

## Varieties of Cognitive Integration

J. Adam Carter (*University of Glasgow*<sup>1</sup>)  
Jesper Kallestrup (*University of Copenhagen*<sup>2</sup>)

*Abstract.* Extended cognition theorists argue that cognitive processes constitutively depend on resources that are neither organically composed, nor located inside the bodily boundaries of the agent, provided certain conditions on the integration of those processes into the agent's cognitive architecture are met. Epistemologists, however, worry that in so far as such cognitively integrated processes are epistemically relevant, agents could thus come to enjoy an untoward explosion of knowledge. This paper develops and defends an approach to cognitive integration—*cluster-model functionalism*—which finds application in both domains of inquiry, and which meets the challenge posed by putative cases of cognitive or epistemic bloat.

### §0. Introduction

Increasingly, in order to meet our cognitive goals, we tend to rely not merely on the powers of our naturally endowed cognitive architecture, but also on extra-organismic elements of the world—and in particular, on technological gadgets with which we regularly and uncritically interact. For instance, it is commonplace to no longer store most of the phone numbers and calendar dates we need in our biological memory. Rather, the norm has become to simply ‘offload’ the storage of such information to smartphones, smart-watches, laptops, Google glass, etc. Philosophers of mind and cognitive scientists call this strategy *cognitive offloading*<sup>3</sup>. As Michael P. Lynch (2014) predicts, increased cognitive offloading coupled with increasingly subtle and physically smaller technology suggests that it is really just a matter of time before the majority of the gadgetry we use for cognitive offloading will be by and large seamless and ‘invisible’<sup>4</sup>.

However, some current forms of cognitive outsourcing are subtly more seamless than others, and this is important from the perspectives of both epistemology as well as the philosophy of mind and cognitive science. In epistemology, this matters because propositional knowledge is widely thought to be grounded in *abilities* of the agent<sup>5</sup>. The extent to which our gadgets are integrated—or disintegrated—with the agent's cognitive architecture matters for the purpose

---

<sup>1</sup> COGITO Epistemology Group, Philosophy, University of Glasgow. E-mail: adam.carter@glasgow.ac.uk.

<sup>2</sup> Faculty of Humanities, University of Copenhagen. E-mail: kallestrup@hum.ku.dk.

<sup>3</sup> See, for example, Dror and Harnad (2008) and Clark (2008).

<sup>4</sup> See, along with Lynch (2016), a review of Bostrom (2009) Bostrom and Sandberg (2009) for discussion of some of the latest intelligence augmentation possibilities. A striking example here is Google, which is already moving beyond Google Glass with the 2014 advent of the Google ‘smart lens’ project, which is bringing to the market smart contact eye lenses with tiny wireless chips inside, and with a wireless antenna thinner than a human hair.

<sup>5</sup> Proponents of this position include Sosa (1999; 2007; 2009; 2011; 2015), Greco (2003; 2010; 2012), Zagzebski (1999), Kelp (2012), Turri (2011), Pritchard (2012) and Kallestrup and Pritchard (2012; 2013; 2014).

of whether her beliefs can be credited to her own abilities<sup>6</sup>. Likewise, the issue of cognitive integration is crucial to discussions outside of epistemology, concerning the viability of the hypothesis of *extended cognition*. As Andy Clark and David Chalmers (1998) have influentially argued, the material realisers of cognitive processes can (literally) include extra-organismic parts of the world, located outside the skull and skin. They concede, however, that not just any item we interact with qualifies; certain conditions of cognitive integration must be met for a part of the world to function as part of an *extended* cognitive process, and recent debates about extended cognition have focused on what exactly these conditions should be<sup>7</sup>.

Interestingly, the bar is typically set higher when the integration of external resources into cognitive architecture leads to states of *knowledge* rather than, say, mere belief. One explanation here is that knowledge is a distinctive achievement worthy of credit and with a separate normative dimension *vis-à-vis* action, assertion and so on. If what goes for states short of knowledge also goes for knowledge, not only do we forgo an explanatory account of what's special about knowledge, we also face the risk of *cognitive bloat* (e.g., Rupert 2004) in the epistemic domain. That is, we may have to embrace an explosion of knowledge in that the integration of gadgets, or other external resources, into our cognitive repertoire could lead to an expansion of knowledge far beyond what would otherwise be within our epistemic reach. Does all that withstand scrutiny? This paper offers a way to navigate through these vexed issues; we propose a novel functionalist approach to cognitive integration, one that spans both the domains of the philosophy of mind and cognitive science as well as epistemology.

Here is the plan. §1 briefly introduces the hypothesis of extended cognition. §2 connects issues to do with cognitive integration in epistemology with parallel debates about cognitive integration in the philosophy of mind and cognitive science. §3 argues why the conditions for cognitive integration in the epistemological case, regarding knowledge, are plausibly more demanding, than they are in the latter case, where the issue is primarily a metaphysical rather than an epistemological one. The upshot would seem to be that one set of cognitive integration conditions is knowledge-relevant, another less demanding set is germane to extended cognition. §§4-7 however challenge this simple 'dualist' picture as problematically coarse-grained and proposes and defends an alternative approach to cognitive integration—*cluster-model functionalism*—in its stead.

## §1. The boundaries of cognition

One of the most provocative and polarizing debates in recent philosophy of mind and cognitive science concerns the *boundaries* of cognition. Here is one kind of divisive question: are the numbers and appointments you have stored in your smart phone, for instance, merely 'similar' to some of the numbers and appointments you've stored in your biological memory, or—more provocatively—are they simply a *part* of your memory? On the traditional *intracranial* picture of cognition, there is a straightforward and principled way to answer this question: cognitive processes take place within the head, period (Adams and Aizawa 2001; 2008; 2010). While, granted, we interact with (non-biologically constituted) external gadgetry, the gadgets are not *literally* parts of our cognition or minds. By contrast, proponents of *extended cognition*

---

<sup>6</sup> See, for example, Pritchard (2010). Cf., Carter and Kallestrup (2018).

<sup>7</sup> Menary (2006) provides one approach to this question. See also Heersmink (2016), Palermos (2014; 2017), Carter et al (2014), Kallestrup and Sprevak (2014), and Pritchard (2010).

(i.e. Clark and Chalmers 1998; Clark 2008; Menary 2006; Palermos 2011; 2014; Rowlands 2009) regard this traditional intracranial picture of cognizing as ‘bio-prejudiced’, outdated and unprincipled. The primary argument for extended cognition takes an egalitarian approach, eschewing material constitution and spatial location of cognition for common-sense functionalism or ‘parity’ as the guiding insight. As Clark and Chalmers (1998, 8) put it,

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’.

To think otherwise, on this line of thought, is to give undue weight to constitution or location relative to functional role, with no non-arbitrary reason for doing so.

## §2. Two kinds of cognitive integration: metaphysical and epistemological

In the two decades since Clark and Chalmers’ (1998) initial paper defending the extended mind, the philosophical objection to their radical proposal which has been perhaps the most challenging is ‘the problem of cognitive bloat’<sup>8</sup>. Even if, as extended cognition proponents tell us, a common-sense functionalist approach to the material constituents of cognition is granted as a viable methodology in the metaphysics of mind, the approach is nonetheless problematic if it renders our cognitive lives too inclusive.

To bring this kind of objection into sharp relief, consider the following pair of cases.

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).

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 first case, featuring Otto, is Clark and Chalmers’ own classic case motivating extended cognition. Otto’s notebook is, with reference to the parity principle noted above, playing a functional role that is isomorphic to biological memory *vis-à-vis* information storage and retrieval and is as such regarded as a *constitutive part* of Otto’s memorial process. Furthermore, just as we attribute to individuals with normally functioning memories dispositional beliefs in

---

<sup>8</sup> Adams and Aizawa (e.g., 2001; 2008) would demur here and suggest that the most challenging problem is what they call the *coupling constitution fallacy*. However, proponents of extended mind have available, within their framework at least, straightforward replies to this problem. See, for example, Clark (2010).

virtue of their storing non-occurrent content, so likewise analogously, the contents of Otto's dispositional beliefs are stored in his notebook.

While Adams and Aizawa and other traditionalists about the mind recoil from this diagnosis, it is precisely what proponents of extended cognition think we should say about the case. Though, with reference to this diagnosis, an obvious worry materializes; as Sean Allen-Hermanson (2012) captures the challenge:

If a notebook counts as part of one's mind, then why not the yellow pages, the internet, or even parts of the natural world that supply information and support cognition? (2012, 792)

Notice that the second in the pair of cases, featuring Telo, is superficially similar to the case of Otto. But, if proponents of extended cognition make a parallel diagnosis in the case of Telo, it looks as though Telo will count as (dispositionally) believing everything in the phonebook, which is (according to both sides of the debate) a very surprising if not unacceptable result—thus, the problem of 'cognitive bloat'.

In order to circumvent a parallel diagnosis in cases like Telo's, Clark (2008) has put forward some conditions that are meant to supplement the core functionalist insight. As Clark sees it, conditions like the following 'trust and glue' conditions should rule-in Otto's notebook as part of Otto's memory, but rule-out Telo's phonebook.

*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."

Unlike Otto, who is meant to be understood as carrying the notebook everywhere, and endorsing it in a way that is second nature, Telo's phonebook is merely something he interacts with occasionally, and plausibly more deliberately, seemingly failing (1) and (3). But perhaps this explanation isn't good enough to assuage the worry. There is, after all, a liberal way of reading (1) – (3) on which Telo *does* plausibly satisfy (1) – (3). Indeed, nothing prevents us from describing Telo and Otto as being on a par when it comes to such features as accessibility, availability and reliability of their respective devices. This motivates a search for a fourth condition, or perhaps a fifth or sixth<sup>9</sup>. For example, one might want to add to (1) – (3) the condition, which Clark and Chalmers themselves suggest, though with some ambivalence<sup>10</sup>:

<sup>9</sup> We shall return to this point in §7. See also Carter and Kallestrup (2016) for more discussion.

<sup>10</sup> *Op. cit.*, 17. Note that as long as the information is entered into the device under one's watch, the relevant point may be upon encoding, but in either case there is no requirement that the agent has formed a belief with a content reflecting that information. For original endorsement may only involve mere acceptance of that content, which is widely regarded as a less demanding attitude than outright belief. Rupert (2004) argues that adopting a past-endorsement condition, such as (4), to avoid untoward bloat in Telo-style cases undercuts the motivation for the extended cognition theory, by offering a picture of cognition that seems to fit better into the explanatory

- (4) That any information retrieved is information originally endorsed at the point of entry.

Such an additional constraint, if good, would yield the desired result in that Telo and his phonebook are ruled out, while Otto and his notebook are still ruled in. In any case, it looks like what the proponent of extended cognition needs is some more substantive, clearer conditions which are suitably *demarcative*—viz., which adjudicate particular cases as ones where an external artifact is to be (constitutively) included as opposed to excluded from a cognitive process. In order to meet such a demarcative demand, the proponent of extended cognition needs a story of what it is in virtue of which an external artifact counts as *integrated* within one's cognitive architecture, when it is.

Such a demand for integration conditions is requested in virtue of the import extended cognition has in the philosophy of mind and cognitive science. But, the last few years have revealed an influx of literature which has explored the ramifications of extended cognition in *epistemology*<sup>11</sup>. If extended cognition is true, for instance, should we be regarded as *knowing* the phone numbers stored in our smart phones?

To sharpen this question, let's define some terminology<sup>12</sup>. For some intracranial process  $C$ ,  $C^*$  is  $C$ 's *extracranial analogue* just in case, by reference to Clark and Chalmers' parity principle,  $C^*$  is a cognitive process if  $C$  is a cognitive process. For example, on Clark and Chalmers' diagnosis, the cognitive process employed by Otto, which involves his notebook, is the extracranial analogue of the process of relying on biological memory for information storage and retrieval. Now, let's define *epistemic* extracranial analogues in a similar fashion. For two otherwise epistemically symmetrical cases of cognition,  $A$  and  $A^*$ ,  $A^*$  is an *epistemic extracranial analogue* of intracranial cognitive process  $A$  just in case,  $A$  and  $A^*$  differ in that the process employed in  $A^*$  is, by reference to the parity principle, the extracranial analogue of the cognitive process employed in  $A$ .

Now, here is a question of epistemological importance: *must we accept as cases of knowledge the extracranial epistemic analogues of whatever intracranial cases of cognition we count as knowledge?* Or, more weakly, if (e.g., following Burge 1993; Graham 2012) an individual  $S$  is defeasibly justified in accepting the deliverances of (biological) memory, then is  $S$ 's epistemic extracranial analogue

---

framework offered by the embedded cognition theory. Rupert's main reason for this claim is that adopting the past-endorsement condition assigns a special status to the boundary between organism and environment, because such endorsement is a conscious process situated within the agent's bodily boundaries. In reply, Clark (2007) admits that the extended cognition theorist will embrace a kind of internal privilege in the shape of what he dubs (HOC): human cognitive processing is "organism-centred", but not "organism-bound". His contention is that while cognition leaks out into the extra-organismic environment, the brain and central-nervous system within the organism remain the core and most active element. Nor does that mean the embedded cognition theory is preferable to the extended cognition theory; both play important explanatory roles between which cognitive scientists can flip depending on the problems they seek to solve; or so Clark (*op. cit.*) argues at length.

<sup>11</sup> This strand of literature has emerged almost entirely since 2010 (e.g., Pritchard 2010; Palermos 2014; Carter and Palermos 2015; Carter and Kallestrup 2016; Carter et al 2014; Carter et al, eds. 2018; Vaesen 2011; Kelp 2013a, 2013b; Jarvis 2015).

<sup>12</sup> This terminology is presented originally in Carter (2013).

$S^*$  defeasibly justified in accepting the deliverances of extended memory (i.e., entries in Otto's notebook)?

This is a perplexing question, partly because it's not obvious to what extent a story of *what it is* in virtue of which an external artifact should count as integrated within one's cognitive architecture should be understood as the *same story* as what it is in virtue of which an external artifact should count as integrated within a *knowledge-generating* cognitive ability.

Call the former kind of integration *metaphysical cognitive integration* and the latter *epistemic cognitive integration*. Here is one very natural *meta-theoretical* view about the relationship between these two kinds of cognitive integration: call it the *univocal view*.

UNIVOCAL VIEW: An external artifact  $E$  is *metaphysically* cognitively integrated if, and only if, it is *epistemically* cognitively integrated<sup>13</sup>.

The UNIVOCAL VIEW is elegant, and it has received some support from Palermos (2014) and has been gestured towards by Pritchard (2010), though in the next section, we will raise some problems for this meta-theoretical proposal. As we'll see, the relationship between *epistemic* cognitive integration and *metaphysical* cognitive integration is best understood as one of asymmetric rather than symmetric entailment. Further, we'll see how denials of the univocal view can come in different strengths, some more problematic than others.

### §3. An objection to the univocal view

The univocal view predicts, not implausibly, that once Otto's notebook satisfies whatever conditions (e.g., 'glue and trust') are needed for his notebook to be ruled-in metaphysically as part of Otto's *extended cognitive process*, then, the deliverances of Otto's consulting the notebook *thereby* (absent any defeaters) are candidates for *extended memory knowledge* (that is: knowledge supported by Otto's extended memorial ability).

Perhaps the primary consideration which marshals support for the univocal view is that an obvious *lack* of cognitive integration seems to not only (i) prevent an external artifact from being included into one's cognitive process, by the lights of extended cognition, but *also* (ii) prevent an individual from counting as *knowing*, by the lights of the *ability intuition* on propositional knowledge (e.g., Greco 2003; 2010; Pritchard 2010; 2012; Kallestrup and Pritchard 2012; 2013; 2014).

According to the ability intuition on knowledge, a subject  $S$  knows that  $p$  only if  $S$ 's believing that  $p$  correctly is, at least to some extent, attributable to  $S$ 's cognitive ability<sup>14</sup>. One way to

---

<sup>13</sup> Pritchard (2010) has defended a version of this kind of view. See also Palermos (2014). Note that it doesn't follow from the fact that if something  $E$  is epistemically cognitively integrated that any belief formed on the basis of  $E$  would thereby qualify as knowledge. Compare: even in the default case of healthy biological memory, the deliverances of memory are oftentimes compresent with defeaters which undermine would-be knowledge. When something is cognitively integrated, it is part of a process that can in the right circumstances generate knowledge unlike (for example) processes like Mr. TrueTemp's the deliverances of which the subject cannot take credit for.

<sup>14</sup> Note that this articulation of the ability intuition is importantly weaker than the formulation which is embraced by *robust virtue epistemologists* (e.g., Greco 2010; 2012; Sosa 2009; 2011; Zagzebski 1996) according to whom

appreciate why a lack of cognitive integration can matter with respect to whether one satisfies the ability intuition on knowledge is with reference to Keith Lehrer's (1990) famous case of Mr. 'TrueTemp':

TRUETEMP: TrueTemp has (though entirely unbeknownst to him) a temperature-detecting device implanted in his head that regularly produces accurate beliefs about the ambient temperature. While TrueTemp reliably forms true beliefs about the temperature, he can't tell you why.

TrueTemp is reliable, but the near-consensus view in mainstream epistemology is that TrueTemp lacks knowledge<sup>15</sup>. After all, with reference to the ability intuition, the temperature-detecting device which causes him to form true beliefs, is independent of TrueTemp's own ability; it is, in an obvious sense *disintegrated* from TrueTemp's own cognitive psychology, or what Greco (e.g., 2003; 2010) calls "cognitive character".

Putting all this together, it seems *prima facie* plausible that the very kind of *lack* of integration of the thermometer in TrueTemp's cognitive architecture which accounts for why (with reference to the ability intuition) TrueTemp lacks knowledge would *not be lacking* were the thermometer to satisfy whatever cognitive integration conditions are demanded for extended cognition. For example, the case of TrueTemp arguably satisfies Clark's 'trust and glue' conditions in terms of the device being reliably available and typically invoked, as well as the retrieved information being easily accessible and automatically endorsed. Any lingering worry that TrueTemp lacks knowledge would be a worry that Clark's *specific* conditions fall short as adequate conditions of integration. Thus, reflection on cognitive integration and the ability intuition in epistemology lends some initial support for thinking that the UNIVOCAL VIEW is correct. And, if it were, this would be a convenient bridge between on the one hand the philosophy of mind and cognitive science, and on the other hand epistemology.

But an objection materializes. Consider first that knowledge asymmetrically necessitates belief. This is not just because knowledge, but not belief, is factive, but also because only knowledge involves an epistemic dimension (roughly: whatever converts true belief into knowledge), the satisfaction of which may call for more stringent conditions of integration. Exactly what that dimension amounts to is a vexed question we shall not settle here, except to note that a cognitive ability is arguably knowledge-apt only if it reliably generates true beliefs.<sup>16</sup> Perhaps then, and contrary to the UNIVOCAL VIEW, integration into one's *knowledge-generative cognitive*

---

knowledge demands that one's cognitive ability be the *primary* or *overarching* explanation for one's cognitive success. As has been suggested in recent work by, among others, Kallestrup and Pritchard (2012; 2013; 2014) simple testimony cases in friendly environments (e.g., Lackey 2007) and barn-façade style cases featuring environmental luck pose challenges to robust virtue epistemology. For a recent defence of the ability condition on knowledge, in conjunction with an extended cognition model of the mental, see Wheeler 2017.

<sup>15</sup> Note that this is not to say that there is any sort of consensus against reliabilist approaches in epistemology more generally. The verdict that TrueTemp knows is one that is issued by flat-footed reliabilism—viz., the view that knowledge is reliably produced true belief. While this version of reliabilism is not popular within mainstream epistemology for well-rehearsed reasons, other forms of reliabilism (e.g., virtue reliabilism) enjoy mainstream appeal, and are not committed to ruling TrueTemp as knowing.

<sup>16</sup> See, for example, Sosa (2010), Greco (2003) and Kelp (2017) for representative discussion.

*ability* requires some further conditions, conditions that needn't be met for an external artefact to be ruled-in merely as a part of one's extended cognitive process<sup>17</sup>.

This idea can be fleshed out with the following case:

MEMORAID: Otto\* is fitted with a remarkable chip called the *Memoraid*, developed to circumvent the need for Alzheimer's patients to use manual notebooks for cognitive offloading. Once the *Memoraid* chip is installed in the subject's prefrontal cortex, it *seems* to one that one is storing one's information normally, in one's biological memory; though, information is actually being stored on the *Memoraid* chip. Likewise, retrieving information from the *Memoraid* chip is subjectively indistinguishable from retrieving information from biological memory. Otto\*'s family felt it would be best not to tell Otto\* that the *Memoraid* chip was installed, and so the entire procedure was done while Otto\* was sleeping. Otto\* is, thus, completely unaware that he is relying on the *Memoraid*—he has never even heard of this technology—and continues to believe that his memorial process involves him storing information in, and retrieving information from, his biological brain.

Here are two important points to glean from MEMORAID:

- (i) Otto\* in MEMORAID is like Clark and Chalmers' Otto in that the *Memoraid* plausibly satisfies conditions for (metaphysical) cognitive integration. The *Memoraid* is, after all, playing the same functional role as Otto's notebook *vis-à-vis* memory storage and retrieval.
- (ii) Otto\* is, *epistemically*, much like TrueTemp. For one thing, both have devices implanted in them, unbeknownst to them, which are reliable. Whereas TrueTemp's thermometer was unbeknownst to him the source of the reliability of TrueTemp's temperature beliefs, Otto\*'s *Memoraid* is unbeknownst to him the source of the reliability of Otto\*'s memorial beliefs. Although Otto\*'s *Memoraid* is reliable, Otto\* fails to *know* (even if he counts as being subjectively justified, or justified by epistemic internalist lights) the deliverances of the *Memoraid*, and this for reasons that are analogous to why TrueTemp fails to know<sup>18</sup>.

---

<sup>17</sup> For further examples see Carter and Kallestrup (2018).

<sup>18</sup> Pritchard (2010, 145), discussing Clark and Chalmers' original case, remarks that 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.' This is the broad insight—one that is familiar from meta-incoherence cases (e.g., Sosa 1983) in the reliabilist literature—behind (ii) in the diagnosis of MEMORAID; Granted, Otto\* would be a candidate for *prima facie* justification as well as knowledge if his belief derived from a *basic* belief-forming process. As Graham (2012) helpfully puts it, perception is a basic belief forming process when functioning normally, because perception has the reliable production of beliefs as an etiological function, and this is because in the case of perception, forming-true beliefs reliably partly *explains* the persistence of the process (2012, 449). Note that TrueTemp's and Otto\*'s belief forming processes would hardly seem to qualify given that *ex hypothesi* they are *sui generis* types that are new additions to TrueTemp's and Otto\*'s cognitive architecture; at least, these types would not be recognized as basic by epistemologists, who generally—and regardless of whether they accept explanations for basicness like that which Graham gives—generally countenance only a handful of basic belief-forming processes.

To the extent that (i) and (ii) are correct, then, plausibly, *even though* Otto\*’s *Memoraid* is (metaphysically) cognitively integrated<sup>19</sup>, it looks as though some further integration condition would need to be met in MEMORAID so that Otto\* is no longer in a relevantly analogous (epistemic) situation as Lehrer’s ‘TrueTemp’, who lacks any appreciation of the existence, let alone the reliability, of the *source* of his beliefs (e.g. meta-incoherence). This would plausibly involve some kind of additional endorsement condition, at least to the effect that Otto\* becomes aware that the *Memoraid* is what is issuing his memorially retrieved beliefs; or, at least, that he no longer believes falsely that the source is something that it is not<sup>20</sup>.

In short, those who find the above diagnosis of MEMORAID appealing will be inclined to think that the conditions for *epistemic* cognitive integration might be more demanding than the conditions for *metaphysical* cognitive integration. Consequently, they will very naturally reject the univocal view in favour of a different kind of meta-theoretic view about the relationship between metaphysical and epistemic, cognitive integration conditions; a view we can call *dualism*.

DUALISM: There is *one* set of cognitive integration conditions for inclusion into a cognitive process, and another, slightly more demanding, set of conditions for inclusion into a *knowledge-generative cognitive ability*, where both sets provide necessary and sufficient conditions for the two respective types of cognitive integration.

But we now want to suggest that even *dualism* is too coarse-grained. The suggestion that there is just one set of cognitive integration conditions necessary and sufficient in the case of integration into a cognitive process, and another set of conditions necessary and sufficient in the case of integration into a knowledge-generative cognitive ability is at best naïve, even though the second set includes as necessary all the conditions in the first set.<sup>21</sup>

#### §4. Beyond (coarse-grained) dualism

We wish to motivate an alternative picture to both the UNIVOCAL VIEW as well as to DUALISM, one on which the attempt to produce necessary and sufficient conditions for the metaphysical and epistemological cognitive integration (respectively) of an external artifact is dropped.

---

<sup>19</sup> At least, with reference to the parity principle, the *Memoraid* is cognitively integrated provided Otto’s notebook is; both are playing the same functional role vis-à-vis memory storage and retrieval as biological memory. To question whether MEMORAID displays cognitive integration is to raise a similar question mark over OTTO and the common-sense functionalism that underpins both. Of course, one difference between TrueTemp and Otto\* is that while the thermometer constitutes a cognitive enhancement on the part of TrueTemp, *Memoraid* merely serves the purpose of replacing an existing cognitive capacity on the part of Otto\*, but such considerations have typically not been part and parcel of any relevant integration conditions. Or we could imagine the *Memoraid* being infallible, or at least more reliable than biological memory, so that this case would also constitute an enhancement of existing cognitive abilities.

<sup>20</sup> At the very least, such a condition will preclude cases where the agent is oblivious to the inclusion of the device/artefact in question. The matter of how to spell out the conditions of endorsement, the character of the conscious experience that features in endorsement and the normative dimension of endorsement, are not something we will attempt to settle here.

<sup>21</sup> We shall return to the univocal and dualist views in §7. A more extreme type of dualism has it that the two sets of cognitive integration conditions have no conditions in common. Such a view should strike anyone as implausible, including those epistemologists who think knowledge is a sui generis mental state, not decomposable into truth, belief and something else.

Basically, the reasons are two-fold. Firstly, reductive analyses of concepts (or properties) in terms of producing conditions necessary and sufficient for their application (or instantiation) have a poor track-record in philosophy. The concepts of knowledge, causation and moral goodness are cases in point. Any initially proposed analysis has been met with counterexamples, as has any more refined analysis aiming to fix the former in light of those earlier counterexamples. Nothing suggests the concept (or property) of cognitive integration is an exception, whether or not we acknowledge, as the dualist insists, two distinct types of cognitive integration. So, if induction in philosophy is anything to go by, such a traditional project should be jettisoned. The second reason is that since there is huge variation in the kinds of external resources that could potentially be cognitively integrated into an agent's cognitive process, not just in material constitution, but also in the functional role these resources play in the relevant cognitive architecture, the conditions required for such integration are likely to also vary significantly in those respects. To think that one set of integration conditions would be necessary and sufficient for *all* resources in *all* cognitive processes is simply not empirically credible. What is more, if we are to also consider conditions for inclusion into a knowledge-generative cognitive ability, there would be even less reason to think that a set of necessary and sufficient conditions apply across all epistemic cases of cognitive integration, such that a proper subset of these conditions apply to all the metaphysical cases.

Instead we contend that the best way to settle the question of whether an external resource is integrated into a cognitive system is to determine whether it plays a characteristic functional role within that system, where that role is to be spelled out in some way other than in terms of necessary and sufficient conditions. Thus Otto's notebook is part of his extended memorial process just in case it plays the same functional role as biological memory plays in most non-Alzheimer patients; where unpacking this role is a matter of detailing folk-psychological, environmental (input) stimuli, behavioural (output) responses and causal relations to other mental states and processes. Importantly, on this functionalist account, the requirements for cognitive integration don't constitute necessary and sufficient conditions.

We will now proceed to develop a positive account of how to understand such a functional role, first in the case of metaphysical cognitive integration and then conclude by turning to epistemological cognitive integration.

## §5. Common sense functionalism

The original case for the extended mind thesis, as proposed in Chalmers and Clark (1998), turned on accepting a commonsense functionalist conception. Basically, some resource counts as a part of a mental state or cognitive process in virtue of what it does and what is done to it, rather than in virtue of what it is made out of, or where it is located. Thus, an ordinary notebook located outside Otto's skin-and-skull counts as part of his memorial process in virtue of being part of a system that plays the functional role that biological memory plays in non-Alzheimer patients. Material composition and spatial location are inessential to the question of whether some resource, e.g. a notebook, partially constitutes some such state or process. What matters is the functional role it plays *vis-à-vis* being caused by input stimuli, causing output responses, and standing in causal relations to other mental states or processes. To uncover the details of a functional role is, on this view, to explicate a tacit common-sense conception of what it characteristically takes for something to play that role. Following Ramsey

(1929), Lewis (1970; 1972) and Jackson (1998), mental and other theoretical properties (or states or processes) can be identified with physical or other observational properties (or states or processes) via a three-stage process.

First, the target property is worked into shape for identification by construing it in terms of the functional role it plays in folk theory. The idea is here that a theoretical property (or term) can be *implicitly defined* by the folk theory in which it occurs, in that its meaning is given by its place in a network of input, output and internal role clauses which together comprise a characteristic functional role.<sup>22</sup> Suppose we list all the platitudes comprising this role, i.e., the common-sense truths which speakers recognise as such if prompted, for some folk theory T as it pertains to term  $t$ . That is our postulate:

$$(i) \quad T(t)$$

By existentially generalization, we then arrived at the *Ramsey sentence* of T:

$$(ii) \quad \exists x T(x)$$

Such Ramsification has the effect of creating a single sentence containing all the platitudes in which  $t$  features without any occurrences of  $t$  itself. We can then define  $t$  by forming the *Carnap sentence* of T as the Ramsey sentence conditional on the postulate:

$$(iii) \quad \exists x T(x) \rightarrow T(t)$$

The Carnap sentence is designed to be *a priori* knowable, as it merely says that  $t$  refers to a realization of T *if* there is at least one. Following Lewis (*op. cit.*), the Carnap sentence can be taken to implicitly define  $t$ , at least if T has a unique realiser. If there is no realiser of T, the Carnap sentence is silent about the reference of  $t$ , and if T is multiply realised, the Carnap sentence says that  $t$  refers arbitrarily to one of the realisers. To fix these problems, Lewis (*op. cit.*) recommends two amendments. The first is to modify the Ramsey sentence such that  $t$  refers to a realization if it is unique:

$$(iv) \quad \exists x [T(x) \ \& \ \forall y (T(y) \rightarrow x = y)] \rightarrow T(t)$$

And the second is to add that  $t$  is empty if T has no realization:

$$(v) \quad \neg \exists x [T(x) \ \& \ \forall y (T(y) \rightarrow x = y)] \rightarrow t = \star$$

where  $t = \star$  means that  $t$  lacks a reference. Putting (iv) and (v) together we arrive at the following definition of  $t$ , where  $\mathfrak{r}$  is the description operator:

$$(vi) \quad t = \mathfrak{r}x T(x)$$

Let  $t$  be a mental term or property ( $M$ ) and let T be folk psychology. Thus, (vi) reads:

---

<sup>22</sup> To simplify things, we shall talk interchangeably about properties and the terms that pick them out.

(vi\*)  $M =$  the property that plays the  $M$ -role

The second stage in this model for psycho-physical identification is that relevant branches of science discover that physical property  $P$  uniquely plays the  $M$ -role:

(vii)  $P =$  the property that plays the  $M$ -role

The third stage is then to identify  $P$  with  $M$ , via transitivity of identity:

(viii)  $P = M$

Thus, it is supposedly *a priori* that memory (if it exists) is the state that plays the memory role, where that role is fleshed out in terms of common-sense truths, such as seamless retrieval of stored information, or what have you.<sup>23</sup> Suppose we then learn by drawing on relevant empirical findings that being a notebook (in conjunction with relevant biological features) is the property that plays the memory role. While that will be true of Otto, different external resources might play the memory role for different Alzheimers patients, not to mention the brain systems (hippocampus and adjacent cortex), which play that role for the rest of us. To avoid the identification of  $M$  with an arbitrary realiser of the  $M$ -role,  $M$  can be relativized to species, sub-species or even individuals. In any case, on the basis of these two premises, we can via inference come to know *a posteriori* that being a notebook (plus such features) is identical with the memory property (in Otto).

Here are two additional remarks on the Ramsey-Lewis-Jackson model. First, we assumed that Carnap sentences are *a priori* knowable, but that is by no means uncontroversial; see e.g. Block and Stalnaker (1999). Fortunately, nothing in the following hangs on this assumption, i.e., our account of cognitive integration is compatible with the relevant common-sense truths being widely known, yet only *a posteriori* so. Second, we focus on *filler-functionalism* whereby mental properties are identical with first-order physical properties. In contrast, *role-functionalism* takes mental properties to be second-order functional properties, e.g., memory is the property of having a (first-order physical) property that plays the memory-role. The latter view is designed to better accommodate the multiple realisability of the mental (though we contend that suitable relativization to species or individuals would suffice on the former view), but also faces familiar problems about the causal inefficacy of second-order mental properties.

## §6. Cognitive integration revisited

According to common-sense functionalism, characterising a property in terms of the functional role it plays in folk theory is a matter of identifying a set of platitudes comprising that role. We took that to be our postulate T( $t$ ). Following Lewis (*op. cit.*), the relevant platitudes

---

<sup>23</sup> One may hold that seamlessness of integration of an information storage device is a feature we know for sure is typically true of human memory without being definitive of the memory role, especially when the relevant notion of integration is epistemic. We recognize the distinction being drawn here but maintain that seamless retrieval of stored information is one among a plurality of common-sense truths constituting the memory role, which means that a state may play that role (in humans) even though such retrieval is not seamless. A specific argument for this claim would take us too far afield, but see Jackson (2000) for more general considerations.

are the truths that everyone knows, and that everyone knows that everyone knows. They are common knowledge in that sense. Importantly, the relevant set of platitudes is not such that  $t$  refers to a property realising  $T$  if and only if all and only those platitudes are true of that property. This set does *not* constitute necessary and sufficient conditions for a property to be picked out by  $t$ . Lewis (*op. cit.*) recommends instead that these platitudes form a *cluster*: a disjunction of conjunctions of most of them. That way  $t$  may pick out a specific property even though some of the associated platitudes are false of that property, as long as *most* of those platitudes are true of it.

Against this background, consider now common-sense functionalism regarding cognitive integration. On such a view, the platitudes pertain to conditions for the integration of some external resource into a cognitive system. We have already mentioned relativisation to species, sub-species or individuals, but we should also expect widespread variation across different *aspects* of a cognitive system. After all, cognitive integration into a *perceptual* system will differ significantly from cognitive integration into a *memory* system, and so on. From now on, cognitive systems are to be understood more narrowly in terms of such sub-systems. To simplify matters, let's also stipulate that only three platitudinous conditions are commonly known, such as Clark's three 'trust and glue' conditions.<sup>24</sup> We can then say that for a cluster of integration conditions  $C1$ ,  $C2$  and  $C3$  (i.e. a disjunction of conjunctions of most of  $C1$ ,  $C2$ ,  $C3$ ), resource  $r$  is integrated into cognitive system  $S$  if and only if  $r$  satisfies  $C1$  and  $C2$ , or  $C2$  and  $C3$ , or  $C1$  and  $C3$ . So, it could be that external resource  $r^*$  is integrated into  $S$  for Otto in virtue of satisfying  $C1$  and  $C2$ , while biological resource  $r$  is integrated into  $S$  for non-Alzheimer's patients in virtue of satisfying  $C1$  and  $C3$ . The key point is that the *same cluster* of integration conditions applies to extended and non-extended cognition. Otto has no less a claim to be a case of memory than ordinary cases of biological memory.

Reflect that on the cluster model we have considered so far, there is a clear-cut boundary between integration and non-integration such that either  $r$  is fully integrated into  $S$ , or else  $r$  is not integrated into  $S$  at all. More precisely, if  $r$  satisfies either of  $C1$  and  $C2$ , or  $C1$  and  $C3$ , or  $C2$  and  $C3$ , then  $r$  is integrated *tout court*; if instead  $r$  satisfies neither of these three conjunctions, then  $r$  fails to be integrated to any degree or extent at all. A near-integration in which  $r$  satisfies only one of  $C1$ ,  $C2$  or  $C3$  is no integration at all. Such clear-cut boundaries may be a palatable implication of our cluster model in so far as mental state realization is concerned. After all, such realization is often a binary matter. Physical state  $P1$  cannot realize mental state  $M$  to a greater extent or a higher degree than physical state  $P2$ .  $P1$  may constitute a *partial* realizer in

---

<sup>24</sup> To be clear, our claim is not that the folk will be able to easily explicate all the various conditions on cognitive integration that pertain to various sub-systems of the cognitive system. For we should not expect the folk to be able to articulate (on the basis of concepts possessed or intuitions elicited) complete and accurate platitudinous beliefs about what it takes for a perceptual, memory, inferential, etc., system to be integrated at the right level of architectural specificity. Rather, the relevant platitudes would need to be refined and systematized by cognitive scientists and philosophers, and perhaps, as we suggested in §5, they would also draw on empirical findings and thus be *a posteriori* knowable. That is not to recommend adopting psycho-functionalism. For we maintain that the folk would, at least if prompted upon careful reflection on possible cases, embrace these platitudes when thus worked into shape. None of the conditions (1) – (6) we consider escape commonsensical assent, but rather represent (improved) characterisations of implicit folk opinions. Here we are grateful to an anonymous referee.

the sense of being necessary but insufficient for the realization of M, but P1 does not thereby realize M to any degree or extent at all.<sup>25</sup>

But, in contrast with mental state realization, cognitive integration is best understood as *gradient*<sup>26</sup>. Thus, if P1, P2 and P3 are conditions on integration, then  $r$  is integrated into S (say for individual  $i$ ) to a greater extent or higher degree than  $r^*$  is integrated into S (for  $i^*$ ) if  $r$  satisfies C1 and C2, but  $r^*$  merely satisfies C1. Likewise,  $r$  is integrated into S (for  $i$ ) to a lesser extent or lower degree than  $r^{**}$  is integrated into S (for  $i^{**}$ ) if  $r$  satisfies C1 and C2, but  $r^{**}$  satisfies C1, C2 and C3. This holds in both the metaphysical and epistemic cases of cognitive integration. So, the cluster model needs revamping such that the more conditions  $r$  satisfies, the more integrated  $r$  is into S, and the fewer conditions  $r$  satisfies, the less integrated  $r$  is into S. A specific number of conditions on integration mark no threshold above which  $r$  is integrated *simpliciter* into S, and below which  $r$  fails outright to be integrated into S. On a more plausible picture, cognitive integration is not an all-or-nothing matter.

Let's therefore revise common-sense functionalism regarding cognitive integration in the following way: for a cluster of integration conditions C1, C2 and C3, if external resource  $r$  is integrated into cognitive system S (for  $i$ ) to extent or *degree*  $n$  in virtue of satisfying exactly one of C1, C2 and C3, then  $r^*$  is integrated into S (for  $i^*$ ) to extent or degree  $n+1$  in virtue of satisfying exactly two of these conditions, and  $r^{**}$  is integrated into S (for  $i^{**}$ ) to extent or degree  $n+2$  in virtue of satisfying all of C1, C2 and C3. Importantly, satisfying an exact number of these conditions constitutes no cut-off point at which  $r$  is considered integrated in a binary sense into S. Unlike our earlier, unadorned formulation, these modifications thus accommodate the *gradiance* of cognitive integration; they set out a continuum along which the extent or degree of such integration is determined.<sup>27</sup>

Three observations are in order at this juncture. First, our revised common-sense functionalism preserves the cluster spirit of the original proposal in that integration to a greater extent or a higher degree is a matter of satisfying (at least) one of the disjuncts that form the

---

<sup>25</sup> Degree of belief is no exception: one can believe to degree .7 that  $p$ , but one cannot more or less believe to degree .7 that  $p$ .

<sup>26</sup> See Sterelny (2010) for an approach to scaffolded cognition according to which the degree of embeddedness is plausibly viewed as gradient. Sterelny's model in which gradiance figures is, to be clear, one that he takes to be an alternative to the extended cognition model, and one which draws from interesting work on niche construction. It is not our objective here to attempt to establish that extended cognition is a better model of the mind than other forms of non-extended embedded cognition. We are rather attempting to make sense of the relationship between the kinds of integration that are apposite to extended cognition and epistemology.

<sup>27</sup> Building on Gaut's cluster view (2000), Longworth and Scarantino (2010) suggest a disjunctive (pluralistic) definition of art in terms of a disjunction of minimally sufficient conditions, i.e. sets of non-redundant properties that are jointly sufficient, for being art. These are not strictly INUS conditions as any given condition may be individually necessary or sufficient. On our model, the disjuncts are also individually sufficient and disjunctively necessary. Note in particular that Meskin's (2007) irrelevant properties objection to Gaut's view would neither apply to our model, as an irrelevant property would not play a causal-functional role towards cognitive integration. However, our model diverts from Longworth and Scarantino by allowing for gradability, as underlined by the variable weightings we assign to different properties. As an anonymous referee observed, that may suggest some affinity with Dean's (2003) prototype theory of art, but ours seeks to offer metaphysically constitutive conditions on cognitive integration; even our talk of "epistemic cognitive integration" pertains to constitutive conditions, and not merely ways of epistemic categorization. Moreover, the prototype theory faces Machery's (2009) selection problem of demarcating those properties that are relevant to categorization, or metaphysically constitutive (of art), from the set of those that are merely typical (of art). We touch on this problem in fn. 23.

disjunction of conjunctions of most of the conditions. Integration to a lesser extent or lower degree is then a matter of satisfying a disjunct of fewer conditions. Second, in the interest of simplicity we are assuming that each of C1, C2 and C3 counts *equally* towards the *maximum* integration of  $r$  into  $S$  via satisfaction of all of C1, C2 and C3. As we shall illustrate in §7, we may wish to waive that assumption in specific cases so as to let one condition carry more weight vis-à-vis integration than some other condition. Any such variable weightings would be explainable by particular properties of the cognitive system in question, rather than any formal features of our model. In that case, gradience need not consist in simply *counting* the number of conditions satisfied. Thus, if in a given case C2 plays a greater integrational role than C1, then a situation is possible where  $r$  is integrated into  $S$  (for  $i$ ) to extent or degree  $n$  in virtue of satisfying C1 and C3, while  $r^*$  is integrated into  $S$  (for  $i^*$ ) to extent or degree  $n+1$  in virtue of satisfying C2 and C3, or C2 and C1, or perhaps even just C2. Third, following Rupert (2004), any appeal to common-sense functionalism must deal with *anti-extended conditionals*, such as ‘if  $x$  is your memory of the football game,  $x$  is not something you can see with the naked eye’. Here we must distinguish folk resistance to the Ramsey sentence  $\exists x T(x)$  itself from folk resistance to an extended system realising  $T$ . True, no Ramsey sentence expressing folk psychological theory should assert or imply visual perception of a memory state, but our conditions (1) – (6) do neither, in fact they are entirely *neutral* with respect to the possibility of extended cognition. The folk may well generate such conditionals, but these should not make their way into the Ramsification of common-sense psychology, as they do not capture any aspects of the relevant causal-functional roles. What such conditionals do express is rather scepticism about wide realizations of  $T$ , but at this juncture we would insist that such scepticism is either ill-founded (e.g. keep in mind that what realises the memory role in Otto is the *combined* system of notebook and parts of his biological brain), or else unwarranted, expressing an unprincipled bio-prejudice or chauvinism.

Bearing in mind what we said in §§2-3 about the metaphysical and epistemological kinds of cognitive integration, we shall now put our preferred functionalist account of cognitive integration to use by considering representative cases of each kind. Cases of particular interest are those that seem to involve cognitive bloat or even an explosion of knowledge.

## §7. Revisiting Cases

In §2 we asked the question of whether Otto and Telo should be considered cases of extended cognition. With reference to Clark’s (2010, 46) ‘trust and glue’ conditions, we took the answer to be affirmative in the first case but negative in the second case. That is to say, Otto’s notebook is (1) reliably available and typically invoked, (2) any information retrieved is automatically endorsed, and (3) such information is easily accessible as and when required, but as described Telo’s phonebook fails (1) and (3). On that basis, intuitions are preserved in that Otto’s notebook is ruled in while Telo’s phonebook is ruled out. Otto’s notebook is a prime example of extended cognition (if anything is), but Telo’s phonebook illustrates a widespread worry about cognitive bloat in that all too easily is cognition expanded far beyond what is reasonably regarded as its natural boundaries. Surely, Telo has no dispositional beliefs about most of the names, addresses and phone numbers in the Verizon phonebook, since only a few

of them have actually crossed his mind.<sup>28</sup> Problem is if (1) – (3) are taken to constitute necessary and sufficient conditions for cognitive integration, as is the case if the univocal or dualist views from §§ 2-3 were true, the cases of Otto’s notebook and Telo’s phonebook could be tweaked such that, counterintuitively, the former is ruled out while the latter is ruled in. Imagine a case of Otto and his notebook in which (2) and (3) are satisfied, but Otto fails as a matter of fact to typically invoke his notebook, perhaps because his Alzheimer’s disease and corresponding notebook entries only pertain to peripheral aspects of his everyday life. Call this agent Otto\*\*. It’s hard to see why infrequency *in itself* should present an insurmountable obstacle to such a device counting as part of his memorial process. Or imagine a case of Telo and his phonebook where (1) – (3) are all satisfied. In particular, Telo carries his phonebook everywhere, and endorses its entries without any hesitation. Call this agent Telo\*. Such amendment would surely not alleviate any concern about untoward cognitive expansion in that Telo\* would be ascribed a vast number of beliefs about phone numbers, etc., as preserved by his extended cognitive processes. The upshot is that (1) – (3) should be deemed neither necessary nor sufficient for an external resource to count as part of a cognitive process.

The cluster model approach, however, better accommodates our intuitions even about these revised cases. Thus, in the case of Otto\*\* and his notebook, not only are original integration conditions (2) and (3) satisfied, let’s also assume that our additional condition (4) from §2 is satisfied:

- (4) That any information retrieved is information originally endorsed at the point of entry.

That means the degree or extent to which the notebook is integrated into Otto\*\*’s cognitive process is high or great, although not quite as high or great as the degree or extent to which Otto’s notebook is integrated into *his* cognitive process. Otto, remember, satisfies all of (1) – (4). Importantly, in so far as we are considering extension of *memorial* (dispositional) beliefs, (4) carries relatively more weight. In the normal run of things, one cannot now remember believing a content one has never endorsed in the past. By parity of reasoning, the case of Telo\* and his phonebook is also a candidate for extended memorial beliefs. Since (1) – (3), but not (4), are satisfied in this case, the extent or degree to which the phonebook is integrated into his cognitive process is relatively small or low. And since sufficient cognitive integration is necessary for such process to preserve beliefs about the contents of the phonebook, assigning (4) relatively more weight helps defuse the threat of cognitive bloat. But perhaps the case of Telo\* (and Telo) and his phonebook might instead be understood as illustrating the extension of a cognitive belief-*generating* (rather than belief-preserving) process. The thought would be that in virtue of satisfying (1) – (3), Telo\* and his phonebook constitute a cognitive process which produces (rather than retains) in Telo\* dispositional beliefs about *all* the information in the phonebook. However, such an interpretation is implausible. Ascriptions of dispositional beliefs (as opposed to occurrent beliefs) presume a medium of storage, such as a notebook for Otto and biological memory for the rest of us.

More plausibly, Telo\* and his phonebook might constitute a cognitive process by which Telo\* forms dispositional beliefs only about *endorsed* contents, which are then stored in his biological

---

<sup>28</sup> He may have a disposition to believe all those contents, but that is different from actually having dispositional beliefs about all of them.

memory. In that case, Telo\* and his phonebook resemble TrueTemp who forms such beliefs about the ambient temperature via an enhanced process. In either case, (4) is irrelevant, since (4) is only applicable to would-be memorial processes. Instead, as mentioned in §3, a different, additional endorsement condition would be called for:

- (5) That the existence of the belief-producing or belief-preserving device in relation to one's cognitive architecture is endorsed.

The rationale for (5) is, as explained in §3, to avoid the kind of meta-incoherence—viz., a lack of any appreciation of the source of one's beliefs would have on one's cognitive architecture. The difference between TrueTemp and Telo\* is that only the latter satisfies (5), and so given the importance of (5) in such cases, the extent or degree of cognitive integration is higher or greater in the case of Telo\*. Bear in mind that our cluster approach to cognitive integration is designed to allow for different conditions to apply in different cases. Invoking (4) and (5) (in addition to (1) – (3)) in cases of extended belief-preserving processes, but not (4) in cases of extended belief-generating cases, is by no means ad hoc. Indeed, our cluster approach offers a neat response to the *hybrid* case of Otto\* unwittingly relying on the fitted Memoraidd chip for storage and retrieval of information. In such a case, both (4) and (5) would arguably apply with some considerable weight for the chip to be *fully* cognitively integrated into Otto\*'s memorial processes. The fact that (5) is not satisfied means that the chip is significantly less than fully integrated into his memorial processes, and certainly much less than Otto's notebook is into his memorial processes.

So far we have discussed the problem of cognitive bloat in the context of *metaphysical* cognitive integration. In the remaining part of this paper we assess how the cluster approach handles this problem in an *epistemic* context. We found it implausible by any reckoning that Telo and Telo\* should be regarded as dispositionally believing the entire Verizon phonebook<sup>29</sup>, but that is consistent with this resource playing a constitutive role in cognitive processes of forming dispositional beliefs about its contents in a piecemeal fashion, provided (1) - (3) and (5) are satisfied. As explained, the latter is also true of TrueTemp. There is thus no barrier in principle to cognitive enhancement in that sense<sup>30</sup>. But could such beliefs also amount to knowledge? Let's assume for the sake of argument that all such beliefs are true, undefeated and properly based. As suggested in §3, upgrading extended beliefs to extended knowledge arguably triggers additional conditions on integration; hence our recommendation that a univocal view be replaced with some form of a dualist view. The reason that the conditions for *epistemic* cognitive integration are more demanding than the conditions for *metaphysical* cognitive integration is that, as noted, knowledge is—very plausibly, any way<sup>31</sup>—a kind of *cognitive achievement* associated

---

<sup>29</sup> This is not to say that it would be implausible to attribute to either *dispositions to believe* the content. Dispositional beliefs, as opposed to dispositions to believe, require past endorsement of the contents in question that is not satisfied in these cases.

<sup>30</sup> For a recent discussion of cognitive enhancement in the context of extended cognition, see Carter and Pritchard (2019).

<sup>31</sup> Note that while the thesis that knowledge is a kind of cognitive achievement is one popular way of unpacking the more general idea that knowledge must in some way arise from ability—and this is a view that has been defended by John Greco (2010)—we don't intend to commit for our purposes to the strong reading of this claim that's defended by robust virtue epistemologists (see fn. 12). That said, it is worth noting that even weaker construals of the ability intuition on knowledge which don't require that knowledge be always and everywhere

with value over and above the value of any state falling short of knowledge<sup>32</sup>. Correspondingly, a knower is worthy of credit over and above whatever credit accrues with being in a state of true belief or even justified true belief<sup>33</sup>. These distinctive epistemic features place additional constraints on the integration of knowledge states. Obviously, epistemologists might disagree over which constraints are needed, but here is a contender:

(6) That the reliability of the belief-preserving or belief-producing device is endorsed.

As (6) pertains to processes that involve the production as well as the preservation of beliefs, it applies to all of our previous examples in so far as they involve knowledge.<sup>34</sup> Just as with (4) and (5), the cluster approach permits the application of (6) in some cases (i.e., those involving knowledge) but not in others. This is important because the worry about cognitive bloat is particularly pressing in the case of extended knowledge. That the kind of common-sense functionalism which underpins the extension of Otto's memorial process into his notebook should lead, by parity of reasoning, to an abundance of dispositional beliefs is very surprising if not problematic, but that such functionalism should also lead to an *explosion* of knowledge is unpalatable. As knowledge is a much more demanding status than that of belief, whatever worry one may have about a proliferation of beliefs on the cheap is more critical when those beliefs are promoted to knowledge<sup>35</sup>. Fortunately, our cluster approach provides the resources to avoid any such *epistemic bloat*, as we may call it.

Consider again the case of Telo\* and his phonebook which was stipulated to satisfy (1) – (3). Suppose we add that epistemic integration condition (6) is also satisfied. Even so, Telo\* cannot be said to know *all* the contents of his phonebook. Knowledge requires dispositional belief, and we argued that Telo\* does not dispositionally believe all those contents. The reason is that such beliefs require memory of their contents, and Telo\* clearly fails to satisfy key condition (4). So, the phonebook is insufficiently cognitively integrated for him to dispositionally believe all of its contents. Consequently, this case presents *no explosion of knowledge*. Satisfying (6) would at best mean that Telo\* possessed a cognitive ability to know these contents in a piecemeal fashion, but such enhancement of his epistemic powers will be embraced by our common-sense functionalist. To allow for a cognitive ability involving an integrated external device, whose reliability is endorsed, to yield knowledge of individually endorsed contents one by one, is consistent with denying that such ability can also generate knowledge *en masse* of hitherto unendorsed contents. The point is here that while neither of (1) – (6) are individually necessary

---

an achievement, as Greco maintains, are nonetheless compatible with the more modest suggestion that knowledge oftentimes involves cognitive achievement.

<sup>32</sup> This is, at least, a very widely shared position. See, for example, Pritchard et al (2010) and Carter and Jarvis (2012).

<sup>33</sup> For more details see for example Kvanvig (2003b) and Pritchard (2007).

<sup>34</sup> One may worry that (6) is unduly intellectual; after all, textbook externalists maintain that reflective access to justifiers is unnecessary for a belief to be justified sufficient for knowledge. But, as mentioned in fn. 12, most reliabilists now reject the claim that mere reliability is all that matters for such justification. Instead virtue reliabilists have embraced the ability intuition we discussed in §3, according to which knowledge requires true belief attributable to cognitive ability. In cases of strange and fleeting processes, such as TrueTemp, the belief is true, not through cognitive ability, but because of some disintegrated device. Only if the agent in question endorses the reliability of the device, and thus takes ownership of its workings, would it be sufficiently integrated into *her* cognitive architecture for the beliefs it produces to count as being because of *her* cognitive ability.

<sup>35</sup> One dissenter on this point is Farkas (2016).

for a device to be integrated to any degree into a cognitive architecture, sufficient integration in virtue of one or more of these conditions may be necessary for the device to count as part of a cognitive ability to which dispositional belief and knowledge can be attributed. Our cluster model is designed to apply to (degrees of) cognitive integration, rather than to those states of belief and knowledge to which any such ability may give rise. Whether cluster-model functionalism is true of belief and knowledge is not a question we shall settle here.<sup>36</sup>

Consider instead the case of TrueTemp whom we stipulated to satisfy (1) – (3). Because TrueTemp, as originally described, fails to satisfy (5), the implanted temperature-detecting device is insufficiently integrated into his cognitive architecture for him to have dispositional beliefs about the ambient temperature. Hence, TrueTemp cannot be said to know those temperatures. That is not to say (5) is a necessary condition for the device to be cognitively integrated to any degree, but only that dispositional beliefs cannot (partially) result from the device unless sufficiently cognitively integrated in virtue of satisfying (5). Of course, we can additionally stipulate that TrueTemp also satisfies (5) and (6) such that he *would* have such knowledge in virtue of possessing an integrated, cognitive ability involving such device. In that case, however, we lose grip on the intuition that TrueTemp mysteriously acquires extraordinary knowledge. Instead the case illustrates an enhancement of epistemic powers which our common-sense functionalist would welcome. Either way, our cluster approach to cognitive integration, as initially motivated by cases in philosophy of mind and cognition, accommodates our intuitions about this much-discussed case in epistemology.

One may object that all the work of blocking cognitive and epistemic bloat is done by specific conditions, which any of the integration models we have discussed could, at least in principle, adopt. True, but our cluster model is better suited for this purpose, as the other two models we discuss fail to deal with these problems satisfactorily. Take first the univocal view according to which an external artefact E is metaphysically integrated iff E is epistemically integrated. Now suppose that E is metaphysically integrated into cognitive process P in virtue of meeting conditions C1, C2...Cn, where C2 prevents P&E from leading to cognitive bloat. An example is Telo\* and his phonebook satisfying (1) – (4). According to the univocal view, E is thus also epistemically integrated into cognitive ability A, but it may be that additional condition Cn+1 is needed to prevent A&E from leading to epistemic bloat. We argued that (6) would do the work in the case of Telo\* and his phonebook. So, if Cn+1 is excluded, the univocal view cannot handle epistemic bloat, but if Cn+1 is included, this view must embrace redundant conditions on metaphysical integration. The dualist view isn't afflicted by this specific dilemma, but it faces a related worry. Suppose again that E is metaphysically integrated into P in virtue of satisfying C1, C2...Cn, where C2 prevents any cognitive bloat resulting from P&E. According to the dualist, C1, C2...Cn are necessary and sufficient conditions for metaphysical integration. The problem is that different conditions may be needed in order to metaphysically integrate E into P in different systems or individuals, and so different

---

<sup>36</sup> In fn. 23 we suggested that once the relevant conditions for metaphysical integration had been suitably refined and systematized, they would command, if prompted upon reflection, common-sense assent. Matters are only slightly more complicated when it comes to conditions on epistemic integration, in that appeal would need to be made to *additional* epistemic conditions, such as (6). Everything else applies *mutatis mutandis*. Thus, only after thorough reflection on possible cases would the folk elicit epistemic judgments from which such conditions can be teased out; perhaps improved by empirical input or expert correction as the method of relying on intuitions about cases may well fall short. The key point is that our common-sense functionalism applies equally to epistemic and non-epistemic concepts concerning integration. Thanks to an anonymous referee.

conditions may also be needed to prevent cognitive bloat as a result of P&E in different systems or individuals. But the dualist cannot say that in a different system or individual, E is metaphysically integrated into P in virtue of satisfying, say C2, C3...Cn+1, where C3 is doing the work of preventing cognitive bloat. To use the same example, we argued that (5) is needed for this purpose if Telo\* and his phonebook rather exemplify a belief-generating process; or we could imagine a different individual for whom (5) would be thus needed even for a belief-preserving process.<sup>37</sup> We conclude that only the cluster view has the requisite flexibility to accommodate such variability in (metaphysical or epistemic) cognitive integration.

## §8 Concluding Remarks

Let's take stock. After mapping out debates in epistemology and philosophy of mind (and cognitive science) about cognitive integration and cognitive extension, respectively, we sought in §2 to connect these hitherto disjoint issues. §3 argued against the monist view that a single set of conditions applies in all cases of cognitive integration; rather, the conditions for cognitive integration in the (epistemic) case of knowledge are arguably more demanding than in (metaphysical) cases of states short of knowledge. While such a dualist picture represents progress by recognising some variation between epistemic and metaphysical types of cases, nevertheless §4 found it wanting, and instead we proposed an alternative approach to cognitive integration—*cluster-model functionalism*. Rather than providing necessary and sufficient conditions for cognitive integration, the common-sense functionalist model has it that different conditions apply in different cases within the two domains to different degrees or extents. After developing this model in some detail in §5, and then applying it to cognitive integration in §6, we demonstrated in §7 how it made intuitively correct predictions about a variety of metaphysical and epistemic cases. Being able to handle such diverse cases lends support to our functionalist approach; or so we contend.

Crucially, the ensuing picture of cognitive integration paves the way for a response to the challenge posed by cognitive and epistemic bloat; or so we argued in §7. The latter are putative cases where an individual is truly ascribed an incredible number of states of (dispositional) beliefs and knowledge on the basis of some form of cognitive extension. These cases were found to be problematic in that the individual has not even entertained the contents in question, and so the question of whether to assent to those contents has never arisen. Nothing in the individual's linguistic behaviour would thus indicate belief, as necessary for knowledge. In response, our functionalist model offers an explanation of why the internal gadget or external resource in question is insufficiently integrated into the relevant cognitive process to enable such untoward cognitive extension. In particular, knowledge results from exercises of (reliable) cognitive abilities, but, on the functionalist model, nothing about the cases we considered suggested that the individuals possessed extended cognitive abilities whose exercise lead to an explosion of knowledge. What would be possible on this model, provided the relevant gadget or resource meets the required integration conditions, is a piecemeal expansion of knowledge (and hence belief) by such enhanced epistemic powers, but any functionalist of

---

<sup>37</sup> Of course, the dualist could relativize to individuals, systems, etc., but then the worry could be rephrased in terms of assigning different weights to the very same conditions. Again, the dualist may avail herself even of such gradability, but increasingly her view begins to look a lot like the cluster view.

the stripe we consider should embrace such relatively more palatable consequence of their view.<sup>38</sup>

---

<sup>38</sup> Thanks to Gloria Andrada, Emma C. Gordon, Mark Sprevak, Mona Simion, and three anonymous referees at *Noûs*. The authors are also grateful to collaborators on the AHRC *Extended Knowledge* project for helpful discussion, including Andy Clark, Orestis Palermos, and Duncan Pritchard.

## REFERENCES

- Adams, Fred, and Ken Aizawa. 2001. 'The Bounds of Cognition'. *Philosophical Psychology* 14 (1): 43–64.
- . 2008. *The Bounds of Cognition*. John Wiley & Sons.
- . 2010. 'The Value of Cognitivism in Thinking about Extended Cognition'. *Phenomenology and the Cognitive Sciences* 9 (4): 579–603.
- Allen-Hermanson, Sean. 2012. 'Superdupersizing the Mind: Extended Cognition and the Persistence of Cognitive Bloat'. *Philosophical Studies* 164 (3): 791–806. doi:10.1007/s11098-012-9914-7.
- Bostrom, Nick. 2009. 'Cognitive Enhancement: Methods, Ethics, Regulatory Challenges'. *Science and Engineering Ethics* 15 (3): 311–41.
- Bostrom, Nick, and Anders Sandberg. 2009. 'The Wisdom of Nature: An Evolutionary Heuristic for Human Enhancement'. In *Human Enhancement*, edited by Julian Savulescu and Nick Bostrom, 375–416. Oup Oxford.
- Burge, T. 1993. 'Content Preservation'. *Philosophical Review* 102 (4): 457–88.
- Carter, J. Adam. 2013. 'Extended Cognition and Epistemic Luck'. *Synthese* 190 (18): 4201–4214.
- . 2019. 'Epistemic Luck and the Extended Mind'. In *The Routledge Handbook of the Philosophy and Psychology of Luck*. Edited by Church, I.M. and Hartman, R. J. Routledge.
- Carter, J. Adam and Ben Jarvis. 2012. 'Against Swamping'. *Analysis*, 72(4), 690-699.
- Carter, J. Adam and Jesper Kallestrup. 'Extended Cognition and Propositional Memory'. *Philosophy and Phenomenological Research* 92 (3): 691-714.
- . 2018. 'Extended Circularity: A New Puzzle for Extended Cognition'. In *Extended Epistemology*, edited by J. Adam Carter, Andy Clark, Jesper Kallestrup, Orestis Palermos and Duncan Pritchard, Oxford: Oxford University Press. pp. 42-63.
- Carter, J. Adam and S. Orestis Palermos. 2015. 'Active Externalism and Epistemic Internalism'. *Erkenntnis* 80: 753-772.
- Carter, J. Adam, Jesper Kallestrup, S. Orestis Palermos and Duncan Pritchard. 2014. 'Varieties of Externalism'. *Philosophical Issues* 24 (1):63-109.
- Carter, J. Adam, Jesper Kallestrup, S. Orestis Palermos and Andy Clark (eds.) 2018. *Extended Epistemology* Oxford: Oxford University Press.
- Carter, J. Adam and Duncan Pritchard. 2019. 'The Epistemology of Cognitive Enhancement', *Journal of Medicine and Philosophy*, forthcoming.
- Clark, Andy. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.
- . 2007. 'Curing Cognitive Hiccups: A Defense of the Extended Mind'. *Journal of Philosophy* 104 (4): 163-92.
- Clark, Andy, and David Chalmers. 1998. 'The Extended Mind'. *Analysis* 58 (1): 7–19.
- Dror, Itiel E., and Stevan Harnad. 2008. 'Offloading Cognition onto Cognitive Technology'. *Cognition Distributed: How Cognitive Technology Extends Our Minds* 16: 1.
- Farkas, Katalin. 2016. 'Know-Wh Does not Reduce to Know That'. *American Philosophical Quarterly* 53 (2):109-122.
- Gaut, Berys. 2000. "'Art" as a Cluster Concept'. In *Theories of Art Today*, edited by Noel Carroll, 25-44, Madison: University of Wisconsin Press.
- Graham, Peter J. 2012. 'Epistemic Entitlement'. *Noûs* 46 (3): 449–82. doi:10.1111/j.1468-0068.2010.00815.x.

- Greco, John. 2003. 'Knowledge as Credit for True Belief'. In *Intellectual Virtue: Perspectives From Ethics and Epistemology*, edited by Michael DePaul and Linda Zagzebski, 111–34. Oxford: Oxford University Press.
- . 2010. *Achieving Knowledge*. Cambridge: Cambridge University Press.
- . 2012. 'A (Different) Virtue Epistemology'. *Philosophy and Phenomenological Research* 85: 1–26.
- Heersmink, Richard. 2016. 'The Cognitive Integration of Scientific Instruments: Information, Situated Cognition, and Scientific Practice'. *Phenomenology and the Cognitive Sciences* 15 (4):1-21.
- Jackson, Frank (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Oxford University Press.
- . 2000. 'Representation, Scepticism, and the A Priori'. In *New Essays on the A Priori*, 320-32, edited by Paul Boghossian and Christopher Peacocke, Oxford: Oxford University Press.
- Jarvis, Benjamin. 2015. 'Epistemology and Radically Extended Cognition'. *Episteme* 12 (04): 459–78.
- Kelp, Christoph. 2013a. 'Extended Cognition and Robust Virtue Epistemology'. *Erkenntnis* 78 (2): 245–52.
- . 2013b. 'Extended Cognition and Robust Virtue Epistemology: Response to Vaesen'. *Erkenntnis*, 1–4.
- . 2017. How To Be A Reliabilist. *Philosophy and Phenomenological Research*. doi:10.1111/phpr.12438
- Kallestrup, Jesper and Duncan Pritchard. 2012. 'Robust Virtue Epistemology and Epistemic Anti-Individualism', *Pacific Philosophical Quarterly* (93): 84-103.
- . 2013. 'Robust Virtue Epistemology and Epistemic Dependence'. In *Knowledge, Virtue and Action: Putting Epistemic Virtues to Work*, edited by Tim Henning and David Schweikard, (Routledge, pp. 209-26.)
- . 2014. 'Virtue Epistemology and Epistemic Twin Earth', *European Journal of Philosophy* (22): 335-57.
- Kallestrup, Jesper and Mark Sprevak. 2014. 'Entangling Externalisms'. In *New Waves in Philosophy of Mind*, London and New York: Palgrave, pp. 77-97.
- Kvanvig, Jonathan. 2003a. 'Simple Reliabilism and Agent Reliabilism'. *Philosophy and Phenomenological Research* 66 (2): 451–56. doi:10.1111/j.1933-1592.2003.tb00273.x.
- . 2003b. *The Value of Knowledge and the Pursuit of Understanding*, Cambridge: Cambridge University Press
- Lackey, Jennifer. 2007. 'Why We Don't Deserve Credit for Everything We Know'. *Synthese* 158 (3): 345–61.
- Lehrer, Keith. 1990. *Theory of Knowledge*. Routledge.
- Lewis, David, 1970. 'How to Define Theoretical Terms', *Journal of Philosophy* 67: 427-446.
- . 1972. 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50 (3): 249-258.
- Longworth, Francis and Scarantino, Andrea. 2010. 'The disjunctive theory of art: The cluster account reformulated', *British Journal of Aesthetics* 50 (2): 151-67.
- Machery, Edouard (2009). *Doing Without Concepts*, Oxford: Oxford University Press.
- Menary, Richard. 2006. 'Attacking the Bounds of Cognition'. *Philosophical Psychology* 19 (3): 329–44.

- Meskin, Aaron. 2007. 'The Cluster Account of Art Reconsidered', *British Journal of Aesthetics* 47, 388–400.
- Palermos, Orestis. 2011. 'Belief-Forming Processes, Extended'. *Review of Philosophy and Psychology* 2 (4): 741–65.
- . 2014. 'Knowledge and Cognitive Integration'. *Synthese* 191 (8): 1931–51.
- Plantinga, Alvin. 1993. *Warrant and Proper Function*. New York: Oxford University Press.
- Pritchard, Duncan. 2007. 'Recent Work on Epistemic Value', *American Philosophical Quarterly* 44: 85–110.
- . 2010. 'Cognitive Ability and the Extended Cognition Thesis'. *Synthese* 175 (1): 133–51.
- Ramsey, Frank. P. 1929. 'Theories'. In *Foundations. Essays Philosophy, Logic, Mathematics and Economics*, H. D. Mellor (ed.), London: Routledge & Kegan Paul, 1978, 101-125.
- . 2012. 'Anti-Luck Virtue Epistemology'. *Journal of Philosophy* 109: 247–79.
- Rowlands, Mark. 2009. 'Extended Cognition and the Mark of the Cognitive'. *Philosophical Psychology* 22 (1): 1–19.
- Rupert, Robert D. 2004. 'Challenges to the Hypothesis of Extended Cognition', *The Journal of Philosophy*, 101(8): 389-428.
- Sosa, Ernest. 1983. 'Nature Unmirrored, Epistemology Naturalized', *Synthese* 55 (1): 49-72.
- . 2009. *A Virtue Epistemology: Apt Belief and Reflective Knowledge (Vol. 1)*. Oxford: Oxford University Press.
- . 2011. *Reflective Knowledge: Apt Belief and Reflective Knowledge (Vol. 2)*. Oxford: Oxford University Press.
- Sterelny, Kim. 2010. 'Minds: extended or scaffolded?' *Phenomenology and the Cognitive Sciences* 9(4): 465-481.
- Vaesen, Krist. 2011. 'Knowledge without Credit, Exhibit 4: Extended Cognition'. *Synthese* 181 (3): 515–29.
- Zagzebski, Linda Trinkaus. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge University Press.