp* is just like TrueTemp except that TrueTemp* endorses the reliability of his thermometer. *However*, TrueTemp* endorses the reliability of his thermometer on the basis of reading a fortune cookie.

Given that the received verdict is that TrueTemp fails to know the deliverances of his thermometer, it is surely implausible to suppose that TrueTemp* *does* know what TrueTemp did not know, simply because TrueTemp* has affirmed, *on the basis of a fortune cookie*, that his thermometer is reliable. It should be very strange indeed to see how the move from the TrueTemp case to the TrueTemp* case is one where the epistemic status of the *deliverances* of the thermometer is enhanced at all²⁹.

In general, moving from not believing a proposition *p* to believing *p* on some improper basis (i.e. for some bad reason) cannot *strengthen* one's epistemic position. TrueTemp's endorsement must thus itself pass a certain kind of epistemic

²⁹Indeed, if anything, there is some pressure to move in the *other* direction and to suggest that an endorsement on the basis of a very unreliable process renders the epistemic status of the deliverances of the process worse than were the reliability of the process not endorsed at all but just blindly trusted. At any rate, the crux of the point here is that when one endorses the reliability of a process, P, the epistemic status of the deliverances of P is not *thereby enhanced* in light of an endorsement of P by an unreliable or defective process.

scrutiny; and indeed, should he endorse his thermometer in an epistemically proper way, then surely we will be less inclined to claim that he fails to know.

What this reveals then is that the strong endorsement condition (4) that Otto must satisfy, as a necessary condition on his notebook constituting a part of his cognitive process, is *not* a condition satisfied by *epistemically defective endorsement*. As the TrueTemp* case shows, no such endorsement of the reliability of a process can serve to raise the epistemic status of its deliverances.

Now we can frame the catch-22: for Otto's notebook to count as part of his cognitive process, Otto must not *merely* endorse the reliability of his notebook, but more specifically, he must do so in the right sort of way—*viz.*, in a way where the endorsement in question is itself the product of reliable cognitive abilities. This suggests that (4) should be replaced with the more demanding (4*):

(4*) That the reliability of the resource be *reliably* endorsed

Clearly, Otto's endorsement of the reliability of his notebook *via* his *failing biological memory* is no better than if TrueTemp simply used an unreliable process to endorse the reliability of his thermometer. (After all, Otto's failing biological memory is, given the effects of Alzheimer's disease, highly unreliable). But since the move from the TrueTemp case to the TrueTemp* case is not one where the epistemic status of the deliverances of the thermometer is enhanced, *neither* is it plausible to think that Otto should enhance in any way the epistemic status of the deliverances of his notebook simply by endorsing its reliability with an epistemically defective, failing biological memory. So, Otto fails to satisfy (4*) as long as his endorsement of the reliability of his notebook relies on his (failing) biological memory³⁰. Yet Otto must satisfy (4*) if he is to count as having knowledge based on an extended memorial process.

But there is an obvious alternative. While Otto's failing biological memory is not a reliable process, his consulting of the notebook is! Thus, as the argument might go, Otto's self-conscious endorsement of the reliability of the notebook must be not merely retained through biological memory³¹ (as this wouldn't increase the epistemic status of the deliverances of his notebook), but *rather*,

³⁰Consider, further, that if Otto's endorsement of the reliability of the notebook is not *preserved* via the notebook, but rather, preserved via biological memory, then the endorsement will be 'fleeting' and unstable; this plausibly undermines the epistemic *integrity* of the endorsement. After all, for the endorsement to have the epistemic significance it needs to have in order to positively affect the epistemic status of the deliverances of the notebook, the endorsement must be in some way a *stable* one. Compare this suggestion with the implausible thought that the epistemic status of the deliverances of the notebook is enhanced, in perpetuity, in virtue of a fleeting assent that is forgotten and as such *not* preserved in memory while Otto continues to rely on the notebook.

³¹It might also be argued that Otto can be deemed to reliably endorse the reliability of the notebook during a 'moment of lucidity' with his biological memory, even if his failing biological memory is not in general reliable. However, it's hard to see how much work this move could do, given that an endorsement in a moment of lucidity would be an unsafe method. (In very close nearby worlds, the endorsement at issue derives from a failing memory).

through the *reliable extended process*—the process of writing down information in the notebook and consulting it later. So, perhaps Otto does satisfy (4*) as long as his endorsement of the reliability of his notebook is acknowledged *via* an extended process involving his notebook.

But now a different sort of problem looms. Were Otto to endorse the reliability of his notebook *via* the notebook, he would be endorsing the process of consulting the deliverances of the notebook as reliable *by appealing to the deliverances of the notebook*. But this strategy degenerates into a notorious kind of epistemic circularity. Compare: suppose you are reading a book about Rome, which includes claims A, B and C on which you rely. You then notice a passage in the book saying that all the claims in the book, including the one you are reading right now, are reliable. Does this increase the epistemic standing of your beliefs in A, B and C? If you didn't know A, B and C before, do you then know them *now*? It's hard to see why. The upshot seems to be that (4*) should be replaced with the even more demanding (4**):

(4**) That the reliability of the resource be *reliably* and *non-circularly* endorsed.

Trouble is, Otto seems incapable of satisfying (4**). But then the extent of the problem comes to focus: in order to not drive an epistemic wedge between cases of extended memory and biological memory, where only the latter qualify as knowledge, it's important to show how the Otto case does not leave the notebook looking like TrueTemp's thermometer. We saw that what is needed here is that Otto satisfies condition (4), i.e. that he endorses the reliability of his notebook. However, we also noted that epistemically *defective* endorsement of the notebook will leave Otto no better than TrueTemp*, who endorses the reliability of his thermometer on an improper basis. What is needed then is condition (4*), i.e. that Otto endorses his notebook *via* a reliable process. That means Otto cannot endorse the reliability of his notebook on the basis of his failing biological memory. Instead Otto can satisfy condition (4*) by epistemically relying on the reliability of the notebook itself. Such an approach, however, seems doomed to fail on grounds of epistemic circularity. That is to say, while consulting the notebook is a reliable process, endorsing the process of consulting the deliverances of the notebook as reliable *by appealing to the deliverances of the notebook* betrays a kind of epistemic circularity. On the assumption that epistemic appeal to either Otto's biological memory or extended memory exhausts the options, the unfortunate upshot is that the proponent of extended cognition has run into an intractable jam.

5

In this final section we shall discuss two possible responses to the catch-22 we have outlined. Basically, unless Otto reliably endorses the reliability of his

notebook in a way that epistemically depends *neither* on his own biological memory *nor* on the notebook itself, we won't be able to rule Otto in while ruling TrueTemp and TrueTemp* out in a way that is epistemically non-circular. So, the pressing challenge for the extended cognition theorist is to explain how Otto could possibly satisfy (4**).

However, before we turn to those two responses, we should pause to deal with a concern about *generality* that may immediately spring to mind. The worry is that our catch-22 depends on the details of the Otto case in ways that don't generalize to other cases of extended cognition. In particular, Otto's biological memory is deficient, but this particular feature seems to play no role in other motivating cases such as Clark and Chalmers' 'Tetris' case.

In response, it's correct that our catch-22 hangs on a particular feature of the Otto case in the sense that Otto cannot satisfy (4*) by relying on his failing biological memory. The only other way for Otto to satisfy (4*) seems to involve appeal to the reliability of his notebook, which then implies that his endorsement of its reliability is circular—in violation of (4**). At this juncture, we could distinguish between two types of cases of extended cognition: those where individuals incorporate an external device to compensate for a natural deficiency, and those where individuals incorporate an external device to improve their otherwise normal cognitive functioning. Our catch-22 would thus hold for all those not insignificant cases that belong to the former category. But, importantly, even those cases of extended cognition that belong to the latter category would be subject to (4**). That is to say, cases of extended cognition involving individuals with impeccable biological memory must *still* endorse the reliability of the pertinent external resource in a way that is both reliable and non-circular. The same epistemic issues thus arise for both groups, i.e. that the endorsement be neither epistemically defective nor epistemically circular. The only difference between individuals who suffer from such cognitive shortcomings and those who do not is that the former have fewer epistemic resources available to meet our additional integration conditions. The conditions themselves are in good standing regardless of any such shortcomings.

5.1 *First Response*

One natural response to the dilemma we've sketched will be to argue that it is possible for Otto to endorse the reliability of his notebook without epistemically relying on his failing biological memory. However, if Otto's epistemic resources are confined to his own cognitive and perceptual faculties, it is difficult to see how he might avoid doing so. Here is why. Observe first that the belief-sustaining process involving the notebook is at best conditionally reliable in that such a process includes other beliefs as inputs. We assume (in keeping with Clark & Chalmers' text) that Otto will record only believed propositions in his notebook. In general, a process is conditionally reliable just in case it issues in mostly true beliefs when those input beliefs are true. An obvious way for him to determine

whether the notebook is reliable in that sense is to see whether the notebook, when employed in belief-sustaining processes, leads to a high ratio of true beliefs given true beliefs as input.

Now consider a range of true propositions $p_1, p_2, p_3 \dots p_n$ all of which Otto believes. Suppose at t_1 Otto writes (a sentence expressing) p_1 down in his notebook. Then at t_2 he retrieves p_1 from the notebook after which he endorses that proposition. In that case, the process involving his notebook of sustaining his belief has produced a true belief given a true belief as input. The notebook successfully sustains Otto's dispositional belief in p_1 . Suppose that the same thing happens at t_3 and t_4 with p_2 . Again the same process has sustained a true belief given as input a true belief. And so on until Otto gets to p_n . There is thus no question that the process is reliable. The question is whether Otto is in a position to *reliably* affirm its reliability. Otto would be able to do so only if he can accurately compare notebook entries at $t_1, t_3 \dots t_n$ with what is recorded in the notebook at $t_2, t_4 \dots t_{n+1}$. That way Otto could amass track-record evidence that the notebook was correct on all these occasions from which he can inductively infer that the notebook is reliable. However, amassing such evidence requires of Otto that at $t_2, t_4 \dots t_{n+1}$ he *remembers* his past notebook entries at $t_1, t_3 \dots t_n$, and since Otto's biological memory is fading, that is something he cannot do³².

Perhaps it is wrong to consider Otto in epistemic isolation. The proposal might be that if we extend Otto's epistemic resources to include *other testifying individuals*, then perhaps it would be possible for Otto to endorse the reliability of his notebook without epistemically relying (in a way that is malignantly epistemically circular) on his failing biological memory. Assuming the reliability of these other individuals, they could, perhaps, make up for the epistemic shortfall posed by the detrimental effects that Otto's Alzheimer has had on his biological memory. After all, we ordinarily rely on others for vast amounts of knowledge of the external world. The current thought is that we can successfully do so whether our memory has deteriorated or not, even in cases such as Otto and his notebook. Let's probe into this proposal in more detail.

Suppose that Otto endorses the reliability of his notebook on the basis of some piece of reliable testimony. It is essential that the testimony be reliable. Otherwise the problem about defective epistemic basing returns, as illustrated by TrueTemp*. Imagine that a trusted friend, whose memory is impeccable, comes to know independently that Otto's notebook is reliable. She reliably conveys that information to Otto who on that basis forms a belief in the reliability of

³²The notebook is obviously an overly simplified, toy example. For instance, Otto may be expected to have a large swath of background information about notebooks, e.g. that any information got there by him writing it down, that any written information doesn't move around on its own, which would help him establish the reliability of this particular notebook. Smartphones may be better candidates for cognitive extension, but few have any substantial background information about their inner workings. Could a high-tech savvy version of Otto ascertain the reliability of such a technological device by directly inspecting its hardware and software features without having to gather track-record evidence? Given that he would have to rely on his already acquired knowledge of which such features could make his phone perform reliably, his fading memory would surely impair any such ability he might otherwise have had.

the notebook. That would certainly suffice as an account of how Otto can reliably and non-circularly endorse the reliability of his notebook. So, it looks as if condition (4**) is, after all, satisfied.

Nevertheless, the extended cognition theorist is not home and dry yet. For one thing, consider that unless Otto records the information afforded by the reliable testimony in the notebook (or somewhere else), then the endorsement of the notebook's reliability on the basis of reliable testimony will be a kind of 'fleeting' endorsement that is soon after forgotten. And a fleeting endorsement is problematic for reasons already considered—*viz.*, the epistemic integrity of an endorsement is undermined if endorsement is unstable. But, of course, Otto can stabilise his endorsement of the reliability of the notebook on the basis of reliable testimony by recording it in the notebook (or somewhere else). However, in doing so, Otto will be in a position where Otto's having a stable endorsement depends on his consulting the notebook to affirm that the notebook is reliable—and thus the circularity worry raises its head.

But even despite this issue, there's another problem with the testimony strategy, which we can appreciate by revisiting the TrueTemp case, only this time with the same twist as in the Otto case. Imagine that TrueTemp forms a belief in the reliability of his temperature-detecting device merely on the basis of a reliable piece of testimony. It is by no means obvious that TrueTemp's reliable beliefs about the ambient temperature *thereby* constitute knowledge. Epistemological internalists insist that beliefs reliably produced (or sustained) by strange and fleeting processes are defeated on grounds of *subjective irrationality*³³. As Bonjour (1985: 61) put it, "if the acceptance of a belief is seriously unreasonable or unwarranted from the believer's own standpoint, then the mere fact that unbeknownst to the believer [it was reliably formed] will not suffice to render the belief epistemically justified." The present point is that the reliability of TrueTemp's belief in the reliability of his thermometer cannot offset subjective irrationality either. Given what else TrueTemp believes, he should have been suspicious of this testimony. After all, he lacks good reasons to think he possesses any temperature-detecting ability; indeed he is aware of no other people possessing such seemingly spurious ability. What is needed is not another reliably produced belief, but accessible reasons, or awareness of evidence, which can render the process in question reasonable from TrueTemp's point of view.

The foregoing has implications with respect to how we might amend condition (4**). If we are to exclude TrueTemp's thermometer as a constitutive part of his cognitive process, *mere* reliable (and non-circular) endorsement of the reliability of the resource will not suffice. We need a third constraint that its reliability be rationally endorsed, where rationality is understood in the subjective sense of being reasonable from an internal perspective:

³³Even reliabilists admit that reliably produced (or sustained) beliefs are subject to various types of epistemic defeat, including mental state defeaters. See for example Goldman's (1986: 62-3, 111-2) non-undermining condition. So, both epistemic internalists and epistemic externalists accept that subjective irrationality can defeat beliefs that are otherwise reliably produced.

(4***) That the reliability of the resource be *reliably, non-circularly* and *rationally* endorsed.

Bear in mind that endorsement is meant to be a necessary condition on cognitive integration of some resource into a cognitive system. If endorsing (the reliability of) some resource is irrational from the point of view of the individual, say if certain background beliefs conflict with the deliverances or putative reliability of that resource, then no endorsement can facilitate its integration into a cognitive system. As cognitive integration involves, among other things, explanatory coherence and inferential consistency, such irrationality is precluded between standing attitudes towards the resource and the cognitive system into which that resource is a candidate for integration.³⁴

The foregoing yields the right results in the cases of TrueTemp and Otto. If TrueTemp's endorsement of the (reliability of the) thermometer were not just reliable and non-circular, but also rational in the sense that he was internally aware of the factors that make his beliefs reliably produced, then he could take them into account when rationalizing his linguistic and physical behaviour. For instance, that may involve abandoning other beliefs so as to restore internal consistency in his belief set. Such rationalization would thus help fully integrate his thermometer into his cognitive system. Assuming the absence of any (non-defeated) defeaters, the true beliefs that are formed on the basis of his thermometer would then count as knowledge. The same is true in the case of Otto. If his endorsement of (the reliability of) the notebook were not just reliable and non-circular, but also rational in the pertinent sense, then he could draw on the factors that make his notebook reliable when rationalizing his linguistic and physical behaviour. Again, such rationalization would play a key role in the integration of the notebook into Otto's cognitive system. Further, the true, reliable and undefeated beliefs that are sustained by the notebook would then count as knowledge.

³⁴One might worry that (4***) begs the question against the extended cognition theorist by requiring that the reliability of the resource be rationally endorsed in the sense of being subjectively reasonable. After all, the extended cognition thesis is a form of active externalism, whereas subjective rationality has typically been adopted by epistemic internalists as a constraint on epistemic justification. However, this worry is misconceived. Firstly, we showed in fn. 24 that subjective rationality is *not* a prerogative of epistemic internalists; even paradigmatic epistemic externalists accept that lack of such rationality constitutes a type of epistemic defeat. Secondly, the charge of (4***) being question-begging is premised on the claim that epistemic internalism and HEC are incompatible. But this is false, unless one already assumes *epistemic individualism*, the thesis that what converts true belief to knowledge supervenes on properties of the biological agent. Once it is appreciated that epistemic internalism and HEC are incompatible only against a background of epistemic individualism—a thesis which has been shown problematic in its own right (cf. Goldberg 2014)—we can easily envision epistemic anti-individualist construals of both *accessibilist* (e.g. Chisholm (1977) and Bonjour (1985, ch. 2)) and *mentalist* (e.g. Conee & Feldman (2001)) versions of epistemic internalism; see Carter & Palermos (2014) for a development of this argument, on which it is suggested that our pretheoretical intuitions regarding the incompatibility of active externalism with epistemic internalism are symptomatic of a tacit yet incorrect identification of epistemic internalism with epistemic individualism.

The important difference between TrueTemp and Otto is that only Otto suffers from a serious cognitive shortcoming. The question is whether Otto would be able to rationally endorse the reliability of his notebook given the suboptimal state of his memory. He would have to be internally aware of the factors that make his belief-sustaining process conditionally reliable, as well as be able to rationalize his behaviour in various ways. Assuming that involves, among other things, revising existing beliefs, Otto would then have to draw extensively on his biological memory. Bear also in mind here that memory is a fundamental cognitive process which subserves most other cognitive functions. So, the prospects for producing an account of how Otto could rationally endorse the reliability of his notebook look dim; or so we contend.

5.2 *Second Response*

The second response is to argue that the epistemic circularity that seems to afflict the process of endorsing the reliability of the notebook *via* its own deliverances is *benign*. Epistemic circularity is to be distinguished from logical circularity, where the conclusion of an argument features explicitly as one of its premises, and rule circularity, where an epistemic rule is employed in an argument for the conclusion that the use of that rule is reliable³⁵. In contrast, epistemic circularity is typically formulated in terms of (knowing) a premise to rely on (the reliability of) a belief source in an argument with the conclusion that that source is reliable³⁶. But we can also formulate a notion of epistemic circularity of a process: a process of endorsing the reliability of some (internal or external) resource cannot rely on the deliverances of that resource. So, in the case of Otto, what is the argument (or process) in question? Here are two scenarios: (a) imagine that when Otto first acquired the notebook it said in the preface: ‘This notebook is reliable’. It is hard to see how Otto could come to know (or even justifiably believe) that the notebook is reliable on the basis of reading that statement in the notebook, even if indeed that statement is true. Consider the following:

- (1) The notebook says that it is reliable
- (2) So, the notebook is reliable

Since having knowledge of (or a justified belief in) premise (1) relies on the reliability of the notebook in an argument with the conclusion that the notebook is reliable, the argument (1) – (2) is blatantly epistemically circular.

³⁵Note that, following Vogel (2008), our take on rule circularity makes it a property of arguments. We can also define a notion of rule circularity of belief: a belief that an epistemic rule is reliable cannot be justified by the application of that rule. See also Fumerton (1995: 180) who thinks we can never “use a kind of reasoning to justify the legitimacy of using that reasoning”.

³⁶For more on epistemic circularity, see for instance Alston (1986), Bergman (2004) and Kallestrup (2012).

Consider instead this scenario: (b) Otto writes down the content p of a true belief at time t_1 , and then retrieves and endorses p at t_2 . Otto infers that his notebook accurately preserved a true content between t_1 and t_2 . At t_3 Otto writes down the content q of a true belief, and then at t_4 he retrieves and endorses q . Otto infers that the notebook accurately preserved another true content between t_3 and t_4 . And so on. All these occasions of producing a true belief given a true belief as input constitute inductive evidence that the notebook is conditionally reliable. On that basis, Otto infers inductively that the notebook is conditionally reliable. Consider the following *track-record argument*:

- (3) At t_1 true (believed) content p was recorded in the notebook, and at t_2 p was retrieved from the notebook and then endorsed.
- (4) So, the notebook preserved a true (and dispositionally believed) content p between t_1 and t_2 .
- (5) Repeat with q at t_2/t_3 , r at t_4/t_5 ... n at t_n/t_{n+1} .
- (6) The conjunction of the instances of true (and dispositionally believed) content preservation in (5) provides inductive evidence that the notebook is conditional reliable.
- (7) So, the notebook is conditionally reliable.

The chief problem with this argument concerns the (deductively valid) step from (3) to (4)³⁷. If the notebook were unreliable, Otto would not be justified in thinking the notebook had preserved a true content between t_1 and t_2 . In that case it would not be more likely than not that the content p , which was true when written in the notebook at t_1 , would also be true at t_2 . We standardly take justification to involve objective probability. This means that Otto is justified in believing (4) on the basis of (3) only if the notebook is reliable. But since that is exactly the conclusion (7) of the argument, it suffers from epistemic circularity.³⁸

³⁷For a similar diagnosis of so-called bootstrapping arguments in general see Kallestrup (2012). Vogel (2008) thinks the bootstraps should be undone at the inductive step, in this case from (5)/(6) to (7), but there are several reasons why that's too late. (i) While the track-record evidence in (5) may suffice for knowledge of (7), (4) counts as some justification for (7). (ii) (7) provides the best explanation of the truth of (5), indeed it is hard to see why (5) should be true unless (7) is also true. (iii) If Otto knows independently that the notebook is either anti-reliable (always wrong) or else reliable, then he can deductively infer (7) from (4). S can thus bootstrap knowledge of (7) while bypassing the inductive step. For more details on (i) – (iii), see Cohen (2002), White (2006) and Titelbaum (2010), respectively.

³⁸A separate but related worry is that Otto's failing biological memory prevents him from competently inferring (4) from (3). That is to say, to conclude that the notebook preserved a true content p between t_1 and t_2 , Otto would at the time of retrieval t_2 need to recall what (if any) content was recorded at t_1 . One might try to finesse this point by suggesting that Otto at t_1 also records the time at which he records p . Then at t_2 if Otto cannot retrieve from biological memory the time (t_1) at which he recorded p then presumably he can just consult the notebook! It should be clear by now why this proposal is problematic: it assumes that between t_1 and t_2 the notebook accurately preserved the time at which p was recorded, and so the question arises of how to justify that assumption in an epistemically non-circular way.

The question is now whether the extended cognition theorist might argue that the epistemic circularity, which the arguments (1) – (2) and (3) – (7) display, is somehow benign. Thus, Pryor (2004), Bergmann (2004) and Davies (2008) take the oddity of epistemically circular knowledge to stem from a lack of dialectical cogency of the afflicted arguments.

For instance, Bergmann maintains that such circularity is malignant only if these arguments occur in what he (2004: 719-720) coins “questioned contexts”. In such contexts a track-record argument is propounded with the purpose of settling the question of whether the conclusion is true, thus resolving any doubt about the conclusion. But since prior doubt about the conclusion undermines knowledge of those premises that are known only if that conclusion is true, a doubter cannot use that argument to rationally overcome her doubt. She is rationally obstructed from acquiring knowledge of the conclusion by going through that argument. Consequently, any attempt to use the argument to rationally convince such a doubter is question-begging. In contrast, track-record knowledge is innocent if the argument occurs in “unquestioned contexts” where the arguer has no initial reservations about the conclusion. In those contexts the argumentative purpose is to make explicit the epistemic commitments of those who harbour no antecedent doubt about the conclusion.³⁹

The first remark to make is that even if this strategy were successful, the proponent of extended cognition would still need to provide an independent response to our first worry about rational endorsement. To adequately explain how Otto’s endorsement avoids being epistemically circular in any malignant sense is distinct from explaining how his endorsement is also subjectively rational. The second, and more important, observation is that this strategy looks unpromising in the case of Otto. The reason is that the arguments (1) – (2) and (3) – (7) are dialectically ineffective *even* for Otto who harbours no prior doubt about the conclusions (2)/(7). Assume momentarily that being in an unquestioned context comprises *agnosticism* in the sense of taking no attitude at all towards (2)/(7). Imagine that Otto is initially agnostic about (2)/(7), he knows the premises (1)/(3) – (6), acknowledges the validity of the argument, and on the basis of correctly inferring (2)/(7) from (1)/ (3) – (6), he forms a belief in (2)/(7). Otto could not thereby come to know (2)/(7). The reason is that Otto’s agnosticism rationally prevents him from taking (1)/ (3) – (6) to support (2)/(7). For if Otto did not already accept (2)/(7), he could not be justified in accepting (1)/ (3) – (6). Hence, the arguments cannot be used by Otto to produce rational acceptance of (2)/(7) from a state of agnosticism about (1)/ (3) – (6). Given that knowledge requires rational acceptance, the arguments cannot be used by Otto in a non-questioned context to acquire knowledge of

³⁹Others have argued that epistemic circularity is harmless if the pertinent method or process *cannot possibly* be justified in any non-circular way. For instance, Boghossian (2001) thinks that we are not unjustified in reasoning through modus ponens just because we cannot provide a non-rule-circular justification for modus ponens. Since the circularity that features in the notebook case lacks this feature (e.g. Inga’s biological memory could provide her with non-circular knowledge of Otto’s notebook), we shall henceforth set this approach aside.

(2)/(7)⁴⁰. The upshot is that our arguments are problematic even when they occur in an unquestioned context.

6

We conclude by placing the result developed in this paper into a wider perspective—first by stressing what we do *not* take ourselves to have demonstrated, and then by highlighting why the foregoing problem outlined here is nonetheless a significant one. We have not argued that HEC is false, nor have we claimed that there is no way to ‘make room’ for this thesis in epistemological theory. More would have to be shown to establish either of these claims.

Rather, we have raised a puzzle that cannot be brushed aside in so far as HEC is to hold water in epistemology; in doing so, we have highlighted as well a novel way that epistemic circularity threatens the possibility of knowledge stemming from extended cognitive processes.

We have, admittedly, focused on a single case in the development of our puzzle. Is this problematic? We think not. Firstly, it should be evident from the discussion that the problem raised will apply *mutatis mutandis* to extended memory as a *type* of extended cognitive process. While Otto may be special in the way his memory is failing, the additional integration conditions we have argued for throughout do not hang on this particular detail of the Otto case. If we are right then these conditions must be met in *all* cases of extended memory. Secondly, even if Otto’s case *were* relevantly different from other extended memory cases, it remains nonetheless that this case—famous now in the growing literature on extended cognition—is not *merely* an example, but *the* classic motivating example that is used to support HEC. In so far as the problem we sketch arises for Otto, then given that Otto’s case is widely understood as a paragon of extended memory, the epistemological puzzles that we have raised must be engaged head-on.⁴¹

7 References

- Adams, F., & Aizawa, K. (2008). *The Bounds of Cognition*, Oxford: Blackwell.
- Alston, W. (1986). ‘Epistemic Circularity’, *Philosophy and Phenomenological Research*, 47(1), 1-30.
- Bergmann, M. (2004). ‘Epistemic Circularity: Malignant and Benign’, *Philosophy and Phenomenological Research* 69, 709-727.

⁴⁰Alston (1986) argues along similar lines, albeit not in the context of extended cognition.

⁴¹Thanks to Duncan Pritchard, Orestis Palermos and Andy Clark for helpful discussion. This article was written as part of the AHRC-funded ‘Extended Knowledge’ (#AH/J011908/1) research project that is hosted by the University of Edinburgh’s Eidyn research centre.

- Beebe, J. R. (2004). ‘Reliabilism, Truetemp and New Perceptual Faculties’, *Synthese*, 140(3), 307-329.
- Boghossian, P. (1989). ‘Content and Self-Knowledge’, *Philosophical Topics* 17, 5-26.
- (2001). ‘How are Objective Epistemic Reasons Possible?’, *Philosophical Studies*, 106(1), 1-40.
- BonJour, L. (1985). *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press
- Carter, J.A. & Kallestrup, J. (2015). ‘Extended Cognition and Propositional Memory’, in *Philosophy and Phenomenological Research*, forthcoming.
- Carter, J.A., Kallestrup, J., Palermos, S.O. & Pritchard, D. (2014). ‘Varieties of Externalism’, in *Philosophical Issues* 24(1): 63-109.
- Carter, J.A. & Palermos, S.O. (2014). ‘Active Externalism and Epistemic Internalism’, *Erkenntnis* (Online First: DOI: 10.1007/s10670-014-9670-5).
- Carter, J.A. & Pritchard, D. (2015). ‘Extended Entitlement’ in *New Essays on Entitlement* (eds.) P. Graham & N. Pedersen, (Oxford UP).
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT press.
- Clark, A. & D. Chalmers (1998). ‘The Extended Mind’, *Analysis*, 58: 10–23.
- Clark, A., (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, New York: Oxford University Press.
- (2010). ‘Memento’s Revenge: The Extended Mind, Extended’. In *The Extended Mind*. Menary (ed.) Cambridge, Massachusetts, MIT press.
- Cohen, S. (2002). ‘Basic Knowledge and the Problem of Easy Knowledge’, *Philosophy and Phenomenological Research*, 65(2), 309-329.
- Conee, E. & Feldman, R. (2001). ‘Internalism Defended’, in Ed. Kornblith, H. *Epistemology: Internalism and Externalism*, Cambridge: MIT Press.
- Clark, A., & Chalmers, D. (1998). ‘The Extended Mind’, *Analysis* 58, 7-19.
- Davies, M. (2008). ‘Two Purposes of Arguing and Two Epistemic Projects’ in I. Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*, Oxford: Oxford University Press, 337-383.
- Fine, K. (2012). ‘Guide to Ground’ in Eds. Correia, F. & Schneider, B. *Metaphysical Grounding: Understanding the Structure of Reality*, 37-80.
- Fumerton, R. (1995). *Metaepistemology and Skepticism*. Rowman & Littlefield Publishers.

- Froese, T., Gershenson, C., & Rosenblueth, D., A. (2013). ‘The Dynamically Extended Mind’, available at:<http://arxiv.org/abs/1305.1958>.
- Goldberg, S. (2014). ‘Interpersonal Epistemic Entitlements’, *Philosophical Issues*, 24, forthcoming.
- Goldman, A. I. (1999). ‘Internalism Exposed’, *The Journal of Philosophy*, 271-293.
- (1986). *Epistemology and Cognition*. Cambridge, MA: MIT Press, 1986.
- (1979). ‘What is Justified Belief?’ In *Justification and Knowledge*, by G Pappas, 1-23. Dordrecht: Reidel, 1979.
- Greco, J. (2010). *Achieving knowledge: A virtue-theoretic account of epistemic normativity*. Cambridge University Press.
- Hutchins, E. (1995). *Cognition in the Wild*, Cambridge, MA: MIT Press.
- Kallestrup, J. (2012). ‘Bootstrap and Rollback: Generalizing Epistemic Circularity’, *Synthese*, 189 (2), 395-413.
- Lewis, D. (1972). ‘Psychophysical and Theoretical Identifications’, *Australasian Journal of Philosophy*, 50(3), 249-258.
- Menary, R. (2006). ‘Attacking the Bounds of Cognition’, *Philosophical Psychology* 19, 329-44.
- (2007). *Cognitive Integration: Mind and Cognition Unbound*, London: Palgrave Macmillan.
- Olsson, E. J. (2007). ‘Reliabilism, Stability, and the Value of Knowledge’, *American Philosophical Quarterly*, 343-355.
- Pappas, G. (2005). ‘Internalist vs. Externalist Conceptions of Epistemic Justification’, *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2013/entries/justep-intext/>>.
- Palermos, O. (2011). ‘Belief-Forming Processes, Extended’, *Review of Philosophy and Psychology* 2, 741-65.
- Pryor, J. (2004). ‘What’s Wrong with Moore’s Argument?’ *Philosophical Issues* 14, 349-378.
- Pritchard, D. (2010). ‘Cognitive Ability and the Extended Cognition Thesis’, *Synthese* 175, 133-51.
- Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes*. New York: Cambridge University Press.
- (2009). ‘Extended Cognition and the Mark of the Cognitive’. *Philosophical Psychology*, 22(1); 1-19.

- Rupert, R. (2004). 'Challenges to the Hypothesis of Extended Cognition'. *Journal of Philosophy*, 101: 389-428.
- Schwitzgebel, E. (2014). 'Belief', *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), forthcoming URL = <<http://plato.stanford.edu/archives/spr2014/entries/belief/>>.
- Spaulding, S. (2012). 'Overextended cognition', *Philosophical Psychology*, 25(4), 469-490.
- Sutton, J., Barnier, A., Harris, C., & Wilson, R. (2008). 'A Conceptual and Empirical Framework for the Social Distribution of Cognition: The Case of Memory', *Cognitive Systems Research*, 1-2, 33-51.
- Titelbaum, M. (2010). 'Tell Me You Love Me: Bootstrapping, Externalism, and No-Lose Epistemology,' *Philosophical Studies*, 149(1), 119-134.
- Vogel, J. (2008). 'Epistemic Bootstrapping', *The Journal of Philosophy*, 518-539.
- Wheeler, M. (2005). *Reconstructing the Cognitive World*, Cambridge, MA: MIT-Press.
- White, R. (2006). 'Problems for Dogmatism', *Philosophical Studies*, 131(3), 525-557.
- Wilson, R.A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition*. Cambridge University Press.
- (2000). 'The Mind Beyond Itself', In D. Sperber (ed.), *Metarepresentations: A Multidisciplinary Perspective*. Oxford University Press.
- Zagzebski, L. (1996). *Virtues of the Mind*, Cambridge: CUP.