For a special issue of *Philosophical Issues* on 'Extended Knowledge'

VARIETIES OF EXTERNALISM

J. Adam Carter, Jesper Kallestrup, S. Orestis Palermos & Duncan Pritchard

University of Edinburgh

ABSTRACT. Our aim is to provide a topography of the relevant philosophical terrain with regard to the possible ways in which knowledge can be conceived of as extended. We begin by charting the different types of internalist and externalist proposals within epistemology, and we critically examine the different formulations of the epistemic internalism/externalism debate they lead to. Next, we turn to the internalism/externalism distinction within philosophy of mind and cognitive science. In light of the above dividing lines, we then examine first the extent to which content externalism is compatible with epistemic externalism; second, whether active externalism entails epistemic externalism; and third whether there are varieties of epistemic externalism that are better suited to accommodate active externalism. Finally, we examine whether the combination of epistemic and cognitive externalism is necessary for epistemology and we comment on the potential ramifications of this move for social epistemology and philosophy of science.

0. INTRODUCTION

The distinction between internalism and externalism is common to both contemporary philosophy of mind and contemporary epistemology, and it is the central topic of ongoing debates in both disciplines. To a certain extent this is unsurprising, given that these two theoretical domains are closely related in that they both focus on the study of cognitive phenomena.

Despite the appearance of such common grounds, however, there are several differences between these two fields of study. Arguably, the most important one is that philosophy of mind studies cognition in a more inclusive way, by considering all kinds of mental states and processes (e.g., experience, beliefs, desires, emotions and so on), whereas epistemology focuses on the cognitive process of knowledge-acquisition and its cognates (e.g., belief and justification). Accordingly, it would be unsafe to assume that the internalism/externalism distinction maps on to both disciplines in the exact same way.

And, indeed, it doesn't. Briefly stated, on the one hand, the internalism/externalism distinction, as it is normally construed within contemporary epistemology, refers to the debate over whether an agent's justification for believing a proposition *p* should always be (at least in principle) accessible to him by reflection alone. In other words, epistemic internalists hold that one's justification should always be internal to one's conscious mind, whereas epistemic externalists deny this claim.[1] On the other hand, the standard way of thinking about the internalism/externalism debate within philosophy of mind is that internalists hold that cognitive processes and mental states reside exclusively within the agent's head, whereas externalists deny this on multiple grounds and with several degrees of departure from the internalist position.

Embodied cognition theorists, for example, hold that several mental processes and states (e.g., experience, emotions and desires) may be constitutively dependent not only on the agent's brain but his body as well, and proponents of content externalism hold that the content of various mental states such as beliefs and desires may at least in part constitutively depend on features of one's physical or social environment. Even more provocatively still, active externalism (be it in the form of the extended mind thesis or the extended and distributed cognition hypotheses) holds that mental states and cognitive processes extend beyond the agent's biological organism to the artifacts or even to other agents that he or she mutually interacts with. Overall, then (and with a great deal of simplification involved), we may say that whereas the difference between epistemic internalists and externalists concerns the accessibility of one's justification through one's (conscious) psychology, the distinction between cognitive internalism and externalism is a debate over the spatial location of the constituents of certain mental states and processes.

However, and despite the fact that the application of the internalism/externalism terminology is not identical across epistemology and philosophy of mind, it should be obvious that it is not entirely unrelated. After all, asking where one's (conscious) psychology might be located is a perfectly intelligible question to ask, and so is the question about whether one could or should be able to be consciously aware of mental states and processes that may not be entirely located within one's head (or even one's biological organism), but which nevertheless seem to justify one's belief in some proposition *p*. Accordingly, our aim in this paper is to explore the possible connections between epistemology and philosophy of mind on the basis of the internalism/externalism debate as it applies to both of them. Specifically, we aim at studying whether epistemic externalism entails cognitive externalism and *vice versa*. Moreover, given the possibility of extended and distributed cognition within philosophy of mind, we are interested in exploring the possible ways in which knowledge can be conceived of as extended or distributed,

and whether there are any forms of epistemic externalism that may be better suited to accommodate the different varieties of cognitive externalism.

Accordingly, in §1, we begin by charting the different types of internalist and externalist proposals within epistemology, and we critically examine the different formulations of the epistemic internalism/externalism debate they lead to. In §2, we present the internalist/externalist debate in philosophy of mind and, then, in light of the above dividing lines, we perform a comparative analysis of the two disciplines in §3. In particular, in §3.1 we focus on the compatibility of content externalism with epistemic internalism; in §3.2 we examine the extent to which active externalism entails a commitment to epistemic externalism; and in §3.3 we explore whether there are varieties of epistemic externalism that are better suited to accommodate any of the available forms of active externalism. Finally, in §4, we investigate whether active externalism is necessary for certain forms of epistemic externalism, and we comment on the potential ramifications of the combination of epistemic and cognitive externalism for social epistemology and the related field of philosophy of science.

1. THE EPISTEMIC INTERNALISM/EXTERNALISM DEBATE

The debate between epistemic internalists and externalists is about whether what confers justification on a belief is necessarily internal to the agent. All epistemic internalists agree that justification consists in reasons or evidence that are somehow internal to the agent's cognitive perspective, and upon which she bases her belief, so that she has a justified belief, but they disagree over how to understand the notion of being internal.

Accessibilism holds that justification is reflectively accessible: whenever an agent holds a justified belief, she is in a position to know just by reflection alone that which justifies her belief. Weakly understood, that requires reflective knowledge of the presence of those factors which justify her belief; more strongly, she must also have reflective knowledge that those factors justify her belief.[2]

In contrast, *mentalism* holds that only mental states can justify beliefs in that only mental states are internal to the agent's mind.[3] Thus, Richard Feldman (2005) holds that justification consists in good reasons, and reasons are mental states, understood broadly to include occurrent and dispositional mental states and events. If two agents are in the same mental states, they are necessarily justificational duplicates. Feldman takes this strong supervenience claim to be supported by a range of cases in which two agents would seem to differ justificationally when

they are in relevantly different mental states. In the absence of any additional commitments, accessibilism and mentalism are logically independent theses. One could hold that the justification-conferring factors must be reflectively accessible, or at least subject to conscious awareness, and yet deny that those factors are mental states. And one could hold that while all such factors are mental states, the agent need not have reflective access to, or indeed even be consciously aware of, those factors.

Formulated in terms of knowledge, accessibilists accept that internal justification—justification that is reflectively recognizable—is necessary for knowledge. So, they hold that whenever an agent knows that *p*, she is in a position to know by reflection, not necessarily that she knows that *p*, but rather that on the basis of which she knows that *p*. Mentalists also accept that internal justification is necessary for knowledge. So, they hold that whenever an agent knows that *p*, she is in mental states which constitute the justifying basis for her knowledge.

Epistemic externalists, on the other hand, deny that justification is always reflectively accessible. It's possible that an agent holds a justified belief without being in a position to know just by reflection the factors that make her belief justified. Externalists think that what makes a belief justified may be external to an agent's cognitive perspective: it consists in an objective relationship between the agent's cognitive faculties and external reality, or in those faculties instantiating certain external properties, which she need know nothing about. For instance, process reliabilists claim that what matters for justification is that the causal process via which the belief was produced is in fact reliable, regardless of whether the agent has any reason or evidence to believe that it is. If the process is reliable, then it is objectively likely that the belief, as produced by that process, is true.[4]

Formulated in terms of knowledge, epistemic externalists claim that knowledge is possible without internal justification. Thus, process reliabilists say that a true belief's origin in a reliable process is sufficient for knowledge, absent certain types of undermining defeat. In short, knowledge can be grounded in what is external to the conscious mind.

Pretty much all epistemic externalists stress the importance of evaluating justified beliefs objectively in terms of probability or truth-conduciveness: what matters is whether the agent is likely to hold a true belief. Epistemic internalists, in contrast, typically emphasize the importance of evaluating justified beliefs subjectively in terms of epistemic responsibility: what matters is whether the agent holds intellectually blameless beliefs, or beliefs that are rational from her perspective.[5]

However, not all internalists are wedded to this deontological conception of justification, according to which an agent has a justified belief when she deserves praise for having the belief,

or when it is her epistemic duty or obligation to form that belief.[6] The disagreement over which conception of epistemic evaluation takes priority is reflected by different judgements about at least two central types of cases. Thus, epistemic internalists maintain that clairvoyance and so-called 'New Evil Demon' cases demonstrate that the reliability of a belief-producing process is neither sufficient nor necessary, respectively, for a belief to count as justified. In the first type of case, an agent forms beliefs as a result of suspect but reliable cognitive faculties, and the internalist concludes that the subject lacks justification for her beliefs even despite the reliability in play. In the second type of case, we are asked to imagine two agents forming identical beliefs in subjectively indistinguishable conditions, but where only in the one case are the beliefs reliably formed (since in the second case the subject is massively deceived). The internalist argues that both subjects enjoy an identical level of justification for their beliefs, even despite the different degrees of reliability in play.[7,8]

Given that traditional Gettier cases were counterexamples to definitions of knowledge as involving an internalist conception of justification as reasons the agent can produce if asked, reliabilists were hoping to solve the Gettier problem by abandoning that conception. What stands in the way of knowledge-undermining luck is rather the reliability of the cognitive process that produced the belief. Still, cases of so-called *environmental epistemic luck* pose a problem for traditional versions of reliabilism. This is not the standard intervening epistemic luck that one finds in the usual Gettier-style cases—such that something causally intervenes between the agent's formation of their justified belief and their cognitive success (i.e., their belief being true)—but rather epistemic luck which simply concerns the environment in which the agent exhibits this cognitive success.[9]

For instance, it looks as if Henry's belief that there's barn, as based on a veridical visual experience of a barn in fake barn county, is the result of a reliable cognitive process, namely visual perception. And yet this belief, so formed, is nonetheless only luckily true, given the nature of Henry's environment.[10] In response, one might insist that the manifestation of a reliable process is somehow environment-relative. Alternatively, reliability can be formulated modally in at least two distinct ways.

Counterfactual reliabilism is the view that an agent knows that *p* only if: were *p* false the agent would not believe *p* as a result of process *r*.[11] *Sensitivity* is a modal condition on knowledge which explains Henry's lack of knowledge: if there were not a barn but instead a barn façade in front of Henry, then he would still believe as a result of visual perception that there's a barn. Alternatively, we can understand the reliability of a process in terms of delivering (mostly) true beliefs in a band of worlds close to the actual world. Thus neighbourhood reliabilism is the view

that in such worlds, all (or nearly all) beliefs formed by process *r* are such that believing that *p* implies *p*.[12] Or we can use a counterfactual to define such a *safety* condition on knowledge: an agent knows that *p* only if had she believed that *p* as a result of *r* in different circumstances, *p* would be true in those circumstances.[13]

Reflect that this other brand of reliabilism can handle the fake barn case equally well: had Henry believed that there's a barn as a result of visual perception in nearby circumstances in which he was visually presented with a fake barn, then his belief would be false. Note that because subjunctive conditionals do not validly contrapose, safety and sensitivity are not logically equivalent. While safety requires merely that *S* track the truth in a range of close *p*-worlds, sensitivity requires that *S* track the truth out to the closest not-*p*-worlds. Still, their common feature is that both require not just actual true belief, but also counterfactual co-variation of belief and truth.

The two views also yield familiarly different results when it comes to sceptical arguments. Your belief that you have hands is both safe and sensitive. Your belief that you are not a brain-in-a-vat is arguably safe but not sensitive.[14] In response to the sceptical argument that you cannot know that you have hands, because you cannot know that you are not a brain-in-a-vat, sensitivity theorists thus reject the underlying closure principle, while safety theorists upheld the neo-Moorean stance that you can know the negation of such sceptical hypotheses.[15]

The literature contains a number of putative counterexamples to both safety and sensitivity as necessary or sufficient conditions on knowledge. Of particular interest are cases where the agent's belief satisfies the model condition in question but with the wrong direction of fit. Imagine an agent who forms beliefs on the basis of some process that would be otherwise unreliable, and yet ends up with safe or sensitive beliefs due to the intervention of some benevolent demon. The master intuition is here that when an agent knows, her belief is true not because of *any* reliable process, but because of the exercise of her cognitive abilities. Knowledge is a cognitive achievement worthy of praise, but if the agent gets things right because of someone else changing the world so as to systematically match her beliefs, then she deserves no credit for hitting the truth.

Consider Ernest Sosa's (2007; 2009) triple-A version of virtue epistemology according to which knowledge is apt belief, where a belief is apt just in case it is accurate (true) because adroit (out of cognitive ability). The because-relation is key to understand how knowledge can be a cognitive achievement on the part of the agent. Suppose an expert archer dispatches an arrow, which is first blown off course by a sudden and unexpected gust, and then diverted back on track

again by a guardian angel to ensure it hits the target. The performance is accurate and adroit but inapt in that the accuracy is not because of the agent's adroitness.[16]

Nevertheless, a lingering worry about environmental luck afflicts virtue epistemology. After all, it seems that Henry not only possesses the right cognitive ability to form beliefs on the basis of visual perception, he also manifests that ability when forming the belief that there's a barn. Virtue epistemologists have offered various responses to cases of this kind. Sosa (2009) concedes that Henry has so-called animal knowledge in the fake barn case, but insists that he lacks second-order reflective knowledge; John Turri (2011) attempts to incorporate a safety condition into the virtue-theoretic framework: knowledge is ample belief, where a belief is ample just in case the safety of the belief is because of the agent's cognitive ability; while a further option is to argue that the manifestation of cognitive ability depends on the appropriateness of one's 'regional' environment, which, in Henry's case, includes the real as well as the fake barns (Palermos *forthcominga*).[17]

Against the backdrop of the foregoing, one might reasonably expect that any attempt to decompose knowledge into belief, truth and some other modal or virtue-theoretic condition is doomed to fail. Timothy Williamson (2000) recommends a different tack. According to his epistemology-first approach, knowing that *p* is not just a mental state, it is the most general factive attitude to a proposition *p* that the agent has if she has any factive attitudes at all. Seeing that *p*, knowing that *p*, and other factive attitudes are such that an agent can bear those mental relations to *p* only if *p* is true. Factive mental states are thus wide (or broad) in virtue of the attitude rather than the content *p* being wide. Knowledge is also what Williamson calls a *prime* mental condition in the sense of not being a conjunction of independent narrow and environmental conditions, say belief, truth and something else. Indeed knowledge causally explains behaviour in ways that cannot be explained by any putative component state that falls short of knowledge such as true belief or justified belief. Given that knowledge plays such an irreducible explanatory role, knowledge should be posited as a *sui generis* mental state. Note also that Williamson (2000) and others have also put knowledge first in accounts of various epistemic norms—norms of assertion, belief, practical reasoning, and so on.[18]

2. THE COGNITIVE INTERNALISM/EXTERNALISM DEBATE

Conceiving of knowledge—typically the primary focus of epistemology—as a cognitive (i.e., mental) phenomenon, epistemological considerations have traditionally focused on the internal features of the individual cognitive agent; cognition, after all—it is largely held—rests within the individual's head.

This last claim, however, has lately been called into question by recent advances within philosophy of mind and cognitive science, and especially the currents of embodied cognition, content externalism and active externalism. To get a grip on what all these theories of mind amount to and how they differ from each other it should be helpful to consider their motivations in a chronological order of appearance. Before proceeding further, however, we should note that their common denominator, and hence the reason why we here group them together, is that they all deny the claim that cognition resides entirely within the individual's head.

Now, the first blow to the approach of internalism—the idea that a complete understanding of our minds can be achieved by an exclusive focus on our brains—came from content (or passive) externalism, which shows that some mental contents fail to supervene on intrinsic facts (i.e., facts that pertain solely to our brains); consequently, the opposite of internalism—*viz.,* externalism about our minds—must be true. Consider the following thought experiment.

Imagine a remote planet, Twin Earth, which is exactly like Earth, except that instead of water (H_2O) it has a different substance, twin-water. Even though twin-water is a different chemical compound, say XYZ, its macro properties are just like those of water: it looks and tastes like water, it can be found in the rivers and oceans on Twin Earth, and so on. Furthermore, imagine two intrinsically identical individuals: *S* who lives on Earth and twin-*S* who lives on Twin Earth, neither of whom knows anything about chemistry. Now, when *S* utters "water quenches thirst" she is expressing her belief that water quenches thirst, a belief that is true if and only if H_2O quenches thirst. To the contrary, having always encountered twin-water and never having encountered or heard of water, when twin-*S* utters "water quenches thirst" our intuition dictates that she does not believe that water quenches thirst. Instead, twin-*S* expresses the belief that twin-water quenches thirst, a belief with different truth-conditions. Accordingly, we seem to have two intrinsically identical individuals, who nevertheless have different beliefs, meaning that some beliefs do not supervene on intrinsic facts. To the contrary, according to this form of externalism, in order to have certain types of intentional mental states, such as beliefs and desires, it is necessary to be related to the environment in the right way.

Now, whereas Hilary Putnam (1975) initially focused on the linguistic content of sentences containing natural kind terms being individuated externally in terms of features of the physical environment, Tyler Burge (1979; 1986) and others were quick to extend Putnam's conclusion in at least three respects. First, if the content of beliefs is fixed by the content of the sentences that the believer uses to express her beliefs, then the content of such mental states will be individuated externally if the corresponding linguistic content is individuated externally. Indeed, if mental states are individuated by their contents, then those states themselves are also individuated externally. Secondly, just as the linguistic content of sentences containing natural kind terms is individuated externally, so is the linguistic content of sentences containing a variety of non-natural kind terms such as 'sofa', 'brisket' and 'red'. Thirdly, it isn't just by varying features of the physical environment while keeping all the intrinsic facts fixed that we can establish the existence of externally individuated content. We can run roughly the same argument by instead varying features of the socio-linguistic environment, hence social externalism. Despite the subtle differences, however, what all these types of content externalism have in common is that they all lead to the conclusion that studying our brains in isolation from their external environments would be insufficient for a complete understanding of our minds.

Not much later, however, several cognitive scientists and philosophers of mind (Varela, Thomson & Rosch 1991; Clark 1997) as well as roboticists (Brooks 1991*a*; Brooks 1991*b*) noted that not even the study of our brains as embedded in their environments is enough. This is because cognition is not just embedded but also embodied in the sense that aspects of the agent's body beyond the brain play a physically constitutive role in cognitive processing (i.e., literally speaking, with respect to several cognitive operations, our bodies are parts of our minds). In particular, for those aspects of an agent's mind where her brain and body are heavily *interdependent*, we should think of the latter as a constitutive element of the agent's overall cognitive system.[19] According to embodied cognition, then, considerations pertaining to the agent's body as well as its interaction with her brain (and central nervous system) are essential for a complete understanding of the human mind.

Now, active externalism, as represented by the extended mind thesis and the extended and distributed cognition hypotheses, is the extreme consequent of the approach of embodied cognition. We should note, however, that we here say 'extreme', only because of its radical conclusions. Indeed, for some it may be counterintuitive to accept that mental states and cognitive systems extend beyond our organisms to the artifacts we mutually interact with or that cognitive processing may be distributed amongst several individuals and their artifacts. The spirit of the approach, however, is very similar to, if not the same as, that of embodied cognition. If we

are willing to accept that our minds are embodied when our brains and bodies heavily depend on each other, there is no principled reason to deny that cognitive processes and states are extended or even distributed in those situations where our brains, the artifacts we employ or the other agents we interact with, are heavily interdependent.

In fact, active externalism in all of its forms has been developed, refined, and defended by many philosophers (Clark & Chalmers 1998; Clark 2007; 2008; Hutchins 1995; Menary 2006; 2007; Theiner 2011; Wheeler 2005, Wilson 2000; 2004).[20] Accordingly, active externalism, in all of its forms, is a viable hypothesis that we believe can generate several ramifications within analytic epistemology. Before concluding this section, however, it should be helpful to highlight some of the different formulations of active externalism and the argumentative techniques that motivate them.[21]

Specifically, active externalism has appeared in the literature under three main labels—*viz.,* the extended mind thesis, the hypothesis of extended cognition and the hypothesis of distributed cognition. Admittedly, there are several possible points of connection between these three formulations, but the means for arriving at them as well as the claims they put forward are sufficiently different to deserve special attention.

Focusing on *cognitive processing*, the hypothesis of extended cognition is the claim that "the actual local operations that realize certain forms of human cognizing include inextricable tangles of feedback, feedforward and feed-around loops: loops that promiscuously criss-cross the boundaries of brain, body and world" (Clark 2007, §2); cognitive processing can and (under the appropriate conditions) literally extends to the agent's surrounding environment. Think about solving a mathematical problem by using pen and paper, or perceiving a chair through a tactile visual substitution system. According to the hypothesis of extended cognition, the involved artifacts are proper parts of the ongoing cognitive processing.

However provocative this claim may sound, the extended mind thesis is usually thought to be more challenging still. Instead of concentrating on cognitive processes, the claim, in this case, is that it is *mental states*—experience, beliefs, desires, emotions, and so on—that get extended. The typical argument (Clark & Chalmers 1998) involves Otto—an Alzheimer's patient—whose dispositional beliefs are taken to be partly *constituted* by his well-organized notebook; his mind, therefore, extends to his notebook.

Finally the third formulation of active externalism—*viz.,* the hypothesis of distributed cognition (Hutchins 1995, Theiner *et al* 2010, Sutton *et al* 2008, Wilson 2005, Heylighen *et al* 2007)—is the most radical of them all. According to this form of externalism, *cognitive processing* may not just be extended beyond the agent's head or organism but even distributed amongst

several individuals along with their epistemic artifacts. Despite its radical conclusion, however, the hypothesis of distributed cognition differs from the hypothesis of extended cognition only in that this time cognitive processes and the resultant cognitive systems extend to include not only artifacts but other individuals as well.

Now to see what the motivational difference is between, on one hand, the hypotheses of extended and distributed cognition, and, on the other hand, the extended mind thesis, notice that whereas the first two hypotheses are concerned with extended cognitive processes, the latter is usually formulated on the basis of extended mental states. This makes it somewhat more provocative, because the existence of extended mental states—such as extended dispositional beliefs—is a claim that is more counterintuitive and thereby less easy to argue for than the claim that there are extended cognitive processes.

Despite their difference in focus (i.e., on processes vs. states), however, all hypotheses have been traditionally motivated on the basis of functionalism, and especially, the sort of common-sense functionalism that is captured by the following principle:

Parity Principle:

> If as we confront some task, part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process (Clark and Clalmers 1998, 8).

To see how we can put this into practice with respect to the extended mind thesis, consider the following thought experiment. First, think about a normal case of a belief stored in biological memory. Inga learns about an interesting exhibition in MOMA. She thinks, recalls that the museum is on 53rd street and starts walking to the museum. Now consider Otto who suffers from Alzheimer's disease; as a consequence, Otto has to rely on information in the environment to help structure his life and so carries a thick, well-organized notebook everywhere he goes. When he learns new information he writes it down, when he needs some old information he looks it up. Otto hears about the same exhibition and decides to go see it. He opens the notebook, finds the address of the museum and starts heading towards 53rd street.

Andy Clark and David Chalmers (1998) claim that Otto walked to 53rd street because he wanted to go to MOMA and believed that MOMA was on 53rd street. What is more, if one wants to say that Inga had her belief before she consulted her memory, then one could also claim that Otto believed that the museum was on 53rd street even before looking up the address in his notebook. This is because the two cases are functionally on a par; given our everyday, common-sense understanding of how memory works, we can make the following claim: "the notebook

plays for Otto the same role that memory plays for Inga; the information in the notebook functions just like the information [stored in Inga's biological memory] constituting an ordinary non-occurrent belief; it just happens that this information lies beyond the skin." (Clark & Chalmers 1998, 13)

Moreover, in order to make this common-sense functionalist claim more plausible, Clark (2010) notes that, just as in the case of biological memory, the availability and portability of the resource of information should be crucial. Accordingly, he offers the following set of criteria to be met by non-biological candidates for inclusion into an individual's mind:

> 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." (Clark 2010, 46)[22]

Philosophers of cognitive science, however, have noted that, contrary to the extended mind thesis, in order to motivate the extended and distributed cognition hypotheses we do not need to rely on the above common-sense functionalism and the three criteria it generates.[23] The reason is that these two hypotheses do not rely for their support on extended mental states, but on extended *dynamical* cognitive processes and the extended cognitive systems these processes give rise to.

This shift in focus allows the employment of the conceptual framework of dynamical systems theory—the most powerful mathematical framework for studying the behavior of dynamical systems in general—and has turned out to be particularly useful: According to dynamical systems theory, in order to claim that two (or more) systems give rise to an overall extended or distributed (or coupled) system, what is required is the existence of non-linear relations that arise out of *mutual interactions* between the contributing parts (Palermos 2014, Chemero 2009, Froese *et al* 2013, Sutton *et al* 2008, Theiner *et al* 2010, Wegner et al. 1985, Tollefsen & Dale 2011). In other words, on the basis of dynamical systems theory, we can claim that in order to have an extended or even distributed cognitive system, all we need is that the contributing members (i.e., the relevant cognitive agents and their artifacts) interact mutually with each other. Accordingly, the extended and distributed cognition hypotheses can be motivated in ways that bypass common-sense functionalism.[24]

To close this section then, content externalism, the current of embodied cognition, the extended and distributed cognition hypotheses, and the extended mind thesis are all different ways that invoke different arguments for claiming that cognition does not reside entirely within

the agent's head. Barring the current of embodied cognition—which allows the agent's metal life to escape only the boundaries of the skull but not those of the body—for the remainder of the paper, we will concentrate on the rest, more outward looking forms of externalism and their relation to the debate over epistemic internalism and externalism.

3. EXTERNALISM IN PHILOSOPHY OF MIND AND EPISTEMOLOGY

In this section, our aim will be to assess the implications of both *passive* (i.e., content) externalism and *active* externalism as they bear on the internalism/externalism divide in epistemology. The section has three parts. In §3.1, we assess whether passive (content) externalism entails epistemic externalism; our method in doing so will be to outline and evaluate three prominent arguments for the incompatibility of content externalism and epistemic internalism: Laurence Bonjour's *accessibility argument*, Paul Boghossian's *Self-Knowledge Argument* and James Chase's *Process Argument*. In §3.2, we take a similar approach: we assess whether (and how) active externalism entails epistemic externalism by evaluating arguments for the incompatibility of (different varieties of) active externalism and epistemic internalism. In §3.3, we consider active externalism's fit with several strands of epistemic externalism, including safety-based accounts as well as process and virtue reliabilist accounts.

3.1. IS CONTENT EXTERNALISM INCOMPATIBLE WITH EPISTEMIC INTERNALISM?

Recall that in §1, we characterised epistemic internalism—construed as a thesis about epistemic justification[25]—as at least committed to the accessibilist thesis, that an agent can determine by reflection alone the factors that would make her beliefs epistemically justified; an associated 'negative' commitment of the accessibilist position is that agents do not diverge in the extent to which their beliefs are justified provided that they do not diverge in what is accessible to them by reflection alone.[26]

As we noted, stronger versions of epistemic internalism about justification (hereafter, J-internalism) can endorse (along with the accessibilist claim) a *mentalist* thesis according to which, as Conee and Feldman put it, "a person's beliefs are justified only by things that are internal to the person's mental life."[27] (Feldman & Conee 2001, 233) But not all accessibilists must endorse mentalism.[28] We can test the bare compatibility of content externalism and epistemic internalism by testing the compatibility of content externalism with the widely endorsed (by internalists) accessibilist thesis.[29]

There are different ways to argue for the 'incompatibilist' thesis,[30] and we'll now consider three of the most-discussed arguments, what we can call: the *Accessibilist Argument*, the *Self-Knowledge Argument*, and the *Process Argument*.

3.1.1 *Bonjour's Simple Accessibilist Argument*

A straightforward argument for the incompatibilist thesis has been suggested by Bonjour (1992) in a brief passage from which we can, following James Chase (2001, 237), extract the following line of reasoning: [31,32]

> (1) If content externalism is true then there can be an agent *S* with belief B such that part or all of the content of B is not internally available to *S*.
> (2) If agent *S* with belief B is such that part or all of the content of B is not internally available to *S*, then the justification relations B stands in with other beliefs of *S*'s are not internally available to *S*.
> (3) If an agent *S* with belief B is such that the justification relations B stands in with other beliefs of *S* are not internally available to *S*, then not all factors relevant to the justification of beliefs of *S* are internally available to *S*.

But, as Chase notes, the consequent of (3) is the denial of J-internalism, and thus:

> (4) If content externalism is true, then J-Internalism is false

Bonjour's accessibility argument is a quick way to reach the incompatibilist thesis, but perhaps it is too quick.[33] One charge against the simple accessibility argument, leveled by Chase (2001, 238), and more recently by Brent Madison (2009),[34] is that the sense of internal availability (e.g., accessibility) precluded by content externalism in (1) is, as Chase puts it, 'not the sense of internal availability at issue in characterizing J-Internalism'. The incompatibilist conclusion (4) thus is dismissed as the product of illicit equivocation.

The equivocation Chase finds objectionable is argued to come out most clearly in new evil demon cases. To appreciate the objection, then, let's consider the example (slightly amended):

> DEMON WATER: Stan is a victim of an evil demon scenario. His belief 'there is water in front of me' (call this belief 'B') is, thus, false. Stan's belief is based on his visual experiences, together with other beliefs, such as 'water is the stuff usually found in lakes' (call this belief 'A') and the belief that he is currently looking at a lake.

If 'water' is *wide*, then Stan will be unable to individuate his belief A ('that water is the stuff usually found in lakes'), and hence won't be able to individuate the justification relations A stands in to B ('there is water in front of me.')

Now here's the rub of the objection. The J-Internalist herself *insists* that victims of evil demon scenarios (such as Stan, in DEMON WATER) are justified in their beliefs. And indeed, that is—*pace* epistemic externalists (e.g., Goldman 1986)—supposed to be the key *point* of these cases, for J-internalists.[35] But then, it follows that the justification relations A ('water is stuff that is usually found in lakes') stands in to B ('there is water in front of me') are *not factors relevant to the*

justification of S's belief B. Rather, the justification relations that are relevant are those that stand between A* and B*, where A* and B* have the same *narrow* content as A and B.[36] Thus, as Chase reasons, what would be a factor relevant to Stan's belief B ('there is water in front of me') will be that *A** beliefs are evidence for *B** beliefs, and "*this fact is internally available to S even in evil demon cases.*" (Chase 2001, 238) Thus, (*contra* (2)), the following conjunction can be true: an agent *S* with belief B can be such that (i) part or all of the content of B is not internally available to *S*; and (ii) yet it is not the case that the justification relations B stands in with other beliefs of *S*'s are not internally available to *S*. Thus, as the argument goes, new evil demon cases illuminate why it is that (2) in Bonjour's accessibility argument should be rejected.

If Chase is right, then notice that he will have effectively diffused an argument capturing what is perhaps the most straightforward way to articulate the incompatibilist insight.[37] But there are other ways to try to make the argument.

3.1.2 *The Self-Knowledge Argument*

Boghossian (1989), like Bonjour, worries that if the content of a thought is determined by its relational properties, then "we can not know our own minds."[38] Also, like Bonjour, Boghossian thinks that the sense in which content externalism prevents us from knowing our own minds is sufficient for precluding us from knowing by reflection alone (as J-internalism's accessibility constraint requires) the factors relevant to justifying our beliefs. Boghossian's argument however takes a different route to this conclusion.

His simple, valid argument has two premises and a conclusion.

(1) If J-Internalism is true, then all self-knowledge is non-inferential.
(2) If all self-knowledge is non-inferential, then content externalism is false.
(3) Therefore, if J-Internalism is true, then content externalism is false.

That premise (1) is true is not immediately obvious. Boghossian opts to defend (1) *via* a regress argument intended as a kind of *reductio* against the pairing of J-internalism with the position that self-knowledge is at least sometimes inferential.

To get the regress (aimed at establishing (1)) up and running, Boghossian opens with a point J-internalists have made against reliabilists, which is that justification requires that, as Bonjour (1985, 38-40) puts it, one "[must] grasp the connection between the evidence and what it is evidence for."[39] Now, in a case where *p* depends on *q*, Boghossian takes it this grasping condition entails the following: that if "I am to be justified in believing that *p*, I must believe that

p as a result both of my recognition that I believe that *q*, and that a belief that *q* justifies a belief that *p*." (Boghossian 1989/2008, 154)

He spells this out—i.e., the conditions that would have to be met for an agent to possess a justified inferential belief (on J-internalism)—more explicitly as follows:

1. *S* believes that *p*.
2. *S* believes that *q*.
3. The proposition that *q* justifies the proposition that *p*.
4. *S* knows that *S* believes that *p*.
5. *S* knows that a belief that *p* justifies a belief that *p*.
6. X believes that *p* as a result of knowledge expressed in 4 and 5. (Boghossian 1989/2008, 154)

But (4) invites a regress. If self-knowledge is inferential, then (4) is satisfied (by the J-internalist's lights) only if *S* knows has some *other* belief. As he writes:

> "But how was knowledge of *this* belief acquired? On the assumption that all self-knowledge is inferential, it could have been acquired only by inference from yet other known beliefs. And now we are off on a vicious regress." (Boghossian 1989/2008, 155)

Given the alleged unacceptability of this regress, then, Boghossian concludes (1): that J-Internalism entails that self-knowledge is non-inferential.

We'll return to this point. But next, let's consider the argument for Premise (2) of the self-knowledge argument—*viz*., the claim that if all self-knowledge is non-inferential, then content externalism is false. Much like the first premise, the second premise also needs unpacking.

The argument offered for (2) involves a 'slow-switching' case meant to establish that, if content externalism is true, then at least some self-knowledge is inferential. And if *that's* right, then premise (2) follows—if all self-knowledge is non-inferential, then content externalism is false.

The slow-switching[40] case runs as follows (slightly amended):

> TYLER: Tyler, unaware he is on Twin Earth, uses the words 'I have arthritis'—to express the thought *I have tharthritis*. (Note: on content externalism, 'arthritis' refers to *tharthritis* on Twin Earth). But in order to know order to know the content of his own thought—i.e., *I have tharthritis*—it is necessary that Tyler be able to exclude a relevant alternative—*viz*., that his thought is *I have arthritis*--something that he can't do (*ex hypothesi*). But then, if content externalism is true, then at least some self-knowledge is inferential as excluding a relevant alternative is part of an inferential process.[41]

And of course, if content externalism entails that some self-knowledge is inferential, then Boghossian gets his premise (2), that if all self-knowledge is non-inferential, content externalism is false.

There is not space to consider all, or even most, lines of resistance offered to (1) and (2) in the *Self Knowledge Argument*. We'll look quickly, though, at some of the notable objections. Firstly, regarding (1): it's at best not clear that the regress doing the relevant work for Boghossian in establishing (1) is actually a *special* problem apart from the 'standard' problem of the regress of justification[42]—which would have to be confronted by *any* theory of epistemic justification.[43] But unless it can be established as a special problem, then the target of the regress is too broad to be compelling as support for (1). Moreover, the objection has been raised that the self-knowledge argument only establishes that, if J-Internalism is true, *some* self-knowledge is non-inferential.[44]

Regarding (2) of the argument. There are several ways we might resist the suggestion that the TYLER case establishes (as it's supposed to) that content externalism entails that some self-knowledge is inferential (and, therefore, that (2): if all self-knowledge is non-inferential, content externalism is false.) One such line of resistance here will be to challenge Boghossian's insistence that (in our case) Tyler must *rule out* the scenario whereby the content of his belief is *I have arthritis* in order to know that the content of his thought is that he has *tharthritis*.[45] Another line of resistance will be to grant that such an alternative must be ruled out, but to then deny that satisfying this epistemic condition must involve (as the argument requires) an *inference*.[46]

3.1.3 *The Process Argument*

The *Process Argument* has been defended by Chase (2001), and challenged by Tony Brueckner (2002). Unlike the previous two arguments we've considered, which have attempted to demonstrate the incompatibility of content externalism and *accessibilist* J-internalism, the *Process Argument* tries to demonstrate the incompatibility of content externalism and *mentalist* J-internalism according to which internal duplicates are justificational duplicates.[47]

The argument begins with a kind of 'process claim' according to which, for two justificational duplicates, a_1 and a_2, and proposition *p*, if a_1 and a_2 believe *p*, then the 'justificatory processes' leading to a_1's belief that *p* and a_2's belief that *p* are identical. Following Chase, call this the Process Claim (PC).

Like the *Self Knowledge* Argument, the *Process Argument* has just two premises and a conclusion:

(1) If PC is false, then mentalist J-internalism is false.
(2) If Content Externalism is true, then PC is false.
(3) Therefore, if content externalism is true, then mentalist J-Internalism is false.[48]

The argument or (1) is that PC follows from mentalist J-internalism 'and a plausible enough claim about the relevance of justificatory processes to the justification of beliefs.' The idea is that, supposing mentalist J-internalism is true, then if two agents are identical in the 'internal physical constitution on which their minds supervene' then they'll be identical in 'all respects relevant to the justification of their beliefs'—and Chase takes it that the justificatory processes will be relevant in this way. So, mentalist J-internalism entails PC, and thus if PC is false, so is J-internalism—premise (1).

The argument for (2) is supposed to establish the conditional claim in (2) by showing (*via* counterexample) that the antecedent in (2) is inconsistent with the negation of its consequent—*viz.*, that content externalism is inconsistent with (the denial of the falsity of) PC. The counterexample runs like this:

> TWIN SUE: Sue is on Earth. Twin Sue is on Twin Earth. Both Sue and Twin Sue reason *via* the following sentences:
>
> **(i)** It is possible to drink water
> **(ii)** Water is liquid
> **(iii)** Hence it is possible to drink a liquid
>
> As a result of this reasoning, Sue and Twin Sue believe that it's possible to drink a liquid and express this in the usual way.[49]

Now, Chase insists that in TWIN SUE, we have a case where if PC is true, then the justificatory processes used by Sue and Twin Sue will be the same;[50] *but* if content externalism is true, then the thoughts operative in their reasoning (in (1) and (2)) to the belief that it's possible to drink a liquid differ,[51] and *ipso facto*, the justificatory processes they use differ. But then, if content externalism is true, PC can't also be true. Hence premise (2): If content externalism is true, then PC is false.[52]

What to make of the *Process Argument?* Unsurprisingly, a hot-wire premise is (2).[53] Consider, as Brueckner (2002, 514) has suggested, that the TWIN SUE case could do the work it needs to do to establish (2) in the *Process Argument* only if (as Chase suggests) PC actually has as an implication that that Sue and Twin Sue have the same justificatory processes. But, as Brueckner (2002, 514) suggests, only a mentalist J-internalist is bound to accept this; a content externalist will simply deny it. After all, the content externalist will insist that their justificatory processes differ because Sue and Twin Sue are reasoning through different beliefs (*water* and *twater* beliefs), respectively, in (i) and (ii).

3.1.4. *Does Content Externalism Entail Epistemic Externalism?*

It is tempting to think that if content externalism were to entail epistemic externalism, then there should be a simple and obvious way to make the point. What the discussion in §3.1 shows, however, is that this simply isn't so. While there are different ways to try to establish the incompatibility of content externalism and epistemic internalism (construed along both accessibilist and mentalist lines), none was a knock-down argument—and in each case, the real philosophical debate was shown to shift quickly to the details of the particular premises relied on, rather than to any clear philosophical thesis to be accepted or denied. A judicious position then is to say that the compatibility of content externalism and epistemic internalism remains a live question.

3.2. DOES ACTIVE EXTERNALISM ENTAIL EPISTEMIC EXTERNALISM?

Recall from §2 that *active* externalism comes in three varieties. In order of provocativeness: (i) extended cognition, (ii) the extended mind thesis, and (iii) the distributed cognition thesis. Because the extended and distributed cognition theses involve the extension of cognitive *processes* while the extended mind thesis concerns extended mental *states*, we will organize our discussion in §3.2 by considering first in §3.2.1 the compatibility of the *extended cognition* and *distributed cognition* varieties of active externalism with epistemic internalism, and then in §3.2.2 the compatibility of the *extended mind thesis* with epistemic internalism.

3.2.1 *Do Extended/Distributed Cognition Entail Epistemic Externalism?*

Because extended cognition is a logically weaker thesis than the distributed cognition thesis, it stands to reason that if extended cognition is not compatible with epistemic internalism, then neither will be the stronger distributed version of the thesis. We'll now investigate the compatibility of the extended cognition thesis (hereafter, EC) with epistemic internalism. Let's begin with accessibilist J-internalism: the thesis that an agent can determine by reflection alone the factors that would make her beliefs epistemically justified. However, as we'll see, the case we appeal to in attempting to show this will be instructive in allowing us to chart some further (in)compatibilities.[54]

A natural opening move here will be to argue that if EC is true, then (*contra* accessibilist J-internalism) it is possible for agents to diverge in the extent to which their beliefs are justified even when they do not diverge in what is accessible to them by reflection alone. Here's a case that would purport to show this.

> TWIN OTTO: Otto's biological memory is failing, and so is Otto*'s, in both cases, due to Alzheimer's. Otto, however, (from Clark & Chalmers's (1998) original example) relies on a notebook to store and retrieve information, in a way that is functionally equivalent to the way he used to (before Alzheimer's) rely on his biological brain. Otto*, however, usually can't be bothered carrying a notebook around (and accordingly, his world is comparatively less organized). Otto and Otto* agree to meet at the Museum of Modern Art on 53rd street. Otto writes this down (as he usually does). Otto*—uncharacteristically for him—buys a notebook and on this occasion writes down that the Museum of Modern Art is on 53rd street. Five minutes later, at time T, neither Otto nor Otto* have the address of the Museum of Modern Art in biological storage, though at T they both know they can simply consult their respective notebooks later when the information is needed.

If the two assumptions are granted, we can (with reference to TWIN OTTO) advance an argument to the effect that EC is incompatible with accessibilist J-internalism. The first assumption is that (i) tacit knowledge (and, thus, tacit epistemic justification) is ubiquitous. Five minutes ago, you knew, and thus were justified in believing, that Paris is the capital of France. The second assumption is that (ii) you knew (five minutes ago) that Paris was the capital of France because you (five minutes ago) could have (in virtue of possessing the cognitive abilities you possess) easily recalled this, had you been asked.

To motivate the argument, consider now that if EC is true, then Otto in TWIN OTTO is justified in believing 'The Museum of Modern Art' is on 53rd street' at time T *so long as* (say) a counterpart of Otto with a normally functioning biological memory would count as having tacit/dispositional knowledge (and thus justification) of this same proposition even when it is not being occurrently entertained. This much would be implied by the parity principle; and our two assumptions.

Now, here's the argument that EC is incompatible with accessibilist J-internalism:

(1) Otto* does not possess justification for the belief 'The Museum of Modern Art is on 53rd street' at T.
(2) If EC is true, then Otto does possess justification for the belief that 'The Museum of Modern Art is on 53rd street' at T. [From the parity principle, plus our two assumptions]
(3) If accessibilist J-internalism is true, Otto is justified in believing 'The Museum of Modern Art is on 53rd street' at T only if Otto* is justified in believing 'The Museum of Modern Art is on 53rd street' at T.
(4) Therefore, if EC is true, accessibilist J-internalism is false. [From 1, 2 and 3]

First, in support of Premise (1): Otto* not only has no occurrent justified belief that the Museum of Modern Art is on 53rd street at T, but also, he has no *tacit* or dispositional justified belief on the matter at T. If EC were false, this would of course be so, because Otto*'s notebook would not in principle be part of his cognitive process. But *even if EC were true*, there would be no grounds for attributing to Otto* a tacit or dispositional justified belief. This is because, for Otto*, the notebook containing the information *is not part of his cognitive process*, even if EC is true. It is after

all not appropriately integrated into his cognitive system.[55] Premise (2) of the argument gains its support from the fact that for Otto, the notebook will (by reference to the parity principle) *de facto* be appropriately integrated into his cognitive system to qualify as part of his cognitive processes.[56] And, with our two initial assumptions, (2) follows. Finally, (3) of the argument is supported by pointing out that Otto and Otto* are in the same position, *vis-à-vis* the location of the Museum of Modern Art, given what is accessible to them by reflection alone.

What to make of this argument? Firstly, it is tempting to cry foul and insist that TWIN OTTO shows that EC entails the falsity of accessibilist J-internalism *only if* EC is committed to the claim that some epistemic *states* are extended. EC is committed only to the claim that cognitive processes are extended. So EC is not committed to the falsity of accessibilist J-internalism.

But this reply is too quick. We can rely on an alternative reading of the Otto case, according to which the tacit/dispositional justification Otto plausibly has at T is not a matter of possessing any extended *mental* state. Rather, it is a dispositional state he has because, on EC, he has a certain cognitive ability in virtue of his extended cognitive process. If such an ability is good enough to ground tacit/dispositional justification in the case of biological memory, then it should be good enough to do so no less (by EC's lights) in the extended case. But then, of course, since Otto* obviously possesses no justification (dispositional or otherwise) for a belief about the location of the Museum at T, it looks like EC entails that accessibilist J-internalism is false, given that Otto and Otto* are in the same position *vis-à-vis* what is internally available to them.

But there is a second, perhaps more powerful, way to resist the argument that TWIN OTTO shows that EC entails the falsity of accessibilist J-internalism. The second route of resistance will claim that the tacit/dispositional justification EC is committed to claiming Otto has (so long his normally functioning counterpart with a biological memory possesses tacit/dispositional justification at T) is not going to satisfy the accessibilist condition. After all, at T, *reflection alone* (this line of argument goes) affords Otto no more than it affords Otto* *vis-à-vis* the target belief, at T, and this is so *even if* (for the reasons cited in support of (2)) EC is committed to counting Otto as having tacit/dispositional justification.

At this stage, however, it would at least first appear that the proponent of the argument that EC entails the falsity of accessibilist J-Internalism has an ace up her sleeve. The ace here (in reply to this second line of resistance to the argument) is to *reject* the claim that Otto and Otto* are in the same position *vis-à-vis* the target belief at T, in light of what they have access to by reflection alone. After all, it could be argued, according to EC, cognitive processes are extended. Reflection is after all a particular type of *cognitive process*. But then (given EC), the entry in the

notebook is available to Otto at T by (extended) reflection alone, though not to Otto*.[57] Of course, an opponent of EC will deny *ex ante* that reflection is extended—but the defender of the argument need only argue that *on EC*, reflection, *qua* cognitive process, could be extended.

But even this 'ace up the sleeve' only goes so far. The reason for this is that this strategy seems to rebut the second line of challenge to the argument but only at the cost of standing in tension of (3) of the argument. Ultimately, it seems like the matter of whether extended cognition is compatible with accessibilist J-internalism is a live question.

However, as we noted, the TWIN OTTO case will also be useful in helping us chart a few other (in)compatibilities. As a general point, though it's not yet obvious whether *active externalism* (per se) entails epistemic externalism because it's not clear whether the weakest version of active internalism (EC) is incompatible with accessibilist J-internalism, it does seem that (perhaps) EC is incompatible with epistemic internalists who endorse *mentalist* J-internalism. Recall that a commitment of mentalist J-internalism is that agents do not diverge in the extent to which their beliefs are justified if they have the same mental states. If we make explicit a further detail in the TWIN OTTO case—e.g., that Otto and Otto* are in the same mental states at T—then:

(1) Otto and Otto* are in the same mental states at T.

And from here, the argument is straightforward. Mentalist J-internalism, along with (1), yield (2):

(2) If mentalist J-internalism is true, Otto and Otto* do not diverge, at time T, in the extent to which they are justified (respectively) in believing that the Museum of Modern Art is on 53rd street (from 1).

Because, on EC, Otto has the correct location of the Museum of Modern Art in memory storage, but Otto* doesn't[58], it is plausible that Otto and Otto* *diverge* in the extent to which they are justified (respectively) in believing that the Museum of Modern Art is on 53rd street. Thus:

(3) If EC is true, then Otto and Otto* diverge in the extent to which they are justified (respectively) in believing that the Museum of Modern Art is on 53rd street.

And, from (1-3) we get:

(4) Therefore, if EC is true, mentalist J-internalism is false.

Is this argument compelling? One might challenge (3) as follows: Otto and Otto* diverge in the extent to which they are justified (respectively) in believing the target proposition only if they both *believe* the target proposition in the first place. But in TWIN OTTO, the target proposition is no longer stored in biological memory for Otto or Otto*. While Otto would count

as believing the target proposition if EC was committed to recognizing extended mental states, EC does *not* have this commitment (this, rather, is the *extended mind thesis*). So, therefore, not even Otto believes the target proposition. Thus, premise (3) is false because (to put it bluntly) Otto and Otto* don't *diverge* in their justification for beliefs they don't have.

At this point, a defender of the argument that EC entails the denial of mentalist J-internalism seems to have two moves. One move is to claim that (3) is true so long as Otto and Otto* diverge in the extent to which they are tacitly/dispositionally justified in the target proposition—*viz.*, were both to entertain the proposition, Otto* would (by EC) count as justified, but not Otto*. Another move will be to draw attention to the fact that if Otto is even *marginally* more justified than Otto*, then (3) is true, and then to insist that, if we embellish the TWIN OTTO case so that both actively consult their respective notebooks (e.g., at time T1), then on EC Otto's belief is supported by *both* perception *and* memory, whereas Otto*'s belief is supported only by perception; further that this difference renders Otto's target belief justified to a greater extent than Otto*'s, in a non-negligible way.[59]

While the manoeuvres we've considered here on behalf of the proponent of an argument for the incompatibility of EC and mentalist J-internalism are not obviously compelling—we did see (in the process of considering the objection to (3) of the argument) that the *extended mind thesis* would be seem, on the face of things, incompatible with mentalist J-internalism. We'll return to this issue.

For now, though, it seems that in terms of *bare compatibility*, the weakest articulation of the active externalist thesis—the thesis of extended cognition—is not *obviously* incompatible with either accessibilist or mentalist J-internalism. What about the *stronger* versions of active externalism?

3.2.2 *Does the Extended Mind Thesis Entail Epistemic Externalism?*

As the previous argument indicated, we could run a simple line of reasoning to the conclusion that the extended mind thesis, according to which mental states are extended, is incompatible with *both* mentalist and accessibilist J-internalism. The quick and easy argument goes like this:

(1) If either accessibilist or mentalist J-internalism is true, then at T, Otto and Otto* are justified to the same extent in the proposition that the Museum of Modern Art is on 53rd street.
(2) If the extended mind thesis is true, then Otto and Otto* are not justified to the same extent at T in the proposition that the Museum of Modern Art is on 53rd street.
(3) Therefore, if the extended mind thesis is true, then accessibilist and mentalist J-internalism are false.

Support for (1) is predictable: on accessibilist J-internalism *and* on mentalist J-internalism, Otto and Otto* are equally justified (which is, not at all) *vis-à-vis* the target proposition, at T, given that they do not diverge in what is reflectively accessible (e.g., to their failing biological brains) nor do they diverge in their mental states, at T. Premise (2) can also be supported straightforwardly: the extended mind thesis (e.g., Clark *&* Chalmers 1998) extends not only Otto's cognitive process, but also his belief, in light of the fact that the notebook has been appropriately integrated into his cognitive system. Since Otto*'s notebook, *ex hypothesi,* is *not* appropriately integrated into his cognitive system, he does not have an extended belief in the target proposition at T. Thus, premise (2): if the extended mind thesis is correct, Otto and Otto* diverge in their justification of the target belief at T because Otto, but not Otto*, has a justified *belief* that the Museum of Modern Art is on 53rd street, at T. And, given (1) and (2), what follows is the incompatibility of *either* accessibilist or mentalist J-internalism and the extended mind thesis.

But notice that the epistemic internalist who is a proponent of the extended mind thesis will challenge (1) in the argument. The objection to (1) goes like this: *if mental states are extended*, then (1) in the above argument is obviously false. After all, if mental states are extended, then *if* mentalist J-internalism is true, Otto and Otto* are not justified to the same extent in their beliefs, at T, because Otto, *but not Otto** meets the condition of *believing* the target proposition at T (he, but not Otto*, would have an extended belief in the target proposition, at T). The same kind of point can be made *vis-à-vis* accessibilist J-internalism: if mental states are extended, then (1) in the argument fails because the location of the Museum of Modern Art would be internal to Otto's, but not Otto*'s mental life.[60]

What this all shows is that the simple argument for the incompatibility of the extended cognition thesis and epistemic internalism seems to go through *only if* the extended mind thesis is not true. If it is true, it's not obvious the extended mind thesis and epistemic internalism are incompatible—we simply would replace (in the traditional formulations) the mental states with extended mental states and allow reflective accessibility to include extended reflective accessibility, and the tension between the views appears to dissolve.

The proponent of epistemic internalism (of either variety) might balk at this suggestion and insist that the traditional formulations of the epistemic internalism (both in its accessibilist and mentalist guises) are necessarily views that preclude extended mental states and extended reflective accessibility. The accessibilist, for instance, might say that 'by reflection alone' what is meant is *obviously* something like: by merely reflecting in a way where the reflective experience is an activity of the agent's *conscious* mind, where the associated reflective mental states supervene on the agent's biological brain.

This raises a more subtle point. Notice that while *mentalist* J-internalism is compatible with the extended mind thesis insofar as mentalist J-internalists insist just that justification supervenes on mental states (which, on the extended mind thesis, will simply be more inclusive than otherwise), there is perhaps a deeper problem associated with the point about accessibilism. The problem is that, to the extent that a presupposition of the accessibilist's claim is that the factors relevant to a belief's justification be accessible *to one's consciousness*, then, to the extent that consciousness itself lacks any kind of extended analogue, it would seem that there could be no pairing of the extended mind thesis and accessibilist J-internalistm. The compatibility of accessibilist-J internalism and the extended mind thesis seems to turn importantly, then, on both (i) whether the accessibilist J-internalist's claim *must* be unpacked, at the end of the day, with reference to consciousness, and whether, if so, consciousness admits of any kind of extended analogue.[61]

3.3. ACTIVE EXTERNALISM AS EPISTEMIC EXTERNALISM

Whilst arguments for the incompatibility of active externalism and epistemic internalism turn out to be on the whole not very satisfying, the point remains that active externalism and epistemic externalism would seem to be, at least *prima facie,* a more natural fit. In this section, we will explore which varieties of externalism in epistemology would square best with varieties of active externalism. In particular, we'll consider pairings of active externalism with *modal* epistemology and *virtue* reliabilism.

3.3.1 *Active Externalism and Modal Epistemology*

Whereas we investigated the *incompatibility* of content externalism and epistemic internalism in terms of J-internalism theses (rather than K-internalism theses), given that K-internalism entails J-internalism, it makes sense to do things the other way around when considering which variety of epistemic externalism is the *best fit* with active externalism.

As we noted in §1, the *process reliabilist* account of knowledge insists that a true belief's origin in a reliable process is both necessary and sufficient for knowledge. While this is a well-known externalist account of knowledge, it is widely taken, at least in mainstream epistemology, to have several serious problems. One problem for the material adequacy of process reliabilism is that the account appears too inclusive; according to process reliabilism, it seems as though *barn*

façade cases (where an agent spots a genuine barn, but easily could have pointed to a nearby façade) are *bona fide* cases of knowledge. More generally, the worry is that process reliabilism fails to preserve the insight that knowledge is incompatible with epistemic luck. A second worry is that the reliabilist account fails to capture the widely shared insight that knowledge must be appropriately connected to cognitive *ability*. Virtue reliabilists offer an externalist account of knowledge that is meant to preserve that insight, and we'll examine the compatibility of active externalism with virtue reliabilism in some more detail in the following subsection.

But first, it will be instructive to assess the compatibility of active externalism with externalist accounts of knowledge that are constructed to guard against the kind of epistemic luck that (as barn façade cases show) mere process reliabilism was unable to preclude. More generally, then, what we'll examine now is the compatibility of active externalism with *modal epistemology*. Kelly Becker (2010, §1) describes modal epistemology as aiming to "articulate conditions on knowing such that no merely lucky true belief—very roughly, a belief that might easily have been false—counts as knowledge."

The two most notable approaches on this score are *sensitivity* and *safety* based accounts of knowledge. On sensitivity-based accounts (e.g., Nozick 1981), one knows that *p* so long as one's true belief has the following modal property: had *p* been false, *S* wouldn't have believed that *p*. (Alternatively: *S* knows that *p* just in case *S* believes that *p*, *p* is true, and *S* does not believe that *p* in the nearest possible world where *p* is false). We will here set sensitivity-based accounts aside for the reason that, in recent years, *safety-based* accounts have played a much more central role in anti-luck debates in mainstream literature than sensitivity-based approaches.[62]

According to the *safety principle* (e.g., Sosa 1999), construed as a condition on knowledge: *S* knows that *p* only if, given the way *S* formed the target belief, *S* couldn't easily have been wrong.[63] This claim has been fleshed in out in a number of different ways. Here, for example, is a recent explication of safety offered by Pritchard (2012*c*):

> *Safety*: *S*'s true belief that *p* is safe iff in *most* near-by possible worlds in which *S* continues to form a belief in the same way as in the actual world, and in *all* very close near-by possible worlds in which *S* continues to form a belief in the same way as the actual world, her belief continues to be true.

Pritchard motivates this construal by appeal to what he calls an anti-luck epistemology, which is a way of thinking about knowledge which explicitly involves unpacking the notion of luck and the specific sense in which knowledge excludes lucky cognitive success. Pritchard argues that this methodology leads not just to a safety condition on knowledge, but moreover to the particular formulation of safety just given. Essentially, the idea is that knowledge requires safety in this

sense because it is completely intolerant to epistemic risk in close possible worlds, with an increasing tolerance to epistemic risk as one moves out to further possible worlds.[64]

So formulated, one might be tempted to offer a complete account of knowledge in terms of safety—*viz*., that knowledge is safe belief. For our purposes, however, we shall examine whether a safety-based account of knowledge, understood as offering a mere necessary condition on knowledge, will be a good fit with active externalism.[65] Let's assess this by considering the safety-based approach alongside the extended cognition (EC) variety of the active externalist program.

It's not immediately obvious to see how EC (as a thesis about the metaphysical nature of cognitive processes) fits together with a safety condition on knowledge. However, points of connection surface once we focus on the fact that, for *S*'s belief that *p* to satisfy the safety condition, we must *hold fixed* certain facts about the way the agent comes to believe the target proposition in the actual world. And at this point, it is fair to assume that among that which is held fixed under the description of the relevant way the belief was formed in the actual world, is the *cognitive process* that the agent employs in the actual world in forming the target belief.

Now the point of connection between EC and safety-based epistemology should be clear. If EC is true, then when (by safety) we assess whether a belief is lucky in a way incompatible with knowing, what we hold fixed as part of the cognitive process the agent deploys in the actual world when forming the target belief can include parts of the world outwith the biological agent herself. This turns out to have some interesting philosophical ramifications. Consider, for instance, the following case:

> NOTEBOOK JOKESTER: Otto** consults his notebook to determine when his doctor's appointment was today, and finds the correct time, noon, written in the book. Unbeknownst to Otto**, his notebook had been stolen by a jokester, who fudged with the times of Otto**'s other appointments that day, changing them all back an hour. The jokester, however, overlooked the doctor's appointment, leaving the original and correct time intact.[66]

Now, when determining whether Otto**'s belief is safe, we should hold fixed the relevant way Otto** forms the target belief in the actual world, and accordingly, we will *hold fixed* the cognitive process Otto** employs in the actual world when considering whether Otto** continues to believe truly in near-by worlds. If EC is true, notice that we will hold *fixed* Otto**'s notebook (which by EC is part of Otto**'s cognitive process) under the description of the cognitive process Otto** employs in the actual world when forming the target belief. But in doing so, notice that we are also holding fixed a notebook that *contains the correct time of Otto**'s doctor's appointment*. But then, Otto**'s belief is surely safe, as the correct entry comes with Otto** as we move out to other worlds and ask whether Otto** continues to believe truly.

Now, it's hard to deny the similarities between NOTEBOOK JOKESTER and barn façade cases, where environmental luck undermines the safety of the target belief. Thus, it is natural to assume that NOTEBOOK JOKESTER is not a case of knowledge but rather a case where Otto**'s belief is undermined by environmental luck. But, as we've seen, EC will have a hard time getting the intuitively right result in cases where the 'bad environment' is ruled-in as part of the cognitive process itself, via the hypothesis of extended cognition. This is not to say, of course, that EC is incompatible with safety-based accounts of knowledge. Rather, the lesson from NOTEBOOK JOKESTER seems to be that, insofar as EC is to be incorporated with mainstream thinking about epistemic luck, problems like the one mentioned here will have to be addressed.

3.3.2 *Active Externalism and Virtue Reliabilism*

So, given that there are good reasons to doubt the compatibility of modal epistemology and active externalism, how could we introduce this latter position within externalist epistemology? The most obvious starting point is the approach of process reliabilism, due to its central focus on the reliability of epistemic *mechanisms, methods, and processes*:

> *Process Reliabilism*
> *S* knows that *p*, if and only if *S*'s true belief is the product of a reliable belief-forming process.

Clearly, a formulation of knowledge along the lines suggested above is well suited to accommodate considerations originating from philosophy of mind and cognitive science whose aim is to provide a mechanistic understanding of the *cognitive processes* that constitute the machinery of the human mind.

There are, however, two serious complications with the view. The first one is that process reliabilism, as it stands, is too weak a condition on knowledge because it allows *any* reliable belief-forming process to count as knowledge-conducive, and this is intuitively incorrect. The second complication is that by making "*de facto* reliability the grounds of positive epistemic status" (Greco 1999, 284-5) process reliabilism misses a very important dimension of our epistemic nature. While it is true that in order to know we do need the way of forming our beliefs to be objectively reliable, this sort of objective justification is not sufficient on its own. What we further need is that we be subjectively justified in the sense that we must be somehow sensitive to the reliability of our evidence.[67] Process reliabilism, however, ignores this dimension of our epistemically sentient nature altogether to the extent that it has been even criticized that it equates us to mere stimulus-response automata. (Fuller 2012)

Now, in order to flag the above shortcomings of process reliabilism, contemporary epistemologists have accentuated the importance of what has come to be known as the *ability intuition* on knowledge—*viz*., knowledge involves cognitive success which is attributable, at least to some significant degree, to the manifestation of the subject's relevant cognitive ability (Sosa 1988; 1991; Plantinga 1993; Greco 1999; 2004; 2007; Pritchard 2009; 2010*a*; 2010*b;* 2012; Haddock, Millar *&* Pritchard 2010, chs. 1-4).[68] And in fact, upgrading process reliabilism with the ability intuition on knowledge gives rise to virtue reliabilism, which is supposed to encompass all of its predecessor's advantages while limiting the knowledge-conducive processes to only the genuinely cognitive ones. Specifically, according to Greco (1999, 2010), a belief-forming process counts as knowledge-conducive only if it is one of the reliable processes that make up or have been integrated into the agent's *cognitive character*.

In this way, we ensure that not all reliable processes will be knowledge-conducive, but only the genuinely cognitive ones. Moreover, on the basis of this approach, we can accommodate the internalist intuition that one must be somehow sensitive to the reliability of one's belief-forming processes, but in an epistemically externalist way that does not involve knowledge or even beliefs about the said reliability. Specifically, Greco (1999, 289) suggests that a promising strategy for doing so is to claim that "a belief *p* is subjectively justified for a person *S* (in the sense relevant for having knowledge) if and only if *S*'s believing *p* is grounded in the cognitive dispositions that *S* manifests when *S* is thinking conscientiously" (i.e., when *S* is motivated to believe what is true). In this way, the agent will employ his reliable cognitive processes only in circumstances that have not been problematic in the past, *and* he will be able to do so without even having *any* beliefs about their reliability.[69] These knowledge-conducive cognitive processes, however, should not be considered in isolation to each other. As Greco further claims, the dispositions/habits that a person manifests when she is thinking conscientiously intertwine with each other and give rise to what we may call one's *cognitive character* (Greco 1999, 290). So, overall, "a belief *p* has a positive epistemic status for a person *S* just in case *S*'s believing *p* results from the stable and reliable dispositions that make up *S*'s cognitive character." (Greco 1999, 287-8)

Now, according to Greco (1999, 2010), in order for a reliable process to be part of one's cognitive character it must be a normal one—strange processes cannot be part of the agent's cognitive character because they are not the kind of processes that a conscientious agent would employ. Also, the relevant process must be a disposition of the agent. First, because it is only dispositions or habits that can really count as character traits. And, second, because in the absence of reasons to believe that the relevant process is reliable, it is only dispositions or habits that one can become aware they are unreliable in certain circumstances and, so—without relying

on any beliefs about their reliability—use them conscientiously in the rest of the circumstances. So, according to virtue reliabilism, in order for a process to be knowledge-conducive it must be part of the agent's cognitive character and in order for that to be the case the relevant process must be a reliable, normal disposition of his.

So, with these conditions in mind, we may now notice that virtue reliabilism appears to be apt for an interpretation along the lines suggested by active externalism. Specifically, the above three features for a process to be part of the agent's cognitive character and thereby knowledge-conducive seem to be very similar to (if not the same as) the common-sense functionalist criteria that Clark has offered in order for an external element to count as part of one's (extended) mind. Specifically, both views state that in order for a process to count as a cognitive ability it must be (a) reliable (i.e., not subject to critical scrutiny), (b) one of the agent's normal dispositions (such that it will be typically invoked), and (c) integrated within the rest of the agent's cognitive system/character (such that it will be easily accessible as if it were part of the agent's organismic cognitive apparatus).[70]

Notice further, however, that even if there is no such close fit between these broad (common-sense functionalist) features, virtue reliabilism may still be tightly connected to active externalism. This is because, despite Greco's insistence on the importance of the normality and dispositionality of the relevant process in order for it to count as a part of one's cognitive character (and thereby as knowledge-conducive), in other places, Greco attempts to further accentuate and shed some light on the *integrated* nature of our cognitive characters by noting that the process of "cognitive integration is a function of cooperation and interaction, or cooperative interaction with other aspects of the cognitive system." (Greco 2010, 152) And conceiving of cognitive integration in this alternative way is the same as the way philosophers of cognitive science account for it, by appealing to dynamical systems theory.[71]

In other words, even if there is no tight connection between virtue reliabilism and active externalism in the form of the extended mind thesis, virtue reliabilism will still appear to be closely connected to active externalism in the form of both the extended and distributed cognition hypotheses; all three views put forward the same criterion (cooperative interaction with the rest of the agent's cognitive system) in order for a process to be integrated to a cognitive system and thereby, according to virtue reliabilism, count as knowledge-conducive. Accordingly, it is possible to have extended as well as distributed (we will expand on this possibility in the section to follow), knowledge-conducive belief-forming processes.

This, however, may not really come as a surprise. Given that virtue reliabilism holds that knowledge must be the product of cognitive ability (however that ability may be realized) and

that the hypotheses of extended and distributed cognition set out to reveal which processes can count as cognitive abilities (wherever they may be located), this close fit between the three theories were to be expected. By making no specifications as to whether knowledge-conducive cognitive abilities should be located within the agent's head, virtue reliabilism has always allowed for the possibility that an epistemic agent's cognitive character may extend to the artifacts she employs or even be distributed amongst herself and other agents she may mutually interact with.

So to close this section, the upshot of the foregoing is a conception of knowledge that we call *extended knowledge*. (Palermos & Pritchard 2013) Rather than understanding knowledge as an internal phenomenon (either in the epistemic or the cognitive sense), we instead get a conception of knowledge which can be *extended* in terms of the subject's interactions with epistemic artifacts and which can also be *distributed* in terms of her social epistemic interactions. In short, the combination of active externalism with virtue reliabilism can motivate the claim that we can have extended and distributed cognitive systems that generate knowledge in the exact same way that our individual/organismic cognitive apparatus does.

4. COMBINING EPISTEMIC AND COGNITIVE EXTERNALISM—RAMIFICATIONS

The possibility of knowledge-conducive cognitive characters that may nevertheless be extended beyond our organismic cognitive capacities can generate interesting ramifications both for traditional and social epistemology—ramifications that may allow us to think about knowledge in new ways. Surprisingly, however, allowing for our cognitive characters to extend may not only be important for moving forward, but also necessary for accounting for knowledge as we already think about it.

According to Greco (1999, 287), in addition to one's organismic cognitive abilities of the brain/central nervous system, a person's cognitive character may also consist of "acquired skills of perception and acquired methods of inquiry including those involving highly specialized training or even advanced technology". The reason for this move is that we need to account for advanced cases of knowledge where one's believing the truth is the product of the operation of epistemic artifacts such as telescopes, microscopes, tactile visual substitution systems and so on.[72] The problem, however, is that in the traditional conception, cognition takes place strictly within the agent's head and so artifacts cannot be parts of one's *cognitive* character.

A possible way out of the problem for virtue reliabilism could be to claim that, in such cases, it is merely the agent's neural/bodily architecture, supporting the agent's skill of using the

artifact, that is the most salient factor in the causal explanation of the agent's cognitive success (i.e., believing the truth). Notice, however, that when an agent employs an epistemic tool, his cognitive success arises on the basis of the *mutual interaction* between his internal processes and the artifact. According to dynamical systems theory, then, the cognitive process that allows the agent to detect the truth is not merely 'aided' or 'assisted' by the artifact but is, instead, constituted by it.[73] What this means is that, in a causal explanation of how the agent acquired his true belief, there will be no way to disentangle the agent's training and skill of using the artifact from his actual engagement with it. [74]

Even if such decomposition were possible, however, we should further notice that the part of the process that allows the agent to detect the truth, or in other words to be sensitive to the facts, is the external component. Think of an untrained agent in possession of a properly working artifact. Obviously, even though he will initially be unable to form any (true or false) beliefs, eventually—provided that he gains sufficient experience—not only will he form beliefs, but he will also reliably enjoy cognitive success. In contrast, imagine now a well-trained agent, but in possession of a faulty artifact. In this case, despite his excellent internal skills, it is evident that he will be unable to reach any (non-lucky) true beliefs, no matter how much he tries. We can therefore claim that in such cases the most (and maybe the only) significant factor that explains the truth-status of the agent's belief is the epistemic artifact. In other words, since the agent's belief is true in virtue of the artifact, the virtue reliabilist must account for it being part of the agent's cognitive system. Due to the fact that cognition is normally supposed to take place within the agent's head, however, it appears that virtue reliabilists can only account for such cases by wedding their view to the hypothesis of extended cognition. Accordingly, combining the extended cognition hypothesis with virtue reliabilism on the basis of their close fit does not seem to be just an available option for epistemologists, but also necessary for dealing with advanced cases of knowledge where the latter is the product of the employment of epistemic artifacts that the agent mutually interacts with. (Palermos 2011; *forthcomingb*).

Apart from the apparent necessity of incorporating the extended cognition hypothesis into epistemology, however, this move can also generate interesting ramifications, especially for social epistemology. For instance, it can lead to the further claim that there could be cognitive characters that do not just extend beyond an agent's organismic capacities, but which are instead *distributed* amongst *several agents* along with their epistemic artifacts.

This is an interesting possibility, because it can allow us to combine an individualistic approach to knowledge, such as virtue reliabilism, with the hypothesis of distributed cognition in order to account for *epistemic group agents*: Groups of individuals who exist and gain knowledge in

virtue of a shared, common cognitive character that mainly consists of a distributed cognitive ability. Such a collective cognitive ability emerges out of the members' mutual (socio-epistemic) interactions and is not reducible to the cognitive abilities possessed by the individual members, thereby allowing us to speak of a group agent in itself. This is important, because by being able to so conceptualize a group of people as a self-standing agent, we can use an *individualistic* condition on knowledge to account for knowledge that is collectively produced and which is, thereby, distinctively social.[75]

In fact, such group agents have already started being studied within cognitive science. A case in point is transactive memory systems (TMSs)—i.e., groups of two or more individuals who collaboratively encode, store and retrieve information. The reason why TMSs are good candidates for distributed cognitive systems—and thereby for epistemic group agents—is that, as John Sutton *et al* (2008) observe, such systems are likely to involve skillful interactive simultaneous coordination of people who can thereby count as a single integrated cognitive system. Therefore, we can use TMSs in order to "conceptualize how people in close relationships may depend on each other for acquiring, remembering, and generating knowledge" (Wegner *et al* 1985*,* 253):

> "Ordinarily psychologists think of memory as an individual's store of knowledge, along with the processes whereby that knowledge is constructed, organized, and accessed. […] With transactive memory we are concerned with how knowledge enters the dyad, is organized within it, and is made available for subsequent use by it." (Wegner *et al,* 256)

Apart from incorporating already existing research from cognitive science, however, the combination of virtue reliabilism with the hypotheses of extended and distributed cognition can generate new avenues for research, some of which have for a long time been inaccessible. An interesting example is the intersection between epistemology and the field of history and philosophy of science. These two intimately related fields have so far been at odds—an awkward situation owing to the fact that the former discipline has traditionally being individualistic whereas the latter has for the most part been socially oriented (hardly anyone could deny the social nature of the scientific process, especially after the publication of Thomas Kuhn's *The Structure of the Scientific Revolutions*, in 1962). The suggested approach, however, could now provide a useful link between the two fields. Science is primarily performed by individual scientists employing their hardware and software epistemic artifacts or by *research teams* operating within scientific labs that are uniquely tailored to fit their purposes. Accordingly, the concepts of extended cognitive characters and epistemic group agents could become very handy for a mainstream epistemological analysis of the scientific progress. Indicatively, as Giere and Moffat (2003, 308) note in their discussion of the scientific revolution of the 16th century,

> "'No 'new man' suddenly emerged sometime in the sixteenth century [...] The idea that a more rational mind [...] emerged from darkness and chaos is too complicated a hypothesis.' [Latour 1986, 1] We agree completely. Appeals to cognitive architecture and capacities now studied in cognitive sciences are meant to explain how humans with normal human cognitive capacities manage to do modern science. One way, we suggest, is by constructing distributed cognitive systems that can be operated by humans possessing only the limited cognitive capacities they in fact possess." (Giere & Moffat 2003, 308)

5. CONCLUSION

As we hope the foregoing illustrates, the intersection of the epistemic and cognitive internalism/externalism debates is an exciting topic that raises a wide variety of questions across both epistemology and philosophy of mind and cognitive science. In this first attempt to tackle these questions (more will certainly follow soon), we ventured to stir up interest and offer a topography of the relevant philosophical terrain with regard to the possible ways in which knowledge can be conceived of as extended. Our tentative conclusion is that even though epistemic internalism is not clearly incompatible with cognitive externalism in the form of content (or passive) externalism, when we substitute the latter with active externalism the tension becomes more apparent. Moreover, there seem to be new philosophical problems that arise for modal epistemology in so far as it is paired with active externalism. Nevertheless, the combination of the latter position—at least in the form of the extended and distributed cognition hypotheses—with the externalist approach of virtue reliabilism does appear to be a promising strategy for conceiving of knowledge as extended.[76]

REFERENCES

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

Adler, J. (2002). *Belief's Own Ethics*, Cambridge, MA: MIT Press.

Alston, W. P. (1986). 'The Deontological Conception of Epistemic Justification', *Philosophical Perspectives* 2, 257-99.

Bach, K. (1985). 'A Rationale for Reliabilism', *Monist* 68, 246-63.

Bach-y-Rita, P., & Kercel, S. W. (2003). 'Sensory Substitution and the Human-Machine Interface', *Trends in Cognitive Science* 7, 541-46.

Becker, K. (2007). *Epistemology Modalized*, London: Routledge.

Becker, K. (2010). 'Modal Epistemology', *Oxford Bibliographies: Philosophy*, DOI: 10.1093/obo/9780195396577-0139.

Bergmann, (2006). *Justification Without Awareness: A Defense of Epistemic Externalism*, Oxford: Oxford University Press.

Black, T. (2002). 'A Moorean Response to Brain-in-a-Vat Scepticism', *Australasian Journal of Philosophy* 80, 148-63.

——— (2008). 'Defending a Sensitive Neo-Moorean Invariantism', *New Waves in Epistemology*, (eds.) V. F. Hendricks & D. H. Pritchard, 8-27, Basingstoke: Palgrave Macmillan.

Black, T., & Murphy, P. (2007). 'In Defense of Sensitivity', *Synthese* 154, 53-71.

Block, N. (1990). 'Inverted Earth', *Philosophical Perspectives* 4, 53-79.

Boghossian, P. (1989). 'Content and Self-Knowledge', *Philosophical Topics* 17, 5-26.

——— (1992). 'Externalism and Inference', *Philosophical Issues* 2, 11-28.

BonJour, L. (1985). *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.

——— (1992). 'Externalism/Internalism', *A Companion to Epistemology*, (eds.) J. Dancy & E. Sosa, 132-36, Oxford: Blackwell.

——— (2002). 'Internalism and Externalism', *Oxford Handbook of Epistemology*, (ed.) P. Moser, 234-64, Oxford: Oxford University Press.

Bonjour, L., & Sosa, E. (2003). *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*, Oxford: Blackwell.

Brooks, R. (1991*a*). 'Intelligence without Representation', *Artificial Intelligence* 47, 139-59.

——— (1991*b*). 'Intelligence without Reason', *Proceedings of 12th International Joint Conference on Artificial Intelligence*, 569–95.

Brueckner, A. (2002). 'The Consistency of Content-Externalism and Justification-Internalism', *Australasian Journal of Philosophy* 80, 512-25.

Burge, T. (1986). 'Individualism and Psychology', *Philosophical Review* 95, 3-45.

Carter, J. A. (*Forthcoming*). 'Extended Cognition and Epistemic Luck' *Synthese.*

Chase, J. (2001). 'Is Externalism about Content Inconsistent with Internalism about Justification?', *Australasian Journal of Philosophy* 79, 227-46.

Chemero, A. (2009). *Radical Embodied Cognitive Science*, Cambridge, MA: MIT Press.

Chisholm, R. M. (1977). *Theory of Knowledge* (2nd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Clark, A. (1997). *Being There: Putting Mind, Body, and World Together Again*, Cambridge, MA: MIT Press.

——— (2007). 'Curing Cognitive Hiccups: A Defense of the Extended Mind', *Journal of Philosophy* 104, 163-92.

——— (2008). *Supersizing The Mind: Embodiment, Action, and Cognitive Extension*, Oxford: Oxford University Press.

——— (2009). 'Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head', *Mind* 118, 963-93.

——— (2010). 'Memento's Revenge: The Extended Mind, Extended', *Extended Mind*, (ed.) R. Menary, 43-66, Cambridge, MA: MIT Press.

Clark, A., & Chalmers, D. (1998). 'The Extended Mind', *Analysis* 58, 7-19.

Cohen, S. (1984). 'Justification and Truth', *Philosophical Studies* 46, 279-96.
Conee, E., *&* Feldman, R. (2004). *Evidentialism*, Oxford: Oxford University Press.
——— (2011). 'Reply to Pritchard', *Evidentialism and its Discontents*, (ed.) T. Dougherty, 440-44, Oxford: Oxford University Press.
Dougherty, T. (ed.) (2011). *Evidentialism and its Discontents*, Oxford: Oxford University Press.
Dretske, F. (1970). 'Epistemic Operators', *Journal of Philosophy* 67, 1007-23.
——— (1971). 'Conclusive Reasons', *Australasian Journal of Philosophy* 49, 1-22.
——— (1996). 'Phenomenal Externalism, or If Meanings Ain't in the Head, Where Are Qualia?', *Philosophical Issues* 7, 143-58.
Engel, M. (1992). 'Personal and Doxastic Justification', *Philosophical Studies* 67, 133-51.
Feldman, R. (2005). 'Respecting the Evidence', *Philosophical Perspectives* 19, 95-119.
Froese, T., Gershenson, C., *&* Rosenblueth, D., A. (2013). 'The Dynamically Extended Mind', available at: http://arxiv.org/abs/1305.1958.
Fuller, S. (2012). 'Social Epistemology: A Quarter-Century Itinerary', *Social Epistemology: A Journal of Knowledge, Culture and Policy* 26, 267-83.
Gallagher, S. (2005). *How the Body Shapes the Mind*, Oxford: Oxford University Press.
Giere, R., *&* Moffat, B. (2003). 'Distributed Cognition: Where the Cognitive and the Social Merge', *Social Studies of Science* 33, 1-10.
Goldman, A. (1976). 'Discrimination and Perceptual Knowledge', *Journal of Philosophy* 73, 771-91.
——— (1979). 'What Is Justified Belief?', *Justification and Knowledge*, (ed.) G. Pappas, 1-23, Dordrecht: Reidel.
——— (1986). *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.
——— (1988). 'Strong and Weak Justification', *Philosophical Perspectives 2: Epistemology*, (ed.) J. Tomberlin, 51-69, Atascadero, CA: Ridgeview.
Greco, J. (1999). 'Agent Reliabilism', *Philosophical Perspectives* 13, 273-96.
——— (2004). 'Knowledge As Credit For True Belief', *Intellectual Virtue: Perspectives from Ethics and Epistemology*, (eds.) M. DePaul *&* L. Zagzebski, 111-34, Oxford: Oxford University Press.
——— (2007). 'The Nature of Ability and the Purpose of Knowledge', *Philosophical Issues* 17, 57-69.
——— (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*, Cambridge, UK: Cambridge University Press.
Haddock, A., Millar, A., *&* Pritchard, D. H. (eds) (2010). *Social Epistemology*, Oxford: Oxford University Press.
Hawthorne, J., *&* Stanley, J. (2008). 'Knowledge and Action', *Journal of Philosophy* 105, 571-90.
Heylighen, F., Heath, M., *&* Van Overwalle, F. (2007). 'The Emergence of Distributed Cognition: A Conceptual Framework', *Proceedings of Collective Intentionality IV*, Siena, Italy: University of Siena.
Hutchins, E. (1995). *Cognition in the Wild*, Cambridge, MA: MIT Press.
Kallestrup, J. (2011). *Semantic Externalism*, London: Routledge.
Kallestrup, J. *&* Pritchard, D. H. (2004). 'An Argument for the Inconsistency of Content Externalism and Epistemic Internalism', *Philosophia* 31, 345-54.
——— (2011). 'Virtue Epistemology and Epistemic Twin Earth', *European Journal of Philosophy*, DOI: 10.1111/j.1468-0378.2011.00495.x.
——— (2012). 'Robust Virtue Epistemology and Epistemic Anti-Individualism', *Pacific Philosophical Quarterly* 93, 84-103.
——— (2013). 'Robust Virtue Epistemology and Epistemic Dependence', *Knowledge, Virtue and Action*, (eds.) T. Henning *&* D. Schweikard, London: Routledge.
Kallestrup, J., *&* Sprevak, M. (*Forthcoming*). 'Entangling Externalisms', *New Waves in Philosophy of Mind*, (eds.) J. Kallestrup *&* M. Sprevak, London: Palgrave Macmillan.

Kiverstein, J., & Farina, M. (2012). 'Do Sensory Substitution Devices Extend the Conscious Mind?', *Consciousness in Interaction: The Role of the Natural and Social Context in Shaping Consciousness*, (ed.) F. Paglieri, The Netherlands: John Benjamins.

Latour, B. (1986). 'Visualization and Cognition: Thinking with Eyes and Hands', *Knowledge and Society* 6, 1-40.

Lehrer, K., & Cohen, S. (1983). 'Justification, Truth, and Coherence', *Synthese* 55, 191-207.

Littlejohn, C. (2009). 'The New Evil Demon Problem', *Internet Encyclopaedia of Philosophy*, (eds.) B. Dowden & J. Fieser, www.iep.utm.edu/evil-new/.

——— (2012). *Justification and the Truth-Connection*, Cambridge, UK: Cambridge University Press.

Luper, S. (1984). 'The Epistemic Predicament', *Australasian Journal of Philosophy* 62, 26-50.

——— (2003). 'Indiscernability Skepticism', *The Skeptics: Contemporary Essays*, (ed.) S. Luper, 183-202, Aldershot: Ashgate.

Lycan, (2001). 'The Case for Phenomenal Externalism', *Noûs* 35, 17-35.

Madison, B. J. C. (2009). 'On the Compatibility of Epistemic Internalism and Content Externalism', *Acta Analytica* 24, 173-183.

——— (2010). 'Epistemic Internalism', *Philosophy Compass* 5, 840-853.

——— (*Forthcoming*). 'Epistemological Disjunctivism and the New Evil Demon', *Acta Analytica*.

Menary, R. (2006). 'Attacking the Bounds of Cognition', *Philosophical Psychology* 19, 329-44.

——— (2007). *Cognitive Integration: Mind and Cognition Unbound*, London: Palgrave Macmillan.

McKinsey, M. (1991). 'Anti-Individualism and Privileged Access', *Analysis* 51, 9-16.

Neta, R., & Pritchard, D. H. (2007). 'McDowell and the New Evil Genius', *Philosophy and Phenomenological Research* 74, 381-96.

Nöe, A. (2003). 'Causation and Perception: The Puzzle Unravelled', *Analysis* 63, 93-100.

——— (2004). *Action in Perception*, Cambridge, MA: MIT Press.

Nozick, R. (1981). *Philosophical Explanations,* Oxford: Oxford University Press.

Olsson (2012). 'Coherentist Theories of Epistemic Justification', *Stanford Encyclopedia of Philosophy,* (ed.) E. N. Zalta, http://plato.stanford.edu/archives/spr2013/entries/justep-coherence.

Palermos, O. (2011). 'Belief-Forming Processes, Extended', *Review of Philosophy and Psychology* 2, 741-65.

——— (2014). 'Loops, Constitution, and Cognitive Extension', *Cognitive Systems Research,* 27: 25-41.

——— (*Forthcominga*). 'Could Reliability Naturally Imply Safety?', *European Journal of Philosophy,* DOI: 10.1111/ejop.12046.

——— (*Forthcomingb*). 'Knowledge and Cognitive Integration', *Synthese.* DOI: 10.1007/s11229-013-0383-02.

Palermos, O., & Pritchard, D. (*Forthcoming).* 'Extended Knowledge and Social Epistemology', *Social Epistemology.*

Pappas, G. (2005). 'Internalist *versus* Externalist Conceptions of Epistemic Justification', *Stanford Encyclopaedia of Philosophy*, (ed.) E. Zalta, http://plato.stanford.edu/entries/justep-intext/.

Plantinga, A. (1993). *Warrant and Proper Function*, New York: Oxford University Press.

Poston, T. (2008). 'Internalism and Externalism in Epistemology', *Internet Encyclopaedia of Philosophy*, (eds.) B. Dowden & J. Fieser, www.iep.utm.edu/int-ext/.

Pritchard, D. H. (2002*a*). 'McKinsey Paradoxes, Radical Scepticism, and the Transmission of Knowledge across Known Entailments', *Synthese* 130, 279-302.

——— (2002*b*). 'Recent Work on Radical Skepticism', *American Philosophical Quarterly* 39, 215-57.

——— (2002*c*). 'Resurrecting the Moorean Response to the Sceptic', *International Journal of Philosophical Studies* 10, 283-307.

——— (2005*a*). *Epistemic Luck*, Oxford: Oxford University Press.

——— (2005*b*). 'Scepticism, Epistemic Luck and Epistemic *Angst*', *Australasian Journal of Philosophy* 83, 185-206.
——— (2007*a*). 'Anti-Luck Epistemology', *Synthese* 158, 277-97.
——— (2007*b*). 'How to be a Neo-Moorean', *Internalism and Externalism in Semantics and Epistemology*, (ed.) S. Goldberg, 68-99, Oxford: Oxford University Press.
——— (2008*a*). 'McDowellian Neo-Mooreanism', *Disjunctivism: Perception, Action, Knowledge*, (eds.) A. Haddock & F. Macpherson, 283-310, Oxford: Oxford University Press.
——— (2008*b*). 'Sensitivity, Safety, and Anti-Luck Epistemology', *Oxford Handbook of Scepticism*, (ed.) J. Greco, 437-55, Oxford: Oxford University Press.
——— (2009*a*). 'Apt Performance and Epistemic Value', *Philosophical Studies* 143, 407-16.
——— (2009*b*). *Knowledge,* London: Palgrave Macmillan.
——— (2009*c*). 'Knowledge, Understanding and Epistemic Value', *Epistemology (Royal Institute of Philosophy Lectures)*, (ed.) A. O'Hear, 19-43, Cambridge UK: Cambridge University Press.
——— (2009*d*). 'Safety-Based Epistemology: Whither Now?', *Journal of Philosophical Research* 34, 33-45.
——— (2010*a*). 'Cognitive Ability and the Extended Cognition Thesis', *Synthese* 175, 133-51.
——— (2010*b*). 'Relevant Alternatives, Perceptual Knowledge and Discrimination', *Noûs* 44, 245-68.
——— (2011). 'Evidentialism, Internalism, Disjunctivism', *Evidentialism and its Discontents*, (ed.) T. Dougherty, 362-92,
——— (2012*a*). 'Anti-Luck Virtue Epistemology', *Journal of Philosophy* 109, 247-79.
——— (2012*b*). *Epistemological Disjunctivism*, Oxford: Oxford University Press.
——— (2012*c*). 'In Defence of Modest Anti-Luck Epistemology', *The Sensitivity Principle in Epistemology*, (eds.) T. Black & K. Becker, 173-92, Cambridge UK: Cambridge University Press.
——— (2013). 'There Cannot be Lucky Knowledge', *Contemporary Debates in Epistemology* (2nd Ed.), (eds.) M. Steup, J. Turri & E. Sosa, §7, Oxford: Blackwell.
——— (*Forthcoming*). 'The Modal Account of Luck', *Metaphilosophy*.
Pritchard, D. H., Millar, A., & Haddock, A. (2010). *The Nature and Value of Knowledge: Three Investigations*, Oxford: Oxford University Press.
Pritchard, D. H., & Smith, M. (2004). 'The Psychology and Philosophy of Luck', *New Ideas in Psychology* 22, 1-28.
Pryor, J. (2001). 'Highlights of Recent Epistemology', *British Journal for the Philosophy of Science* 52, 95-124.
Putnam, H. (1975). 'The Meaning of "Meaning"', *Language, Mind and Knowledge*, (ed.) K. Gunderson, 131-93, Minneapolis, IND: University of Minnesota Press.
Roush, S. (2005). *Tracking Truth: Knowledge, Evidence and Science*, Oxford: Oxford University Press.
Rupert, D. R. (2004). 'Challenges to the Hypothesis of Extended Cognition', *Journal of Philosophy* 101, 389-428.
——— (2009). *Cognitive Systems and the Extended Mind*, Oxford: Oxford University Press.
Sainsbury, R. M. (1997). 'Easy Possibilities', *Philosophy and Phenomenological Research* 57, 907-19.
Sosa, E. (1988). 'Beyond Skepticism, to the Best of our Knowledge', *Mind* 97, 153-88.
——— (1991). *Knowledge in Perspective: Selected Essays in Epistemology*, Cambridge: Cambridge University Press.
——— (1999). 'How to Defeat Opposition to Moore', *Philosophical Perspectives* 13, 141-54.
——— (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford: Oxford University Press.
——— (2009). *Reflective Knowledge: Apt Belief and Reflective Knowledge, Volume II*. Oxford: Oxford University Press.

Steup, M. (1999). 'A Defense of Internalism', *The Theory of Knowledge: Classical and Contemporary Readings* (3rd Ed.), (ed.) L. Pojman, 310-21, Belmont, CA: Wadsworth.

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.

Theiner, G. (2011). *Res Cogitans Extensa: A Philosophical Defense of the Extended Mind Thesis*, Bern, Switzerland: Peter Lang GmbH, Europaischer Verlag der Wissenschaften.

Theiner, G., Allen, C., & Goldstone, R. (2010). 'Recognizing Group Cognition', *Cognitive Systems Research* 11, 378-95.

Tollefsen, D., & Dale, R. (2011). 'Naturalizing Joint action: A Process-Based Approach', *Philosophical Psychology* 25, 385-407.

Turri, J. (2011). 'Manifest Failure: The Gettier Problem Solved', *Philosophers' Imprint* 11, available at: http://quod.lib.umich.edu/cgi/p/pod/dod-idx?c=phimp;idno=3521354.0011.008.

Vahid, H. (2003). 'Content Externalism and the Internalism/Externalism Debate in Justification Theory', *European Journal of Philosophy* 11, 89-107.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.

Vogel, J. (2000). 'Reliabilism Leveled', *Journal of Philosophy* 97, 602-23.

——— (2007). 'Subjunctivitis', *Philosophical Studies* 134, 73-88.

Wegner, M., Giuliano, T., & Hertel, P. (1985). 'Cognitive Interdependence in Close Relationships', *Compatible and Incompatible Relationships*, (ed.) W. J. Ickes, 253-76, New York: Springer-Verlag.

Wheeler, M. (2005). *Reconstructing the Cognitive World*, Cambridge, MA: MIT Press.

Williams, M. (1991). *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism*, Oxford: Blackwell.

Williamson, T. (2000). *Knowledge and its Limits*, Oxford: Oxford University Press.

——— (2007). 'On Being Justified in One's Head', *Rationality and the Good*, (eds.) M. Timmons, J. Greco & A. Mele, 106–122, Oxford: Oxford University Press.

Wilson, R. A. (2000). 'The Mind Beyond Itself', *Metarepresentations: A Multidisciplinary Perspective*, (ed.) D. Sperber, 31-52, New York: Oxford: University Press.

——— (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition*, New York: Cambridge University Press.

NOTES

[1] Note that we will be exploring the epistemic externalism/internalism distinction in its own right below, and in the process considering alternative formulations of this distinction.

[2] For two influential defences of accessibilism, see Chisholm (1977) and Bonjour (1985, ch. 2). For some helpful recent discussions of accessibilism and its role in the wider epistemological externalism/internalism debate, see Steup (1999), Pryor (2001, §3), Bonjour (2002), Pappas (2005), and Poston (2008).

[3] The *locus classicus* for defences of mentalism is Conee & Feldman (2004). For a set of critical discussions of this proposal, see Dougherty (2011). For a recent critical discussion of the relative merits of mentalism and accessibilism in the context of epistemological disjunctivism, see the exchange between Pritchard (2011*c*) and Conee & Feldman (2011).

[4] The *locus classicus* for defences of process reliabilism is Goldman (1979; cf. Goldman 1986). We shall return to process reliabilism in §3.3.2.

[5] The *locus classicus* for discussions of epistemic deontology is Alston (1986). See also Pryor (2001, §4) and Adler (2002).

[6] Conee & Feldman (2004) is a case in point.

[7] The clairvoyance case is due to Bonjour (1985). For the seminal discussion of the new evil demon cases, see Lehrer & Cohen (1983; cf. Cohen 1984). There is a wealth of literature on new evil demon cases. For some of the key texts and recent discussions of this topic, see Bach (1985), Goldman (1988), Engel (1992), Bonjour & Sosa (2003, 159-61), Neta & Pritchard (2007), Pritchard (2011; 2012*b*), Littlejohn (2012), and Madison (*forthcoming*). For an excellent survey of recent work on the new evil demon, see Littlejohn (2009).

[8] Note that a complication we are setting to one side for the purposes of this paper is the relevance of *epistemological disjunctivism* to the epistemic externalism/internalism debate. While this view accepts accessibilism and is arguably compatible with mentalism (the view as it is usually understood also accepts some sort of deontic requirement on justification), it rejects the internalist reading of the new evil demon case—*viz*., that the two subjects enjoy identical levels of justification. See Pritchard (2008*a*; 2011; 2012*b*). See also Neta & Pritchard (2007).

[9] This distinction between intervening and environmental epistemic luck is due to Pritchard (2009*a*; 2009*b*; 2009*c*; 2012*a*; cf. Pritchard, Millar & Haddock 2010, chs. 1-4). For further discussion of the phenomenon of epistemic luck, see Pritchard (2005*a*). For further discussion of the notion of luck more generally, see Pritchard & Smith (2004) and Pritchard (2005*a*, ch, 5; *forthcoming*).

[10] This example is due to Goldman (1976), who in turn credits it to Carl Ginet.

[11] See Vogel (2000; 2007).

[12] See Vogel (2000; 2007). We shall return to sensitivity- and safety-based accounts of knowledge in §3.3.1.

[13] Although one can think of sensitivity- and safety-based proposals as being a way of developing reliabilism, we should note that they are typically motivated entirely independently. For more on sensitivity, see Dretske (1970; 1971), Nozick (1981), Roush (2005), Becker (2007), Black & Murphy (2007), and Black (2002; 2008). Versions of safety have been defended by Luper (1984; cf. Luper 2006), Sainsbury (1997), Sosa (1999), Williamson (2000), and Pritchard (2002*b*; 2005*a*; 2007*a*; 2009*d*; 2012*a*; 2012*c*). For a critical overview of the various issues related to sensitivity- and safety-based epistemology, see Pritchard (2008*b*).

[14] See Nozick (1981) for the classic statement of the idea that one's beliefs in the denials of sceptical hypotheses are insensitive. Note that this claim has been disputed. See Williams (1991, ch. 8) and Black (2002; 2008).

[15] For more on the Moorean safety-based response to radical scepticism, see Sosa (1999) and Pritchard (2002*b*; 2005*a*; 2005*b*; 2007*b*). For a survey of recent work on the topic of radical scepticism, which covers both Moorean and non-Moorean approaches to the problem, see Pritchard (2002*a*).

[16] Note that another very important formulation of virtue reliabilism is agent reliabilism (Greco 1999, 2010). We focus on this view in §3.3.3 because of its aptness for an interpretation along the lines suggested by the hypotheses of extended and distributed cognition.

[17] See Kallestrup & Pritchard (2011) for an argument to the effect that no plausible relativisation of cognitive abilities to an environment could exclude environmental epistemic luck. See also Kallestrup & Pritchard (2012; 2013).

[18] See especially Hawthorne & Stanley (2008).

[19] This interdependence is usually cast out in terms of "sensorimotor dependencies." Consider the following passages:

> "The basic claim of the enactive approach is that the perceiver's ability to perceive is constituted (in part) by sensorimotor knowledge (i.e., by practical grasp of the way sensory stimulation varies as the perceiver moves)." (Noë 2004, 12)
>
> "[*Perception*] is not a process in the brain, but a kind of skillful activity on the part of the animal as a whole." (Noë 2004, 2)
>
> "Perception is not something that happens to us or in us, it is something we do." (Noë 2004, 1)

Sensorimotor dependencies are relations between movements or change and sensory stimulation. It is the practical knowledge of loops relating external objects and their properties with recurring patterns of change in sensory stimulation. These patterns of change may be caused by the moving subject, the moving object, the ambient environment (e.g., changes in illumination), and so on. For more recent elaborations of the idea, see Hurley & Noë (2003), Noë (2003; 2004), Gallagher (2005), and Chemero (2009).

[20] Which is not to say, of course, that the view is without its critics. The discussion of the objections facing active externalism is well beyond the scope of the present paper, however. Nevertheless, in brief, many of them point either to the dissimilarity between the inner cognitive processes and the external elements that are supposed to be parts of one's cognitive system (e.g., Rupert 2004; Adams & Aizawa 2008), or to the perceptive rather than introspective manipulation of those external elements. Others deny the mark of the cognitive to the alleged extended cognitive processes (e.g., Adams & Aizawa 2008), or claim that there cannot be a science of active externalism (e.g., Rupert 2004; Adams & Aizawa 2008). For a short discussion and reply to most of these objections, see Menary (2006).

[21] Rupert (2004, 397), Clark (2008, 78) and others have explicitly assumed that active externalism and content externalism are independent doctrines, but Sprevak & Kallestrup (*forthcoming*) challenge this received view by arguing that this independence claim is not obviously true. The relationship between those two views is much more complex than previously suspected.

[22] While this paper was first published in 2010, it has been available online since 2006. The 'glue and trust' criteria, however, had already made their appearance in Clark & Chalmers (1998), although the phrasing was somewhat different.

[23] In Clark & Chalmers (1998, 17) the authors consider a further criterion:

> "Fourth, the information in the notebook has been consciously endorsed at some point in the past, and indeed is there as a consequence of this endorsement."

As the authors further note, however, "the status of the fourth feature as a criterion for belief is arguable (perhaps one can acquire beliefs through subliminal perception, or through memory tampering?)", so they subsequently drop the said criterion.

[24] This does not mean that the hypotheses of extended and distributed cognition are incompatible with common-sense functionalism, or that they are anti-functionalist on the whole. Insofar as a cognitive process is a function, these two hypotheses are compatible with functionalism. For more details see Palermos (2014).

[25] As we noted in §1, formulated in terms of *knowledge*, epistemic internalists insist that internalist justification is necessary for knowledge. Accordingly, if content externalism is incompatible with internalism about justification, then it will be so as well with an epistemic internalist's account of knowledge. Thus, it makes sense to assess the *compatibility* of content externalism and epistemic internalism with reference to the internalist thesis *vis-à-vis* justification. This is how we will be proceeding.

[26] Note that some versions of the accessibilist position are characterised in terms of 'awareness' (e.g., Bergmann 2006). See Madison (2010) for an overview.

[27] Correlatively, a commitment of mentalism is that agents do not diverge in the extent to which their beliefs are justified insofar as they have the same mental states.

[28] Or, *vice versa*, as these are logically independent theses. However, most all internalists will endorse the accessibilist thesis, and in this respect, it is the least controversial formulation.

[29] See Pritchard (2011) for a mapping of different varieties of epistemic externalism as functions of different combinations of epistemic internalist theses that can be denied. (Note the discussion here of a further commitment of classical epistemic internalism, which Pritchard calls the 'discriminability' thesis).

[30] See here Vahid (2003, §1) for some helpful discussion here. We follow Vahid in referring to the position that content externalism is incompatible with epistemic internalism as 'the incompatibilist thesis.'

[31] Bonjour remarks that:

> "The adoption of an externalist account of mental content would seem to support an externalist account of justification in the following way: if part or all of the content of a belief is inaccessible to the believer, then both the justifying status of other beliefs in relation to that content and the status of that content as justifying further beliefs will be similarly inaccessible, thus contradicting the internalist requirement for justification." (Bonjour 1992, 136)

On this line of thinking, the content of my belief 'X is F" where X is an externally individuated (i.e. wide) content is not reflectively accessible in such a way that 'X is F" can never, by the standards of J-Internalism, be something that justifies any other belief I have.

[32] Note that the variables have been amended from Chase's original presentation for ease of exposition. See Chase (2001, 327) for a slightly different way of extracting Bonjour's argument.

[33] Note that Williamson (2007, 107-108) offers a similarly straightforward argument for the incompatibility of content externalism and epistemic internalism construed along *mentalist* lines, with the operative point being that Oscar and Twin Oscar can differ in their justified beliefs (given content externalism) despite being internal duplicates, a difference that Williamson takes to be incompatible with the mentalist claim that internal duplicates are

justificational duplicates. It is interesting to note that Madison (2009, §3, esp. 180-2) objects to Williamson's argument for the incompatibility of content externalism with mentalist J-internalism for essentially the same reasons (as we show) Chase sites in objecting to Bonjour's argument for the incompatibility of content externalism with *accessibilist* J-internalism.

[34] Note that Madison (2009) engages primarily with an incompatibilist thesis submitted by Williamson (2007)—see endnote 33.

[35] Recall that the intuition in new evil demon cases is meant to be that one agent knows if her internal duplicate does. But of course the epistemic externalist (e.g., the reliabilist) seems committed to supposing that the deceived agent's beliefs (because unreliable, say) are not justified. Therefore, to the extent that the intuition that both internal duplicates are equally justified is right, epistemic externalism is in trouble. See §1 for more discussion on this point.

[36] Compare here with Madison's (2009, 182) point that the internally grounds relevant to epistemic J-Internalism's accessibility requirement are 'what is necessarily in common between twins in Twin Earth scenarios'—viz., narrow content that is 'subjectively indistinguishable' in the two cases. It is a difficult philosophical question how to articulate with precision what is subjectively indistinguishable and 'necessarily in common' in these cases, in part because, on some views—e.g., Dretske (1996)—what is subjectively indistinguishable is also widely individuated, and so would not be 'necessarily in common' between the twins in Twin Earth scenarios. See here, along with Dretske (1996), Lycan (2001) for the view that *qualia* are wide, where qualia in Lycan's sense are 'the introspectible qualitative phenomenal features that characteristically inhere in sensory experiences.' (Lycan's 2001, 18) See also Block's (1990) 'Inverted Earth' argument.

[37] See Vahid (2003, 101-3) for a criticism of Chase's critique of Bonjour's argument.

[38] This is how Boghossian (1989, 5) summarises his overarching position, which is (as Chase notes) primarily aimed at rejecting the claim that we can know *a priori* the contents of our thoughts. His self-knowledge argument, which generates the incompatibility thesis, is thus in the service of a slightly different dialectical goal. It is important to note that Boghossian's overarching objective—a claim about whether according to content externalism we can know *a priori* the content of our own thoughts—can be located as part of a debate that (following McKinsey 1991) is framed in terms of the compatibility of content externalism and the *privileged access thesis*. As Kallestrup (2011, 157) puts it, this is the thesis that a thinker *S* has *a priori* access to the propositional contents of her occurrent thoughts, where by '*a priori*' we stipulate the inclusion of knowledge based on introspective deliverances. Along with Kallestrup (2011), see also Pritchard (2002) for further discussion.

[39] Cited also in Boghossian (1989/2008, 153-154).

[40] See, for instance, Burge (1988). See §2 for further discussion. *Cf.* Ludlow (1995) for an argument to the effect that the phenomenon of slow-switching is more prevalent than far-fetched thought experiments suggest.

[41] Boghossian (1989/2008, 157-64). See also Chase (2001, 233-34) for a summary. But see also Kallestrup & Pritchard (2004, 351-54) for a variant of this kind of argument.

[42] See Olsson (2012, §2) for a clear presentation of the general problem.

[43] See here Boghossian (1989/2008, 154). Boghossian is aware of the worry and insists that his regress is not just an instance of the general regress. Chase is not convinced:

> "Boghossian's argument for P1 does involve a more complicated set of generating analyses than the standard argument, but it also retains the essential structural features of the regress argument…the class limitation isn't doing any particular work here other than restricting talk to a subset of our general beliefs; certainly it does no work that would prevent a foundationalist, or indeed a coherentist reply." (Chase 2001, 232)

[44] On this point, see Chase (2001, 232) and Vahid (2003, 92-93). Vahid actually (*Op. cit.*, 93) claims that the regress argument does not show '*either* that all or some self-knowledge is non-inferential.' (Vahid 2003, 93, our italics)

[45] This line could be pursued either by challenging relevant alternatives accounts of knowledge or by challenging, more specifically, the claim that the belief in question in the example case constitutes a 'relevant' alternative.

[46] As Vahid notes:

> "The activity of reasoning or inferring is not essential to the relevant alternatives theory which is actually characterized by the satisfaction of certain counterfactuals." (Vahid 2003, 97)

Vahid here references Goldman's (1986, 46) remark that a 'true belief (*p*) fails to be knowledge if there are any relevant alternative situations in which the proposition *p* would be false, but the process used would cause *S* to believe *p* anyway'.

[47] Chase articulates mentalism, which he calls *J-Int*$_2$ as follows:

> "For all agents a_1 and a_2 and worlds w_1 and w_2, if a_1 in w_1 and a_2 in w_2 are identical in the internal physical constitutions on which their minds supervene, then a_1 and a_2 are identical in all respects relevant to the justification of their beliefs." (Chase 2001, 239)

[48] We paraphrase the argument as Chase (2001, 241) presents it. In particular, we are using "mentalist J-internalism" where Chase is using his favoured articulation of mentalist J-internalism which (as we previously noted in endnote 45) he calls J-Int$_2$.

[49] Chase (2001, 241-242). See also Brueckner (2002, 513-514) for a presentation of Chase's argument for (2).

[50] Chase's (2001, 241-242) remarks here are rather quick. He says, 'If you believe in propositions, then you could say that Sue and Twin Sue are now expressing the same proposition; the beliefs they are reporting have the same content, they are the same type belief (although of course not the same belief token). Moreover, as in all Twin Earth cases, Sue and Twin Sue are identical in internal physical constitution. Hence, by (PC) the justificatory processes Sue has used to obtain the belief that it is possible to drink a liquid is identical to the justificatory process Twin Sue has used.'

[51] Sue is reasoning through premises (e.g., 1 and 2) about *water* while Twin Sue is reasoning through premises about *twater*.

[52] See Chase (2001, 242).

[53] One issue of which to be mindful throughout is the sense in which the philosophical dispute here turns on the distinction between subjective and objective justification.

[54] A brief remark: whereas the topic of the compatibility of *passive* content externalism and epistemic internalism has (as we saw in §3.1) some philosophical precedent in the literature, the same is not so for the compatibility of versions of the (comparatively more recent) active externalism and epistemic internalism. We will thus examine what we take to be the most natural arguments one might make for the *incompatibility* of EC and epistemic internalism (beginning with accessibilist J-internalism), and see whether they hold water.

[55] Compare here Clark's (2011) 'glue and trust' conditions.

[56] Of course, it would be in *virtue of* satisfying Clark's (2011) 'glue and trust' conditions (and not in virtue of satisfying the parity principle) that we would offer an explanation for why some part of the world is appropriately integrated in Otto's cognitive system. The idea here is just that, by reference to the parity principle, we can suppose that the relevant integration conditions will be satisfied; hence, with reference to the parity principle, a part of the world that counts as part of a cognitive process should (on Clark's view) *de facto* satisfy his independently specified integration conditions.

[57] One might attempt to say that by reflection what is meant is a kind of non-extended introspection. But a proponent of extended cognition will then make the same point, that if introspection is a cognitive process, there is no in principle barrier to extended introspection.

[58] Again, even though Otto* has the correct entry written in his *notebook*, he does not have the correct entry in memory storage, because even on EC, Otto*'s notebook is not part of his memory, as it is not appropriately integrated into his cognitive system.

[59] It's worth noting that it's not *obvious* that the justification in question would increase in a non-negligible way. Consider after all, a normal case where Finn remember that there is a fence in the yard, and then, at T, walk into the yard to perceive what his memory clearly confirmed, that there is a fence in the yard. Now, contrast Finn with his duplicate, Finn* who is like Finn except that Finn* has never been in the yard before T, and so has no memories of the fence being in the yard. Nonetheless, at T, Finn* in broad daylight, perceives that there is a fence in the yard, by looking right at one. Is Finn justified to a greatest extent in believing there is a fence in the yard than Finn*? It's not obvious. But of course, the line of argument under consideration for (3) of the argument relies in such reasoning.

[60] It should be pointed out that the line of argument suggested here does *not* rest on the suggestion that the extended mind thesis could be rendered compatible with epistemic internalism, *traditionally construed*. Rather, the extended mind proponent is insisting here that *if the extended mind thesis* is true, then both versions of epistemic internalism (construed in a way that allows for extended states) is obviously compatible with the extended mind thesis. The positions are thus not necessarily incompatible. But, again, this compatibility is preserved only by drawing attention to the fact that the versions of epistemic internalism compatible with the extended mind thesis endorse extended states in such a way that will render the position nearly unrecognizable to traditional epistemic externalists.

[61] Clark (2009) denies that consciousness extends beyond the brain. For a recent argument to the opposite direction see Kiverstein & Farina (2012).

[62] One notable reason why sensitivity principles are not popular is that they appear to require that one be willing to part ways with the intuitive closure principle on knowledge: if *S* knows *p*, and *S* competently deduces *q* from *p*, thereby forming a belief that *q* on this basis and while retaining knowledge that *p*, then *S* knows that *q*. See Pritchard (2008*b*) for discussion of this issue.

[63] Of course, if *safety* is a necessary condition on knowledge, then notice that (unlike on standard process reliabilism) the target belief in the barn façade case does not qualify as knowledge. After all, holding fixed that one forms a belief about a barn by pointing at what looks like a barn, in an environment where there are many barn facades, *S* believes falsely in many near-by worlds.

[64] For more on anti-luck epistemology, see Pritchard (2005*a*; 2007*a*; 2012*c*). See also Pritchard (2008*b*; 2009*d*; 2013).

[65] Pritchard (2012*a*) suggests that 'robust' versions of the safety approach, according to which safety is both necessary and sufficient for knowledge, fail to adequately capture the sense in which knowledge should be said to arise from ability.

[66] This example is taken from Carter (*forthcoming*).

[67] Notice, however, that in order to remain fast to externalism, reliabilism must find a way to satisfy a condition of 'subjective sensitivity to the reliability of one's evidence' in a way that will not require *knowledge of or even beliefs about* the said reliability.

[68] The ability intuition on knowledge was, initially at least, often introduced in order to do away with knowledge undermining epistemic luck:

> "To say that someone knows is to say that his believing the truth can be credited to him. It is to say that this person got things right due to his own abilities, efforts and actions, rather than due to dumb luck, or blind chance, or something else." (Greco 2004, 111)

[69] The fact that people manifest highly specific, finely tuned dispositions to form their beliefs in certain ways but not in others amounts to an implicit awareness of the reliability of those dispositions.

> "For example suppose that it seems visually to a person that a cat is sleeping on the couch, and on this basis she believes that there is a sleeping cat on the couch. Suppose also that this belief manifests a disposition that the person has, to trust this sort of experience under these sorts of conditions, when motivated to believe the truth. Now, suppose that much less clearly, it seems visually to the person that a mouse has run across the floor. Not being disposed to trust this kind of fleeting experience, the person refrains from believing until further evidence comes in. The fact that the person, properly motivated, is disposed to trust one kind of experience but not the other, constitutes sensitivity on her part that the former is reliable. There is a clear sense in which she takes the former experience to be adequate to her goal of believing the truth, and takes the latter experience not to be. And this is so even if she has no beliefs about her goals, her reliability, or her experience." (Greco 1999, 290)

A similar argument can be found in Sosa (1991, 60-63).

[70] For a detailed defense of the claim that Clark (2010) and Greco (2010) put forward the same criteria for a process to count as part of one's cognitive system/character, see Palermos (2011; *forthcominglı*).

[71] For more details on cognitive integration and its effects on the process of knowledge acquisition see (Palermos *forthcominglı*)

[72] See Bach-y-Rita & Kercel (2003) for a recent review on tactile visual substitution systems.

[73] For an objective criterion of constitution according to dynamical systems theory, and how it applies to cognition see Palermos (2014).

[74] Remember that according to virtue reliabilism knowledge must be belief that is true in virtue of cognitive ability, where, according to Greco, "in virtue of" must be understood in causal explanatory terms. Even though several proponents of virtue reliabilism agree on this general causal-explanatory understanding of the view, there is disagreement on whether the relevant cognitive ability should be the "most salient" (Greco 2010) or merely a "significant" (Pritchard 2010*a*) factor in the causal explanation of how the agent acquired his true belief.

[75] For a more detailed explanation of how virtue reliabilism may be applied to epistemic group agents see Palermos & Pritchard (2013).

[76] This paper was produced as part of the AHRC-funded 'Extended Knowledge' research project (AH/J011908/1), which is hosted at Edinburgh's *Eidyn* Research Centre. The authors are grateful to the AHRC for its support of this project.