



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Belief-Forming Processes, Extended

Citation for published version:

Palermos, SO 2011, 'Belief-Forming Processes, Extended' Review of Philosophy and Psychology, vol. 2, no. 4, pp. 741-765. DOI: 10.1007/s13164-011-0075-y

Digital Object Identifier (DOI):

[10.1007/s13164-011-0075-y](https://doi.org/10.1007/s13164-011-0075-y)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Review of Philosophy and Psychology

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



BELIEF-FORMING PROCESSES, EXTENDED

S.Orestis Palermos

University of Edinburgh

ABSTRACT. We very often grant that a person can gain knowledge on the basis of epistemic artifacts such as telescopes, microscopes and so on. However, this intuition threatens to undermine virtue reliabilism according to which one knows that p if and only if one's believing the truth that p is the product of a reliable cognitive belief-forming process; in an obvious sense epistemic artifacts are not parts of one's overall cognitive system. This is so, unless the extended cognition hypothesis (HEC) is true. According to HEC when parts of the environment become properly coupled to the agent's brain then they too can be considered constitutive parts of the overall cognitive mechanism—i.e. cognition potentially extends to the world surrounding the agent. Interestingly, HEC and the broader framework of virtue reliabilism share some intriguing similarities, which render these two views mutually supportive. Making these similarities explicit provides a principled account of the way in which our knowledge-conducive cognitive characters may extend beyond our natural cognitive capacities by incorporating epistemic artifacts.

Introduction

Think about Travis. He is in a room. He moves his head around and through his practical understanding of sensorimotor contingencies he perceives a chair.¹ He walks confidently towards the chair and sits down. Travis, however, is blind and has been equipped with a Tactile Visual Substitution System (TVSS) for the past five years.² Thus, Travis is well trained and accustomed to perceiving shapes and objects that are recorded by a mini video camera attached to his sunglasses that then converts the images to tactile stimulation on his tongue. Subjects like Travis are not only able to perceive shapes and objects and orient themselves, but they have also occasionally

¹ For a full account of how sensorimotor knowledge is constitutive of perception see (Noë 2004). “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). “What the perception is, however, 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 (changes in illumination) and so on.

² See Bach-y-Rita and Kercel (2003) for a recent review on TVSS.

offered reports of feeling as if they were *seeing* objects, that is as if they were enjoying the phenomenal qualities of the original sense modality that is being substituted.

So the question is: does Travis *know* there is a chair in front of him? It seems our intuitions should be on his side. However, this and similar intuitions according to which we may grant that a person can gain knowledge on the basis of epistemic artifacts threaten to undermine a prominent trend within contemporary epistemology, that of virtue reliabilism. Briefly, according to virtue reliabilism, one knows that *p* only if one's believing the truth that *p* is the product of a reliable cognitive belief-forming process. But, in an obvious sense, epistemic artifacts are not part of one's overall cognitive system.

This is so, unless the hypothesis of the extended cognition (HEC) is true. According to HEC, not all cognitive processes are entirely realized inside the skin of the cognizing organism. When parts of the environment become properly coupled to the agent's brain, then they too can be considered as constitutive parts of the overall cognitive mechanism—i.e. cognition potentially extends to the world surrounding the agent.

That contemporary philosophy of mind may become highly pertinent to epistemology should not come as a surprise. After all, the possession and acquisition of knowledge is a highly cognitive phenomenon. Remarkably, despite the growing appeal of HEC within philosophy of mind, there have been almost no attempts in the literature to investigate its epistemological ramifications, and yet clearly such a view may have dramatic implications for epistemology in terms of how it radically changes our conception of the inherently cognitive processes of knowledge acquisition.

Duncan Pritchard (2010c), however, has recently argued that a virtue reliabilistic necessary condition on knowledge fits 'snugly' with HEC. The aim of the present paper is not only to show that HEC is necessary for virtue reliabilism but, also, that these two views share some intriguing similarities which render them mutually supportive. Furthermore, making those similarities explicit will provide us with a principled account of the way in which our knowledge-conducive cognitive characters may extend beyond our natural cognitive capacities by incorporating epistemic artifacts.

In order to become familiar with virtue reliabilism's compatibility with HEC we need to pay close attention to the fundamental tenets of both views. In the first and

second section, I discuss virtue reliabilism and HEC, respectively. In the third section, I perform a comparative analysis of the two views, which is followed by the consideration of several thought experiments in section four.

i) The Ability Intuition on Knowledge

The guiding idea of robust virtue reliabilism is that knowledge is the product of cognitive ability. Call this the ability intuition on knowledge.³ As Greco (2004, 111) explains, “to say that someone knows is to say that his believing the truth can be credited to him. It is to say that the person got things right due to his own abilities, efforts and actions, rather than due to dumb luck, or blind chance, or something else”.

Usually, robust virtue reliabilism is cashed out as a refinement of process reliabilism according to which knowledge is the product of a reliable belief-forming process, where a reliable belief-forming process is one that generally tends to result in true rather than false beliefs. Process reliabilism is an externalist approach to knowledge in that it “denies that evidential relations must be necessary and denies that one must know that one’s evidence is reliable” (Greco 1999, 284-5); if forming a belief on a certain kind of evidence constitutes a reliable belief-forming process, it does not matter that the evidence is only contingently reliable. So, contrary to the traditional account of knowledge as internally justified true belief, one needs not know by *reflection alone* that one’s beliefs are formed in a reliable fashion. Process reliabilism makes *de facto* reliability the grounds of positive epistemic status.

However, as Greco notes (1999, 285), “it is not enough that one’s belief is formed in a way that is objectively reliable; one’s belief must be formed in a way that is subjectively appropriate as well. Nevertheless, in order to remain fast to externalism Greco claims that subjective justification must be accommodated in a way that does not involve knowledge of reliability, or even beliefs about reliability. Accordingly, he proposes (1999, 289) 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

³ Hereafter, ‘virtue reliabilism’ will signify the view that the ability intuition on knowledge is only a necessary condition on knowledge whereas ‘*robust* virtue reliabilism’ is the view that the satisfaction of the ability intuition is both necessary and sufficient for knowledge. The idea that knowledge must be grounded in cognitive abilities can be traced back to the writings of Ernest Sosa (1988; 1993) and Alvin Plantinga (1993). For more recent advocacies see Greco (1999; 2004; 2007) and Pritchard (2009; 2010a; 2010b; 2010c). Pritchard, however, holds that the ability intuition is not the only guiding epistemological intuition and he thus avoids to formulate a version of robust virtue reliabilism. Instead he proposes a kind of anti-luck virtue reliabilism. See (Pritchard 2010a, 2010b). See also ft. 11.

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 addition, Greco suggests, the dispositions/habits that a person manifests when she is thinking conscientiously are the stable properties of her *cognitive character*. So, not all reliable belief-forming processes are knowledge conducive; rather “it is those processes that have their bases in the stable and successful dispositions of the believer that are relevant for knowledge and justification” (Greco 1999, 287). “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). In other words, one is sensitive to the reliability of one’s beliefs merely by employing one’s cognitive dispositions/habits for formulating them; so long as a reliable belief-forming process is one of the dispositions/habits that make up one’s cognitive character, it can thereby be implicitly regarded as trustworthy.

Robust virtue reliabilism is thus a refinement of process reliabilism in that it accommodates the ability intuition on knowledge. In order for a belief to be both subjectively and objectively justified it’s not enough that it is the product of a reliable belief-forming process; it must be the outcome of the agent’s stable and successful *cognitive* belief-forming processes that make up his/her cognitive character.⁴ On this view, one’s cognitive character consists of one’s cognitive faculties of the brain/central nervous system (CNS) including one’s natural perceptual cognitive faculties, one’s memories and the overall doxastic system. In addition, it can also consist of acquired habits of thought, “acquired skills of perception and acquired methods of inquiry, including those involving highly specialized training or even advanced technology” (1999, 287). Accordingly, robust virtue reliabilism is usually formulated as follows:⁵

Robust Virtue Reliabilism

⁴ Another reason why Greco appeals to the notion of one’s cognitive character is his attempt to do away with the problem of the strange and fleeting processes. See (1999, 286-9).

⁵ Notice that Greco calls his view ‘agent reliabilism’. I have here preferred the alternative name ‘robust virtue reliabilism’ for two reasons. First, Greco himself does not explicitly endorse such a strong formulation of the ability intuition on knowledge. Instead, Greco holds that *S* knows that *p* if and only if *S*’s reliable cognitive character is an important—but not necessarily the most important—necessary part of the total set of causal factors that give rise to *S*’s believing the *truth* regarding *p* (see Greco 1999, 287-8; 2004, 123; 2010,12). (However, his critics argue that he implicitly endorses a strong version of robust virtue reliabilism along the lines of the above formulation. See (Lackey 2008) and (Vaesen 2010). See also (Kelp, forthcoming)). Second, I wanted to make explicit that this approach falls under the broader trend of virtue epistemology, “since the stable and successful dispositions of a person are appropriately understood as virtues” (Greco 1999, 287).

S knows that *p* if and only if *S*'s reliable cognitive character is the most important necessary part of the total set of causal factors that give rise to *S*'s believing the truth regarding *p*.

As far as the ability intuition on knowledge is concerned, the thinking behind robust virtue reliabilism is this: if *S*'s true belief that *p* is the product of some reliable cognitive ability, then we may also safely conclude that *S*'s cognitive character figures most importantly in the causal explanation of how *S* came to believe the truth regarding *p*, and thus that *S* knows that *p*.

The reason why robust virtue epistemologists are inclined towards such a strong virtue-theoretic account of knowledge is their attempt to do away with the knowledge-undermining epistemic luck involved in Gettier cases.⁶ As Gettier demonstrated, one's justified belief may turn out to be true without thereby counting as an instance of knowledge. In the typical scenario, one's belief, which is the product of faulty reasoning, *just happens* to be true for reasons that are extraneous to one's justification. Or again, one may come to believe the truth on the basis of a lucky guess. Contrast this with cases of success through ability. "There is a sense of "luck" on which lucky success is precisely opposed to success through virtue or ability" (Greco 2007, 58). When one's success is the product of one's ability then clearly one's success cannot have been a lucky one. Accordingly, robust virtue epistemologists hold that when one knows, one's intellectual success is the product of a reliable cognitive ability.⁷ In other words, they claim, the cognitive success must be primarily creditable to one's cognitive character. (Alternatively: one's cognitive character must be the most salient factor in the causal explanation of how one acquired one's true belief).

While robust virtue reliabilism fares well with respect to a wide variety of thought experiments, including standard Gettier cases, it is nonetheless not free of problems. Let us first consider two thought experiments.

⁶ See (Gettier 1963). As an anonymous referee has noted, familiar Gettier cases can be addressed by process reliabilism, so why the need for the virtue theoretic element? There are, however, Gettier cases such as the Temp case (see section four) where while one's way of forming beliefs is reliable, it is not responsive to the facts, but the other way around. If, however, one's belief-forming process is a cognitive ability, then one's beliefs are formulated in a way that will guarantee their responsiveness to the facts; hence the need for strengthening process reliabilism by accommodating the ability intuition on knowledge.

⁷ Notice, here, that the claim is that the cognitive success must be the *product* of a reliable cognitive belief-forming process. It is not the weaker claim that cognitive ability must have been involved in the acquisition of one's true belief, since this can be satisfied far too easily in ways that do not exclude luck.

Barney⁸

Barney is driving through the country and happens to look out of the window into a field. In doing so, he gets to have a good look at a barn-shaped object, whereupon he forms the belief that there is a barn in the field. This belief is true, since what he is looking at is really a barn. Unbeknownst to Barney, however, he is presently in ‘barn-façade country’ where almost every object that looks like a barn is a convincing fake. Had Barney looked at one of the fake barns, then he would not have noticed the difference. Quite by chance, however, Barney just happened to look at the one real barn in the vicinity.

Barney comes to truly believe that he is looking at a real barn by employing his cognitive abilities; he is looking directly at a real barn. Therefore, we have a case where despite the fact that Barney’s true belief is solely formed on the basis of his cognitive ability, his cognitive success cannot be called knowledge given that his belief could have so easily been false; Barney is in a barn-façade environment. Put another way, the objection is that we have cognitive success that does not amount to knowledge even though it is solely formed on the basis of Barney’s reliable cognitive character.

Consequently, it has been argued that robust virtue reliabilism cannot ultimately do all the work that is expected to do. Despite robust virtue epistemologists’ initial expectations, the ability intuition on knowledge seems unable to *fully* accommodate the equally important intuition that knowledge must not be due to luck, *viz.* the anti-luck intuition on knowledge. Instead, it seems that in order to deal with knowledge-undermining luck, virtue reliabilists must also incorporate a separate anti-luck condition into their theory of knowledge. The upshot appears to be that these two intuitions about knowledge “impose independent epistemic demands on our theory of knowledge” (Pritchard 2010a, 2).

Accordingly, friends of anti-luck epistemology claim that any adequate theory of knowledge must explicitly have as a central component an anti-luck epistemic condition such as the safety or the sensitivity principle.⁹ In contrast to the ability

⁸ The Barney case is described in Goldman (1976) and credited to Carl Ginet. This formulation of the example is due to Pritchard (2009, 12).

⁹ The *sensitivity principle* is usually formulated as follows: If *S* knows that *p*, then *S*’s true belief that *p*, is such that, had *p* been false, *S* would not have believed *p*. The classic defenses of the sensitivity principle can be found in Dretske (1970) and Nozick (1981). The *safety principle* is usually understood thusly: if *S* knows that *p*, then *S*’s true belief that *p*, is such that *S*’s belief that *p* could not have easily been false. For recent defenses of the safety principle see Sosa (1999, 2000) and Pritchard (2002b,

condition on knowledge, which arguably addresses the problem posed by knowledge-undermining luck only indirectly, these modal conditions on knowledge are primarily targeted to capture the anti-luck requirement, as they are explicitly concerned with the responsiveness of one's belief to relevant counterfactual circumstances (such as the scenario in which Barney looks at a barn-façade instead of a real barn).¹⁰ Arguably, then, for virtue reliabilism to be a fully adequate account of knowledge it may have to be supplemented by a specific anti-luck condition on knowledge such as safety or sensitivity.¹¹

Therefore, on the face of the Barney counterexample robust virtue reliabilism appears to be an insufficient condition on knowledge. However, the following case, demonstrates that it is too strong as well.

Jenny¹²

Jenny gets off the train in an unfamiliar city and asks the first person that she meets for directions. The person that she asks is indeed knowledgeable about the area, and gives her directions. Jenny believes what she is told and goes on her way to her intended destination.

2008). For a very good discussion concerning the relation between the ability and the anti-luck intuition on knowledge see Pritchard (2010a).

¹⁰ Notice, however, a subtle difference: In regular Gettier cases, the knowledge-undermining luck is very direct, in that the luck concerns the relationship between the belief and the fact. One's belief is erroneously formed but there is a lucky fact that renders it true. Luck intervenes between the belief and the fact. On the contrary, in cases like the one Barney is in, the knowledge-undermining luck is quite indirect; indeed, it is specifically *environmental*. Barney does look at a real barn and believes that there is a real barn in front of him. There is nothing wrong with the way he forms his belief. However, given the environmental conditions he cannot acquire knowledge in this way. Luck interferes with the environment in such a way that even a well-formed belief will be a lucky one if true. So, it seems that while the ability condition on knowledge can deal with the normal (non-environmental) knowledge-undermining *epistemic* luck involved in Gettier cases, it may not be able to deal specifically with *environmental* knowledge-undermining luck.

¹¹ Consider for example *Anti-Luck Virtue Epistemology*: *S* knows that *p* if and only if *S*'s safe belief that *p* is the product of her relevant cognitive abilities (such that her safe cognitive success is to a significant degree creditable to her cognitive agency) (Pritchard 2010a, 20). Again, in (Pritchard 2010b, 76) we can read: "knowledge is safe belief that arises out of the reliable cognitive traits that make up one's cognitive character, such that one's cognitive success is to a significant degree creditable to one's cognitive character". Notice, however, that Greco (2007, 2008) has put forward a 'subject-sensitive, interest-dependent' contextualism for judging which belief-forming processes are reliable each time that may successfully deal with the Barney and similar cases, thereby avoiding the need for an explicit anti-luck component in a complete account of knowledge.

¹² The Jenny case is adapted from Jennifer Lackey's 'Morris case' (2007, 352). This formulation of the example is due to Pritchard (2009,68). Apart from the hero's name, however, nothing else really changes.

Unless we want to deny a great amount of knowledge that we suppose we have, we must admit that Jenny gains knowledge in this way. However, even if we read the example both in such a way that Jenny is in an epistemically friendly environment (i.e. the city that Jenny visits had better not be renowned for its dishonest informants) and by presupposing that Jenny is sensitive to some epistemologically relevant factors, such that she can distinguish between potentially reliable and clearly unreliable informants and information, we would still have to conclude that her cognitive character is not the most important factor in the causal explanation of how she believes the truth.¹³ Instead, if credit is to be attributed for her true belief, it should be primarily directed to her informant. (Alternatively: The most salient factor in the causal explanation of how Jenny believes the truth is her informant). So, according to robust virtue reliabilism Jenny lacks knowledge that she in fact possesses.¹⁴

However, apart from these two counterexamples that show robust virtue reliabilism to be both a too strong and too weak account of knowledge, there is also another problem looming on the side. Recall that in addition to one's organismic cognitive abilities of the brain/CNS, 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" (Greco 1999, 287). The reason for this move is that virtue reliabilism needs to account for advanced cases of knowledge in which one's believing the truth is the product of the operation of epistemic artifacts such as telescopes, microscopes, or as in the case of Travis,

¹³ For more details on the Jenny case, see (Pritchard 2010a, 18).

¹⁴ Traditionally, accounts of testimonial knowledge are divided in two main trends. The first one is called reductionism and it is the view that a hearer is justified in believing a speaker's testimony if and only if she has non-testimonial positive reasons in favor of the speaker's reports, such that her justification for accepting them is reducible to basic sources of knowledge such as sense perception, memory and inductive inference. It is often thought that Hume was committed to a form of reductionism (Hume 1977). Contemporary proponents of the view include Faulkner (2000) and Fricker (1994). The second trend is called non-reductionism and it is the view that one is by default justified in believing one's testimony unless one has negative reasons for doing so. On this view, testimonial justification cannot or need not be reduced to more basic sources of knowledge. Reid (1983) is thought to be the first to defend the view. Contemporary proponents of non-reductionism on testimonial knowledge are, amongst others, Burge (1993), Weiner (2003) and Audi (1998). Recently, however, Jennifer Lackey (2008) has put forward a dualist account of testimonial knowledge, which accommodates both of the aforementioned views.

(Robust) virtue reliabilism appears to accord only with reductionism on testimonial knowledge, due to its strong demand that the cognitive success should be primarily creditable to the hearer's cognitive character.

TVSS. In the traditional conception, however, cognition takes place strictly within an agent's head and so artifacts cannot be parts of one's *cognitive* character.

One way to sidestep this problem for robust virtue reliabilism could be to claim that, in such cases, it is merely the agent's training and skill of using the artifact, as mirrored in the agent's neural/bodily architecture, that is the most salient factor in the causal explanation of acquiring the cognitive success.¹⁵ Notice, however, that when an agent employs an epistemic tool, his true belief arises as a product of the *interaction* between his internal processes and the artifact. What this means is that the agent's cognition is not merely 'aided' or 'assisted' by the artifact, but, instead, the very content of his beliefs, which he will later express in propositional terms, arises out of his ongoing engagement with the artifact (think about Travis' phenomenal experience, for example). It thereby appears that it will be impossible to disentangle the agent's training and skill of using the artifact from his actual engagement with it in a causal explanation of how the agent acquired his true belief.¹⁶ Crucially, however, even if such decomposition were possible, notice in addition that the part of the process that allows the cognitive agent to detect the truth, or in other words to be sensitive to the facts, is the external component. To make this point clear, consider, on one hand, an untrained agent in possession of a properly working artifact. In that case it is obvious that even though the agent will initially be unable to form any beliefs, eventually, provided that he gains sufficient experience, he will enjoy cognitive success. On the other hand, think about a well-trained agent (like Travis), but in possession of a faulty artifact. In this case, despite the agent's excellent internal skills, it is evident that he would be unable to reach any (non-lucky) true beliefs, no matter how much he tried. It, therefore, seems that, in such cases, the most salient factor that explains the agent's cognitive success is the epistemic artifact. In other words, since the epistemic artifact is the most salient part of the *integrated* belief-forming process that produces one's true believing, the virtue reliabilist must account for it being part of one's cognitive system. However, given that cognition is normally supposed to take place within the agent's head, Greco has no principled way to show why such artifacts may count as proper parts of one's cognitive character. Therefore, there

¹⁵ Thanks to an anonymous referee for pressing this point.

¹⁶ Provisionally this point may seem ambiguous but it will hopefully become clearer in the discussion of section two where I explore the phenomenon of continuous reciprocal causation between the organismic agent and the epistemic artifact.

seems to be a worrying tension between the ability intuition on knowledge and such a broad understanding of the notion of one's cognitive character.

In order to solve these problems for the ability intuition on knowledge, Pritchard (2010c) has recently proposed a virtue reliabilistic necessary condition on knowledge, namely COGA_{weak}.

COGA_{weak}: If S knows that p, then S's true belief that p is the product of a reliable belief-forming process, which is appropriately integrated within S's cognitive character such that her cognitive success is to a significant degree creditable to her cognitive agency. (Pritchard 2010c, 136-7)

Obviously, the Barney case is not a problem for COGA_{weak} since it is only a necessary condition on knowledge; even if Barney's cognitive success is significantly creditable to his cognitive agency, he may nevertheless lack knowledge because he has failed to satisfy some supplementary condition on knowledge such as the safety principle. Moreover, COGA_{weak} can easily accommodate the Jenny case; although the cognitive success is not primarily creditable to Jenny's cognitive agency—but rather to the stranger who delivers the reliable information—Jenny, in so being able to distinguish between clearly unreliable and potentially reliable informants and information, has the right sort of abilities and employs them in the right sort of way so as to accept the stranger's information. In this way, believing the truth is significantly creditable to her cognitive agency. Therefore, according to COGA_{weak}, Jenny can gain knowledge.

What is of import here is to notice the lenient demands of COGA_{weak} regarding the creditability of the cognitive success to one's self. In contrast to robust virtue reliabilism where believing the truth must be primarily creditable to one's cognitive character and thereby to one's self, COGA_{weak} loosens the required dependence of the cognitive success on one's *cognitive agency* thereby allowing significant part of the credit to be attributable to other factors as well. Therefore, according to COGA_{weak}, even though the most salient factor that explains Jenny's cognitive success is the informant's contribution, Jenny's cognitive abilities render her cognitive agency significantly creditworthy, thereby allowing her to gain knowledge in this way.¹⁷

¹⁷ Notice, then, that the cognitive success being *primarily* creditable to one's cognitive agency and it being the product of cognitive abilities is not exactly the same thing. That is, one's cognitive success

Moreover, according to COGA_{weak}, so long as one forms true beliefs on the basis of some process in such a way that believing the truth can be significantly creditable to one's cognitive agency then one may be said to know the target proposition. Since COGA_{weak} is a formulation of the ability intuition on knowledge, what this means is that so long as one's true belief is significantly creditable to one's cognitive agency, then the process by which one came to form one's belief can be said to have been appropriately integrated within one's cognitive character, thereby counting as a *bona fide* cognitive ability.¹⁸ Therefore, in contrast to robust virtue reliabilism, COGA_{weak} offers a principled way to account for the acquisition of knowledge on the basis of epistemic artifacts, while retaining the ability intuition on knowledge.

We thus see that apart from accommodating our intuitions with respect to the thought experiments that trouble robust virtue reliabilism, COGA_{weak} can also create some important logical space for explaining how we may come to acquire knowledge on the basis of epistemic artifacts, which lie outside our heads but which may nevertheless count as parts of our cognitive characters. However, considering artifacts as part's of one's cognitive character such that their employment can count as a *bona fide* cognitive ability is a rather radical claim, which had better not been left unsupported. Accordingly, there remains to see whether there are available any independent resources for properly conceptualizing the use of artifacts in a way that is continuous to our understanding of the organismic cognitive abilities.¹⁹ And should there be such an account, what does it take for those artifacts to be appropriately integrated within one's overall cognitive character other than our intuitions on the degree of creditability of one's cognitive success to one's cognitive agency?

ii) The Hypothesis of Extended Cognition

can be the product of one's cognitive abilities even if it is not *primarily* creditable to one's cognitive agency.

¹⁸ Pritchard makes this point in (2010c, 137; en. 7). Although this might generally be a good way to start judging whether a process can count as a *bona fide* cognitive ability, notice that Pritchard uses the terms 'cognitive agency' and 'cognitive character' interchangeably thereby running the risk of rendering his criterion circular or at least unsafe. As I argue in section four, where I discuss several thought experiments, Pritchard has indeed been led astray by his criterion with respect to the Temp case. Therefore, we are in need of an alternative way to judge whether a process has been *appropriately* integrated within one's cognitive character. I offer one in section three.

¹⁹ That is, even though we might be epistemologically motivated to accept the employment of artifacts as *bona fide* cognitive abilities, such a claim would be very weak in the absence of any metaphysical support.

Most probably, the only way to argue for the inclusion of artifacts within one's cognitive character is through the consideration of the hypothesis of the extended cognition. According to HEC, "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).

HEC is a form of active externalism, which should be distinguished from meaning (or passive) externalism as presented in the writings of Putnam (1975) and Burge (1986). Active externalism, which is different from the aforementioned traditional forms of externalism in that it concerns the aspects of the environment that actively drive one's cognitive loops, has been defended by many philosophers and appears in the literature with as many names as the number of its proponents: the extended mind (Clark and Chalmers 1998), environmentalism (Rowlands 1999), locational externalism (Wilson 2000; 2004), cognitive integration (Menary 2007) and so on. However, for reasons of simplicity and theoretical affinity to the epistemological framework of virtue reliabilism, I will here discuss the view on the basis of the terminology and argumentative lines that appear in Clark and Chalmer's initial proposal (1998).

It is true that in its relatively short history HEC has faced an impressive number of objections. 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 (Rupert 2004; Adams and Aizawa 2008) (some of them stemming from a misunderstanding of the parity principle)²⁰, or to the perceptive rather than introspective manipulation of the external elements. Others deny the mark of the cognitive to the alleged extended cognitive processes (Adams and Aizawa 2008), or claim that there cannot be an extended cognitive science (Rupert 2004; Adams and Aizawa 2008) and so on.²¹ The list could go on for a while.

²⁰ 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). Parity principle is only meant as an intuition pump in order to overcome internalist prejudices, when judging whether external processes could in principle be parts of one's overall cognitive system. Many critics, however, have given much stronger interpretations that generate counterintuitive results. For a very good discussion on the objections raised against the Parity Principle see (Menary 2007; 2010a; 2010b).

²¹ For a short discussion of most of these objections see (Menary 2006).

However, it is also true that many of the objections have been convincingly confronted, thereby making HEC a viable hypothesis with a growing appeal in several disciplines such as philosophy of mind, AI, developmental psychology, and possibly epistemology as well. In any case, the discussion of the objections facing HEC is well beyond the scope of the present paper. However, in my exposition of the extended cognition hypothesis I will, eventually, address two interrelated objections, namely the coupling-constitution fallacy and the ‘cognitive bloat’ worry, which as it will gradually become apparent are closely connected to my aim of delineating the hard core of HEC.

According to HEC, when parts of the environment become properly coupled to an agent’s brain then they too can be considered as constitutive parts of the overall cognitive mechanism –*viz.* cognition potentially extends to the world surrounding the agent. Put another way, “in these cases, the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right” (Clark and Chalmers 1998, 8). Clark has also termed this two-way interaction as ‘continuous reciprocal causation’ (CRC): “CRC occurs when some system S is both continuously affecting and simultaneously being affected by activity in some other system O” (Clark 2008, 24). In such cases, according to Dynamic Systems Theory, in order to account for, understand, and describe (i.e. model) the temporal evolution of the two systems S and O, one must further postulate a coupled system E consisting of both S and O. In cases of extended cognition, one’s internal cognitive capacities are combined (i.e. mutually interact) with some environmental element O to form an extended cognitive whole, E, whose behavioral competence will drop if one removes the external component, just as it would drop if one removed part of its brain.²²

²² I am thankful to an anonymous reviewer for pointing out that the above point might be interpreted as an epistemic argument for HEC; in short, that HEC is true because we must consider the contribution of external materials if we are to fully understand human cognition. This is a clearly fallacious argument but it is just a misinterpretation. My point is rather conceptual: since the best available mathematics (i.e. conceptual tools) we have in order to model systems that exhibit CRC postulate a further coupled (extended) system, we should accept the hypothesis of extended cognition as true.

As Julian Kilverstein has noted in personal communication, however, the same point may be interpreted as a metaphysical one, concerning the nature of the systems involved in cases of CRC. As Varela writes: "If one says there is a machine *M* in which there is a feedback loop through the environment, so that the effects of its output affect its input [i.e. *M* reciprocally interacts with some environmental aspect], one is in fact talking about a larger machine *M'* which includes the environment and the feedback loop in its defining organization" ((Varela 1979, 158), quoted by Hurley (1998, 404)). The reason is that the effects of the environment on *M* are partly defined by *M*'s ongoing states *and*

Consider, for example, the use of pen and paper when trying to solve a complex, say, a three-digit multiplication problem such as 987 times 789. It is true that few, if any, of us can solve this problem by looking at or contemplating on it. We may only perform the multiplication by using pen and paper to externalize the very problem in symbols. Then, we can serially proceed to its solution by performing simpler multiplications starting with 9 times 7. In this way, the pen and paper compensate for our limited working memories allowing us to perform a task that it is otherwise infeasible. If one should try to describe how a regular human mind may perform such a cognitive task then, apart from the states and properties of a typical human brain, one should also factor in both the normative aspects of the notational/representational system involved, and the properties and ongoing states of the mediums with which the manipulation of the representations was performed.²³

Or, think about the role of language when writing a philosophy paper. According to Clark, language too is “an external epistemic artifact designed to complement, rather than recapitulate or transfigure, the basic processing profile we share with other animals” (1998, 169).

As I write down this essay, “I am continually creating, putting aside, and re-organizing chunks of text. I have a file, which contains all kinds of hints and fragments, stored up over a long period of time, which may be germane to the discussion. I have source texts and papers full of notes and annotations. As I (literally, physically) move these things about, interacting first with one, then another, making new notes, annotations and plans, so the intellectual shape of the chapter grows and solidifies. It is a shape which does not spring fully developed from inner cogitations. Instead, it is the product of a sustained and iterated sequence of interactions between my brain and a variety of external props” (Clark 1998, 173).

The moral, Clark claims, is that public language and text play more than just a preserving-and-communicating-ideas role:

“Instead, these external resources make available concepts, strategies and learning trajectories which are simply not available to individual un-augmented brains. Much of the true power of language lies in its underappreciated capacity to re-shape the computational spaces which confront intelligent agents” (Clark 1998, 173).^{24 25}

vice versa. This dense and complex *interaction* amounts to an overall behavior which is in fact internal to a larger system *M'* comprising of both *M* and the relevant environmental aspect.

²³ For the importance of the normative aspects of the external representational systems in understanding cognition see (Menary 2007).

²⁴ For a discussion on Clark’s view regarding language see (Wheeler, 2004). For a straightforward criticism, see (Rupert, 2010).

To return to the discussion of the core tenets of HEC, let us now take a look at a further example that Clark and Chalmers (1998) use in order to argue for the more provocative Extended Mind Thesis from the existence of extended mental states. 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 take a look at 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, thus, 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. Clark and Chalmers 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; "the notebook plays for Otto the same sort of 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" (1998, 13).

Although the postulation of extended mental states is not necessary for making the case for HEC, thereby allowing us to bypass the admittedly long debate that Otto

²⁵ But if one allows for such an understanding of public language and text, then important conceptual space is created to lengthen the list of epistemic artifacts in interesting ways. For, as Robert Logan (2003, 275) claims, "speech, writing, math, science and computing form an evolutionary chain of languages. Each of these activities can be considered as a separate language because each allows us to think differently, create new ideas and develop new forms of expression. Another consideration is that each of these five forms of language possesses its own unique semantics and syntax and hence qualifies as a language in itself according to criteria set by classical linguistics". While much more remain to be said on this matter, concentrating on the case of scientific theories, it is interesting that philosophers of science such as Imre Lakatos write: "[scientists] *use our most successful theories as extensions of our senses*" (Lakatos 1970, 107, emphasis in the original). Given the appropriate theorizing, however, this may turn out to be more than just a metaphor. That is, it could be the case that scientific theories, like public language and text, are software epistemic artifacts that extend one's cognitive abilities beyond one's natural cognitive capacities.

has generated, the discussion of this example is helpful as it has produced some very important intuitions on what is required for an external artifact to count as a putative part of one's overall cognitive economy. In particular, investigating the case in more detail, Clark (2010) notes that the availability and portability of the resource of information might be crucial. Accordingly, he offers the following set of additional criteria to be met by non-biological candidates for inclusion into an individual's cognitive system (2010, 46):²⁶

- 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."

These criteria have also come to be known in the literature as the 'glue and trust' criteria and they are primarily meant to ensure the effect of 'transparent equipment': "equipment (like the carpenter's hammer) with which we are so familiar and fluent that we do not think about it in use, but rather rely on it to mediate our encounters with a still-wider world" (Clark 2006, 106). Put another way, an external element is part of one's ongoing cognitive loops when it is not part of the problem space but is instead one of the mediums manipulated in order to complete the cognitive task at hand.

It has been pointed out, however, that the 'glue and trust' criteria may be far too easily satisfied and, therefore, they are not enough to ensure that some external element can count as part of one's cognitive system.²⁷ For example, Rupert (2004, 401-5) argues that when a person has regular access to a phonebook, or a directory service through the use of her cellular phone she can be said to satisfy the criteria that Clark has set forth. However, it would be highly counterintuitive to conclude that the phonebook, or the directory service is part of that person's overall cognitive system

²⁶ This paper was first published in *The Extended Mind*, (2010), Menary (ed.) Cambridge, Massachusetts, MIT press, but it has been available online since 2006.

²⁷ In (Clark and 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". However, as the authors further note, "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.

allowing her to have non-occurrent true beliefs about the phone numbers of everyone whose number is listed.

Therefore, if any external element that both satisfies the ‘glue and trust’ criteria and causally affects one’s cognitive processes is to count as part of one’s cognitive system, we are going to be led to a ‘cognitive bloat’ (Clark 2001, Rowlands 2009, Roberts *manuscript*) whereby cognition will seem like leaking all the way out in implausibly many directions. Note, however, that this obvious mistake would be another instance of committing the (in)famous ‘coupling-constitution’ fallacy, according to which “it simply does not follow from the fact that process X is in some way causally connected to a cognitive process that X is thereby part of that cognitive process” (Adams and Aizawa 2008, 91). Accordingly, instead of arguing for the constitutive contribution of the external artifacts to one’s overall cognitive economy, one should simply endorse the much less provocative idea that cognition is many times *dependent* on external elements. Consequently, one should better opt for a less radical position which has come to be known as the Hypothesis of Embedded Cognition (HEMC) (Rupert 2004, 393).

HEMC: Cognitive processes depend very heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place.

This hypothesis is close to HEC as it acknowledges the dependence of cognition on its environment. However, it is a much more conservative view because it denies that environmental aspects are proper parts of cognition; external factors may only serve as tools and props to cognition, which is restricted within the organismic brain or, at most, the organismic body as a whole. Thus, according to HEMC, cognition is organism-bound, potentially aided by environmental factors, but not extended to them. And should one try to analyze the causal dependence of the internal elements on any external factors, one could simply proceed by accumulating the properties of the distinct parts that are involved in this one-way linear dependence.

At this point, however, we should return to one of the core claims of HEC that we noted earlier, namely the continuous reciprocal causation between the internal and

external elements that are involved in cases of cognitive extension. It is only when this phenomenon occurs that we can safely talk about extended cognition. As Clark suggests (2007, 38), “when we confront a recognizably cognitive process, running in some agent, that creates outputs (speech, gesture, expressive movements, written words) that recycled as inputs drive the process along”, these inputs should count as constitutive parts of the overall cognitive process.²⁸ In other words, we should think of extra-organismic parts as constitutive to the overall cognition generating mechanism when they “emerge as interacting parts of a distributed cognitive engine participating in cognitively potent self-stimulating loops, whose activity is as much an aspect of our thinking as its result” (Clark 2007, 39).

The reason is that, in such cases, the outer and the inner contributions come together in a highly complex, probably non-linear, two-way dependence. According to the Dynamic Systems Theory, the two components are treated as a *coupled system* in its own right because “the equation describing the evolution of each component contains a term that factors in the other system’s current state (technically, the state variables of the first system are also parameters of the second, and vice versa)” (Clark 2008, 24).²⁹ Therefore, the cases in which HEC is interested cannot be analyzed in the one-way causal dependency that HEMC suggests.³⁰

Echoing Susan Hurley (2010), whether external elements might be proper parts of one’s cognitive system should not be a metaphysical debate on the nature of the mind but rather a matter of successful scientific explanation. The internalist should not

²⁸ Notice that, as the above quote suggests, Clark treats CRC only as a sufficient condition on cognitive extension. Given, however, that CRC not only helps to overcome several problems for the hypothesis of extended cognition, but it also helps to distinguish the view from rival alternative theories (i.e. from HEMC, see also footnote 30) that might allegedly offer the same causal-explanatory accounts for a given phenomenon, I wish to accentuate its importance by treating it as a necessary condition on cognitive extension.

²⁹ For more details see (Van Gelder 1995, 355-8, 373). See also ft. 20.

³⁰ An anonymous referee has pointed out that the HEMC theorist may be happy to accept that an agent might be in a continuous and reciprocal causal relation with some aspect in the environment, but still deny cognitive extension. However, this seems not to be an option for the HEMC theorist because given the conceptual framework of Dynamic Systems Theory, in such cases, HEMC collapses into HEC. In fact, Rupert, in chapter 7 of his book (2009), concedes that Dynamic Systems Theory can provide strong support to the hypothesis of extended cognition in just the way I have been here describing (Rupert 2009, 131-4). It is telling that none of the dynamic models that he considers in favor of HEMC concerns a two-way interaction between the organism and some environmental aspect. For more details see (Rupert 2009, 137-149).

judge from a pre-scientific and metaphysical standpoint, without independent arguments, whether cognition may be constitutively dependent on external elements, or not, because this is exactly what is at issue here. “To avoid thus begging the question, we should not operate with prior assumptions about where to place the causal-constitutive boundary, but wait on the results of explanation” (Hurley 2010, 106).³¹

In the framework of the continuously and reciprocally interacting dynamical systems that motivates HEC, the boundaries are not “exogenous to explanatory aims. In cognitive applications, the state space can extend to include dimensions whose variables are bodily and environmental as well as neural, as brain, body and environment interact in mutually shaping patterns” (Hurley 2010, 130). Only time will show whether such externalist approaches to cognition will keep being explanatorily helpful. In any case, however, “the issues between internalism and externalism should be resolved bottom up by such scientific practice, not by advance metaphysics” (Hurley 2010, 107). HEMC and HEC will have to compete with each other on the empirical field.³²

Returning to the ‘cognitive bloat’ worry, it seems that no CRC occurs in the case of the phonebook or the directory service. That is, the person employing those resources does not generate any outputs on their basis, which recycled as inputs drive her ongoing cognitive loops along. Those resources are simply causally, as opposed to constitutively, related to the agent in that they deliver her information whose formulation was entirely independent to her cognitive agency. That is, there is no two-way causal interaction between the agent and the epistemic artifact. If an account were to be provided for those cases, then it would have to be a linear one, probably in testimonial terms.

³¹ Adams and Aizawa (2008) claim that the mark of the cognitive is the manipulation of representations with underived content, which is plausibly not a feature of any external process, and they thus avoid begging the question against externalism when they put forward the causal-constitution fallacy. However, it is not clear what Adams and Aizawa have in mind and how the said criterion is supposed to promote internalism. Thus a long debate has been generated. Indicatively, see (Clark 2008), (Clark 2010*b*), (Menary 2006), (Adams and Aizawa 2010), (Ross and Ladyman 2010).

³² Notice, however, that both hypotheses may turn out to be successful, though HEMC might be so only with respect to more mundane cases.

Consequently, then, apart from the ‘glue and trust’ criteria, in order to account for the constitutive status of external elements within one’s overall cognitive mechanism, we also need the phenomenon of continuous reciprocal causation between the outer and the inner parts to take place. These 3+1 criteria seem to jointly ensure the integration of the external artifacts within one’s overall cognitive mechanism, thereby overcoming the ‘coupling-constitution’ fallacy and the ‘cognitive bloat’ worry.

iii) **Belief-Forming Processes, Extended**

Returning to the epistemological discussion, let us see how the above points might be of interest. First, however, we should refresh our memory with respect to the criteria that a belief-forming process must meet so as to deliver knowledge, according to the broader virtue reliabilistic framework.

Pritchard, firstly and foremost, points to the fact that the relevant belief-forming process must be appropriately integrated within the agent’s cognitive character. Second, we want the relevant belief-forming processes to be reliable, where a reliable process is one that tends to lead to true rather than false beliefs. Recall, however, that according to reliabilism, one does not need one’s evidence to be necessarily reliable; if forming a belief on a certain kind of evidence constitutes a reliable belief-forming process, it does not matter that one’s evidence is only contingently reliable. Moreover, according to virtue reliabilism, the agent on his part does not need to know or have any beliefs about the reliability of his belief-forming process; it is sufficient that she is sensitive to the reliability of her way of forming beliefs simply by it being one of the cognitive dispositions/habits that she employs when she is thinking conscientiously. Finally, then, not all reliable belief-forming processes are knowledge-conducive; rather “it is those processes that have their bases in the stable and successful *dispositions* of the believer that are relevant for knowledge and justification” (Greco 1999, 287, the emphasis is added). Combining the three: *a belief-forming process is knowledge-conducive if it is a reliable belief-forming disposition properly integrated within the agent’s cognitive character*. Now, let us recall what the three criteria to be met by non-biological candidates for inclusion into

an individual's cognitive system are, as offered by Clark (2010, 46):

1)“That the resource be reliably available and typically invoked”. That is, the agent should habitually and easily invoke the external resource. In other words, its employment must be a *disposition/habit* of the agent's overall cognitive mechanism.

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”. That is, the information in the resource must be *regarded* as reliable, and not be necessarily reliable, so long as its employment results into an equally trustworthy belief-forming process as the one of forming beliefs on the basis of one's own biological memory.³³

One might object, however, that being reliable is not the same as being *trustworthy* (i.e. being regarded as reliable).³⁴ In response, notice first that Clark identifies the notion of trustworthiness of a process with the idea of being “more-or-less automatically endorsed” or, in other words, “not usually being subject to critical scrutiny”, which means that the target process must have not been (for the most part) problematic in the past. Moreover, the processes under consideration are supposed to be cognitive dispositions/habits of the agent, which he has *repeatedly* employed in the past, and so had they been problematic the agent would have noticed that and responded appropriately. Accordingly, a *trustworthy* belief-forming process, in Clark's account, will be one that tends to lead to true rather than false beliefs, which is to say that it will be objectively reliable in the virtue reliabilist's sense. What the agent will deem reliable will be what is objectively reliable, i.e. that which has *not* been (for the most part) problematic in the past.³⁵

³³ That is, the process does not need to be, due to any underlying logical or quasi-logical relations, 100% reliable. Memory, for example, is supposed to be reliable even though one may misremember, and so it is only contingently reliable.

³⁴ The idea here is that trustworthiness (i.e. being regarded as reliable) might sometimes supervene on values other than objective reliability (i.e. the tendency to lead to success rather than failure). See the next footnote for a description of just such a case.

³⁵ To elaborate a bit more on this point, an anonymous referee has objected that in the United States, there are many individuals who trust a particular cable TV "news organization", but this news organization often provides misleading information. Viewers of this cable TV "news" channel trust this medium, but it is not reliable in the epistemologist's sense. So, it may now appear that Clark's conditions and the epistemologist's reliability condition come apart. In response, let me draw your attention to two points. First, as noted above, Clark's claim that in order for something to be trustworthy it must not usually be subject to critical scrutiny implies that the agent *would* check the object of his trust on the face of discrepancies. In contrast, the viewers of the American TV news channel appear that they *would not*, and this is a kind of *blind trust* that is quite different from the kind of trust Clark has in mind. Accordingly, the fact that there might be kinds of trust that have nothing to

Now, notice this negative way of deeming processes reliable with which Clark concurs, i.e. that a trustworthy process is one that is *not usually* subject to critical scrutiny such that it is more-or-less automatically endorsed. What this means is that the agent does not need to have any beliefs about why or whether the belief-forming process is trustworthy; it suffices that it has not repeatedly caught his negative attention in the past. This is in perfect agreement with Greco's demand that one's subjective justification need not rely on any beliefs but simply on one's motivation to believe the truth. For example, one may trust one's vision in appropriate circumstances, just because vision has not been particularly problematic in the past. It suffices that one is motivated to believe the truth and will thereby employ the belief-forming process that has not in the past (generally) failed to be conducive towards this end, and, crucially, one will do so without even thinking about it. This methodological point, in turn, appears to be in line with the spirit of the glue and trust criteria which, after all, are meant to ensure the effect of transparent equipment: equipment that we are so fluent and familiar with that we have no beliefs about it in use.

3) "That information contained in the resource should be easily accessible as and when required". That is, the information must be directly available to the agent whenever necessary without difficulty; the agent must be able to employ the resource as if it was part of his organismic cognitive mechanism. In other words, the resource must have been *appropriately integrated* within the agent's overall cognitive mechanism.

We thus see that in order for non-biological elements to be included into one's cognitive system, their deployment must meet the same criteria as the ones that the belief-forming processes must satisfy so as to count as knowledge-conducive. However, we also need to make a further remark concerning the *appropriate* integration of the external elements within an agent's cognitive character.

do with objective reliability, should not generate any problems for the identification of Clark's 'trustworthiness' condition with the epistemologist's objective reliability criterion.

Second, even if on Clark's account, the American TV "news" channel were to count as part of one's cognitive economy—were it to also satisfy the rest of the criteria for cognitive extension—one could note that this is an issue about which epistemology and philosophy of mind can, and should exchange normative considerations. That is, even if, when philosophizing about what may count as part of an agent's mind, we do not focus on the nature of a good mind, epistemology could point out that a *conscientious* mind should try to believe what is true and thereby employ the resources, which are reliable in the epistemologist's sense. Nevertheless, Clark seems to accommodate this normative dimension of the mind through his "not-usually-subject-to-critical-scrutiny" understanding of trustworthiness, thus having been in line with the reliabilist tradition in epistemology, all along.

As we noted in the previous section, an external element will not count as properly integrated within one's cognitive system unless it is also reciprocally coupled to the agent's organismic cognitive faculties. That is, in order for an artifact to be a constitutive element of one's ongoing cognitive loops, the agent must deliver on its basis outputs which recycled as inputs will drive his cognitive character along. Therefore, in order for a belief-forming process to be appropriately integrated within one's cognitive character, the phenomenon of CRC must be manifested between the target process and one's organismic cognitive faculties.

The underlying reason for this is that HEC clearly acknowledges the central role of the persisting biological organism in recruiting and maintaining the extended organization in order to eventually accomplish its very own cognizing: "Human cognitive processing (sometimes) extends to the environment surrounding the organism. But the organism (and within the organism the brain/CNS) remains the core and currently the most active element. Cognition is organism centered even when it is not organism-bound" (Clark 2007, 49). Note, then, that using the phenomenon of CRC to judge whether an artifact has been appropriately integrated within one's cognitive character is in line with Pritchard's suggestion to consider whether the cognitive success will eventually be significantly creditable to one's cognitive agency; it is the agent's organismic cognitive faculties (i.e. the brain/CNS) that are first and foremost responsible for the recruitment of the external elements on whose basis the agent will deliver outputs which recycled as inputs will drive her cognitive character further along so as to eventually form a true belief with respect to some proposition *p*.

We thus see that the 3+1 criteria are common currency for both COGA_{weak} and HEC, thereby rendering both accounts complementary. Reasonably, then, it may also be claimed that there is no principled theoretical bar disallowing extended belief-forming processes to count as knowledge-conducive cognitive abilities. That is, provided the satisfaction of the 3+1 criteria, one's cognitive character may extend beyond the organismic cognitive abilities that make up one's cognitive agency, by incorporating epistemic artifacts. Put differently, given the right conditions, the deployment of epistemic artifacts can count as knowledge-conducive cognitive abilities, allowing us to gain knowledge on their basis in accordance with the ability intuition on knowledge.

iv) Case studies

So let us finally see how the 3+1 criteria rule with respect to several versions of two core thought experiments that Pritchard discusses in (2010*c*), namely the Temp and Alvin case.

Temp³⁶

Temp's job is to keep a record of the temperature in the room that he is in. He does this by consulting a thermometer on the wall. As it happens, this way of forming his beliefs about the temperature in the room will always result into a true belief. The reason for this, however, is not because the thermometer is working properly, since in fact it isn't—it is fluctuating randomly within a given range. Crucially, however, there is someone hidden in the room next to the thermostat who, unbeknownst to Temp, makes sure that every time Temp consults the thermometer the temperature in the room is adjusted so that it corresponds to the reading on the thermometer.

Intuitively and according to $\text{COGA}_{\text{weak}}$ Temp lacks knowledge of the temperature. As Pritchard explains, even though the way in which Temp forms his true beliefs is reliable, it in no way reflects his cognitive abilities. Therefore, since Temp fails to meet the ability intuition on knowledge, $\text{COGA}_{\text{weak}}$ disallows knowledge to Temp. One may further argue that the reason why Temp fails to meet the ability intuition is that the relevant belief-forming process lies outside his head. Consider, however, Alvin.

Alvin³⁷

Alvin has recently developed an unusual brain lesion, a guaranteed side effect of which is that it prompts him to randomly, but reliably, form true beliefs about the product of fairly complicated arithmetical sums.

Alvin's belief-forming process is indeed reliable. However, even though it lies under Alvin's skin, Pritchard argues that Alvin lacks knowledge of the mathematical propositions. The reason, Pritchard claims, is that as in the Temp case, Alvin's cognitive success has nothing to do with Alvin's cognitive abilities but is rather the "fortunate consequence of the otherwise unfortunate fact that he has a brain lesion" (2010*c*, 136). I agree. However, one may fairly wonder why think that the way Alvin forms his beliefs is not one of his cognitive abilities? After all, it is part of his brain.

³⁶ (Pritchard 2009, 48)

³⁷ Adapted from Pritchard (2010*c*, 136) who attributes it to Plantinga (1993).

My best response would be that Alvin's belief-forming process fails to satisfy the first and the second 'glue and trust' criteria. That is, Alvin's belief-forming process 1) is not one of his dispositions (i.e. one of his habitual cognitive routines) and thereby 2) it cannot have been deemed trustworthy.³⁸

Next, Pritchard moves back to the Temp case and asks how would our intuitions change if Temp came to know what is the true source of the reliability of his belief-forming process and that it is reliable. Pritchard (2010c, 138) argues that it would make a great difference because in becoming aware of the source of the reliability Temp can now take cognitive responsibility for this cognitive success. Accordingly, his cognitive success is primarily creditable to his cognitive agency and thereby, by Pritchard's suggestion,³⁹ the relevant belief-forming process has been appropriately integrated within his cognitive character. Therefore, Temp can gain knowledge in this way.

At this point, however, the 3+1 criteria and the results of Pritchard's suggestion to regard a belief-forming process as appropriately integrated within one's cognitive character when the cognitive success is significantly creditable to one's cognitive agency come apart. For, in the way Temp forms his beliefs, no continuous reciprocal causation is being observed between Temp and the thermometer on the wall. Instead, the thermometer delivers information to Temp in a one-way causal—as opposed to constitutive—dependence. Notice I am not saying that Temp cannot gain knowledge when he knows the true source of the reliability of his belief-forming process and that it is reliable. He can. I am rather saying that this is not a case of extended cognition. Instead, a more appropriate epistemic description of this case would be in testimonial terms, whereby Temp has positive reasons to accept and/or no undefeated defeaters to deny the reading of the thermometer.⁴⁰

³⁸ Notice that Alvin's lesion is not one of the belief-forming processes he would employ was he motivated to believe the truth. Or, in other words, forming beliefs on the basis of the brain lesion cannot be deemed trustworthy. Alvin cannot automatically endorse the products of this process; it's not the case that, from Alvin's 'point of view', the resource is not *usually* subject to critical scrutiny simply because it is a recently acquired one.

³⁹ Recall that according to Pritchard when some cognitive success is significantly creditable to one's cognitive agency, then the belief-forming process by which the belief was acquired can be said to have been appropriately integrated within one's cognitive character.

⁴⁰ Notice, then, that while it is the case that whenever CRC takes place the cognitive success will be significantly creditable to one's cognitive agency, things do not work the other way around.

Next, however, Pritchard (2010c, 138) asks the same question about Alvin: what if Alvin comes to know both what the true source of the reliability of the belief-forming process is and that it is reliable? Pritchard rightly argues that now Alvin can gain knowledge. In so being aware of the facts he can now satisfy the first and second ‘glue and trust’ criteria.⁴¹ After all it would be very implausible to deny CRC in this case. Alvin’s lesion is to be found within his brain and so, most likely, the affected area interacts very densely with the other parts of his brain in order to produce the relevant outputs. However, becoming aware of the source of the reliability of a belief-forming process is a very strong condition on knowledge and it is exactly what externalist approaches such as (robust) virtue reliabilism set out to resist. Accordingly, Pritchard goes on to further explore whether one’s belief-forming processes can be integrated within one’s cognitive character in weaker ways.

According to Pritchard (2010c 146), one factor that seems to play a central role regarding our intuitions on the integration of a belief-forming process is whether it has always been present or whether it was added at a later juncture. Thus, Pritchard prompts us to imagine Tempo who is fitted from birth with a highly reliable device, which records the ambient temperature. Moreover, Tempo has grown up in a society where it is completely natural for one to consult the temperature-recording device in order to form beliefs about the ambient temperature.

Pritchard claims that “interestingly, in a case like this it seems entirely unnecessary for Tempo to know that this is a reliable belief-forming process or what the source of the reliability before he gain knowledge via this process” (2010c, 146). I think he is right. Forming beliefs in this way is a disposition for Tempo and, plausibly, there is CRC involved. Having always been fitted with the device Tempo has a practical understanding of how his actions will affect his temperature beliefs and vice versa. For instance, he has a practical understanding that when he goes closer to a heater or the fireplace the quick silver is supposed to rise and that when he moves away it will drop. Or, that if the temperature changes while he is sitting still then some warm object must be near by, or a window has been opened.⁴² The temperature-recording device is not just a thermometer on the wall for Tempo. Instead, he is

⁴¹ That is, in being aware of the facts, Alvin has surely deemed his belief-forming process reliable and, as time goes by, it can start being one of his dispositions/habits.

⁴² Put another way, Tempo has acquired knowledge of the sensorimotor contingencies that accompany his continuous interaction with the device.

continuously interacting with the device as he moves around generating outputs which recycled as inputs drive his cognitive character along with respect to his temperature-movement related beliefs. Moreover, through all these interactions, it could be safely assumed that the device has been deemed reliable. Had the device told Tempo that it is cold while he is next to the fireplace, he would not have trusted it. Therefore, the device has been appropriately integrated within his cognitive character even though Tempo might not even be aware of its existence.

But then Pritchard asks what if the agent is fitted with the device at a later stage? So, imagine Tempo* who comes out of a comma with this device fitted, while being non-culpably unaware that this device has been artificially implanted in him (Pritchard 2010c 148). Such a change, argues Pritchard, cries out for the agent to take a reflective stance on the epistemic standing of this change, and in its absence we cannot say that Tempo* can gain knowledge on the basis of his newly fitted device. Interestingly, however, Pritchard claims that as time goes by this intuition lessens. For if Tempo* has been fitted with the device, say for three years “there is now a track-record of beliefs formed via this process which have generally cohered with the beliefs formed via Tempo*’s cognitive abilities (and if they hadn’t cohered, we may suppose, then Tempo* would have spotted this and responded accordingly)” (Pritchard 2010c, 148). That is, the new belief-forming process has now both become a disposition for Tempo* and has been deemed trustworthy. Plausibly, moreover, within these three years Tempo* has become able to reciprocally interact with the device such that it can count as his cognitive ability. In other words, given that Tempo* has been fitted with the device for a sufficiently long period of time, our intuitions on his ability to acquire knowledge on its basis become more supportive, as he may have plausibly satisfied the 3+1 criteria.

Before closing this section, one last remark is in order here. When discussing the Temp case where the thermometer is hanging on the wall and Temp knows what the source of the reliability is, I claimed that it would better be analyzed in testimonial terms. So, one might fairly wonder why not try to analyze any instance in which an agent is in cognitive contact with some epistemic artifact in testimonial terms? Although this might seem a promising strategy, it is not going to work for the same reason that HEMC cannot account for every case of an agent’s deployment of an

artifact. For, in many cases, the agent's true belief depends very deeply on his ongoing interaction with the epistemic artifact in such a way that no causal explanation of how the agent *formed* his true belief will be possible or available in testimonial terms.

Consider, for instance, telescopic observation. Making observations through a telescope is a dynamic process that requires a great deal of experience on how to manipulate the artifact and a great amount of background knowledge in order to understand what one is looking at. By moving around the telescope while adjusting the lenses, the agent delivers outputs (shapes on the lens of the telescope) which recycled as inputs drive the agent's cognitive character along, so as to identify—that is see—recognizable objects in the space (e.g. stars, planets, comets, galaxies et cetera). The epistemic artifact actively drives the agent's cognitive mechanism in a continuous and highly interactive way. Therefore, agent and telescope should be considered as a coupled system, and the overall process as a case of extended cognition and not as merely a case of an agent using an instrument. That is, the interaction between the agent and the telescope is not linear such that the two systems can be neatly decomposed by having their function described in terms of distinct inputs and outputs from the one to the other. Accordingly, it would be a vain attempt to analyze one's knowledge of stellar facts in terms of (artificial) testimony, whereby the telescope provides the agent with fully articulated pieces of information, which she must accept or deny.

Conclusion

I have argued that COGA_{weak} and HEC are complementary accounts. On one hand, COGA_{weak} is an attempt to accommodate the ability intuition on knowledge (i.e. knowledge must be the product of cognitive ability). HEC, on the other hand, is an attempt to recognize a process as part of one's cognitive system despite whether it is wholly realized within an agent's head, or not. Interestingly, the criteria set forth by COGA_{weak} in order for a belief-forming process to count as knowledge-conducive are the same as the criteria suggested by HEC in order for a process to be part of one's overall cognitive mechanism (i.e. be one's cognitive ability). I have been referring to them as the 3+1 criteria: i) the process must be reliably available and typically invoked (i.e. it must be one of the agent's dispositional/habitual cognitive routines),

as such ii) it must be more-or-less automatically endorsed and not usually be subject to critical scrutiny (i.e. it must be trustworthy/reliable), iii) it should also be easily accessible as and when required (i.e. it must have been appropriately integrated within the agent's overall cognitive mechanism/character), and iv) there must be a continuous reciprocal causation between the target process and the agent's natural cognitive faculties.

This agreement on the fundamental tenets of the two views should be interesting on its own. However, it further provides us with a principled account of the way in which epistemic agents may extend their knowledge-conducive cognitive characters beyond their natural cognitive faculties by incorporating epistemic artifacts. Such an account, I claimed, is necessary for any formulation of the ability intuition on knowledge, which wishes to explain our acquisition of knowledge on the basis of epistemic artifacts.

Finally, to return to the question this paper opened with, does Travis *know* there is a chair in front of him? Provided that Travis has been equipped with the TVSS for a sufficiently long period of time, perceiving the environment through it should count as one of his dispositions/habits, which he must have deemed trustworthy even if he may know nothing about the true source of the reliability. Moreover, given that the mechanism is constantly—or, say, every morning—attached to his body he has acquired knowledge of the relevant sensorimotor contingencies on the basis of which he may continuously and reciprocally interact with it; that is, the TVSS has been appropriately integrated within his cognitive character. Therefore, the 3+1 criteria are indeed satisfied and so, according to virtue reliabilism and the underlying ability intuition on knowledge, Travis can be knowledgeable with respect to his environment, despite the fact that his knowledge-conducive cognitive ability does not lie entirely in his head.

References

- Adams, F., & Aizawa, K. (2008). *The Bounds of Cognition*, Blackwell Publishing Ltd.
- (2010). 'Defending the Bounds of Cognition'. In *The Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.

- Audi, R. (1998). *Epistemology: A Contemporary Introduction to the Theory of Knowledge*, London: Routledge.
- Bach-y-Rita, P., and S. W. Kercel. (2003). 'Sensory substitution and the human-machine interface', *Trends in Cognitive Science* 7, no.12 541-6.
- Burge, T. (1993). 'Content Preservation', *The Philosophical Review*, 102, 457-88.
- (1986). 'Individualism and psychology', *Philosophical Review*, 95: 3-45.
- Clark, A., and Chalmers, D. (1998). 'The Extended Mind', *Analysis* 58, no. 1: 7-19.
- Clark, A., (1998). 'Magic Words, How Language Augments Human Computation'. In *Language and Thought: Interdisciplinary Themes*, P. Carruthers and J. Boucher (Eds) (Cambridge University Press: Cambridge, 1998)
- (2001). 'Reasons, Robots, and the Extended Mind', *Mind and Language*, 16;2: 121-145.
- (2006). 'Soft selves and Ecological Control'. In *Distributed Cognition and the Will*, D. Spurrett, D. Ross, H. Kincaid and L. Stephens (eds). MIT Press, Camb. MA
- (2007). 'Curing Cognitive Hiccups: A Defence of the Extended Mind', *The Journal of Philosophy*, 104: 163-192. Also available at <http://hdl.handle.net/1842/1719>.
- (2008). *Supersizing The Mind*. Oxford University Press.
- (2010a). 'Memento's Revenge: The Extended Mind, Extended'. In the *Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- (2010b). 'Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aizawa'. In *The Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- Dretske, F. (1970). 'Epistemic Operators', *Journal of Philosophy* 67, 1007-23.
- Faulkner, P. (2000). 'The Social Character of Testimonial Knowledge', *The Journal of Philosophy* 97, 581-601.
- Fricke, E. (1994). 'Against Gullibility'. In *Knowing from Words*, (1994) B. K. Matilal & A. Chakrabarti (eds.), 125-61, Dordrecht: Kluwer.
- Gettier, E. (1963). 'Is Justified True Belief Knowledge?', *Analysis* 23, 121-3.

- Goldman, A. (1976). ‘Discrimination and Perceptual Knowledge’, *Journal of Philosophy* 73, 771-91.
- Greco, J. (1999). ‘Agent Reliabilism’, *Philosophical Perspectives*, 13: Epistemology (1999) James Tomberlin (ed.), Atascadero, CA: Ridgeview Press, 273-296.
- (2004). ‘Knowledge As Credit For True Belief’. In *Intellectual Virtue: Perspectives from Ethics and Epistemology*, M. DePaul & L. Zagzebski (eds.), Oxford: Oxford University Press.
- (2007) ‘The Nature of Ability and the Purpose of Knowledge’, *Philosophical Issues* 17, 57- 69.
- (2008) ‘What’s Wrong with Contextualism?’, *The Philosophical Quarterly* 58, 299-302.
- (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press.
- Hume, D. (1977). *An Enquiry Concerning Human Understanding*, E. Steinberg (ed.), Indianapolis: Hackett.
- (1975). *An Enquiry Concerning Human Understanding and concerning the Principles of Morals*, third edition, L.A. Selby-Bigge, ed., (Oxford: Clarendon Press, 1975)
- Hurley, S. (2010). ‘The Varieties of Externalism’. In the *Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- (1998). *Consciousness in Action*. Cambridge, MA: Harvard University Press.
- Kelp, C. (forthcoming) ‘Extended Cognition and Robust Virtue Epistemology’.
- Lackey, J. (2008). *Learning From Words: Testimony as a Source of Knowledge*, Oxford: Oxford University Press.
- (2007). ‘Why We Do not Deserve Credit for Everything We Know’, *Synthese* 158, 345-61.
- Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programmes. In *Criticism and the Growth of Knowledge*, edited by Imre Lakatos and Alan Musgrave, Cambridge University Press, 1970.
- Logan, K. R. (2003). The Extended Mind: Understanding Language and Thought in Terms of Complexity and Chaos Theory. In *Humanity and the Cosmos*, Daniel McArthur & Cory Mulvihill (eds). Also available at <http://www.upscale.utoronto.ca/PVB/Logan/Extended/Extended.html>

- Menary, R. (2006). 'Attacking the Bounds of Cognition', *Philosophical Psychology*. Vol. 19, No. 3, June 2006, pp. 329-344.
- (2007). *Cognitive Integration: Mind and Cognition Unbound*. Palgrave MacMillan.
- (2010a). 'Introduction: The Extended Mind in Focus'. In *The Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- (2010b). 'Cognitive Integration and the Extended Mind', In the *Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- Noë, A. (2004). *Action in Perception*. Cambridge, MA:MIT Press.
- Nozick, R. (1981). *Philosophical Explanations*, Oxford University Press, Oxford.
- Plantinga, Alvin. (1993). *Warrant and Proper Function*. New York: Oxford University Press.
- Pritchard, D. H. (2010a). 'Anti-Luck Virtue Epistemology', manuscript, available at <http://www.philosophy.ed.ac.uk/people/full-academic/duncan-pritchard.html>.
- (2010b). 'Knowledge and Understanding', in A. Haddock, A. Millar & D. H. Pritchard, *The Nature and Value of Knowledge: Three Investigations*, Oxford: Oxford University Press.
- (2010c). 'Cognitive Ability and the Extended Cognition Thesis', *Synthese*, 175: 133-151.
- (2009). *Knowledge*, London: Palgrave Macmillan.
- (2008). 'Sensitivity, Safety, And Anti-Luck Epistemology'. In *The Oxford Handbook of Skepticism*, (ed.) J. Greco, (Oxford: Oxford University Press).
- (2002). 'Resurrecting the Moorean Response to Scepticism', *International Journal of Philosophical Studies* 10, 283-307.
- Putnam, H. (1975). 'The Meaning of "Meaning"'. In K. Gunderson (ed.), *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press.
- Reid, T. (1983). *Essay on the Intellectual Powers of Man*, in R. E. Beanblossom & K. Lehrer (eds.), *Thomas Reid's Inquiry and Essays*, Indianapolis: Hackett.
- Roberts, T. (*manuscript*). 'Taking Responsibility for Cognitive Extension'.
- Ross, D. and Ladyman, J. (2010). 'The Alleged Coupling-Constitution Fallacy and the Mature Sciences'. In *The Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.

- Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes*. New York: Cambridge University Press.
- (2009). ‘Extended Cognition and the Mark of the Cognitive’. *Philosophical Psychology*, 22(1); 1-19.
- Rupert, D. R. (2004). ‘Challenges to the Hypothesis of Extended Cognition’. *Journal of Philosophy*, 101: 389-428.
- (2009). *Cognitive Systems and the Extended Mind*, Oxford University Press.
- (2010). ‘Representation in Extended Cognitive Systems: Does the Scaffolding of Language Extend the Mind?’. In the *Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- Sosa, Ernest. (1988). ‘Beyond Skepticism, to the Best of our Knowledge’, *Mind*, New Series, vol. 97, No.386, 153-188.
- (1993). ‘Proper Functionalism and Virtue Epistemology’, *Nous*, Vol. 27, No. 1, 51-65.
- (1999). ‘How to Defeat Opposition to Moore’, *Philosophical Perspectives*, 13, 141-54.
- (2000). ‘Skepticism and Contextualism’, *Philosophical Issues* 10, 1-18.
- Vaesen, K. (2010). ‘Knowledge Without Credit, Exhibit 4: Extended Cognition’. *Synthese*.
- Van Gelder, T. (1995). ‘What Might Cognition Be, If Not Computation?’, *The Journal of Philosophy*, Vol. 92, No. 7 (Jul., 1995), pp. 345-381.
- Varela, F (1979). *Principles of Biological Autonomy*. New York. North-Holland.
- Weiner, M. (2003). ‘Accepting Testimony’, *The Philosophical Quarterly* 53, 256-64.
- Wheeler, M. (2004). ‘Is Language the Ultimate Artifact?’, *Language Sciences* 26: 693-715.
- Wilson, R. A. (2000). ‘The mind beyond itself’. In D. Sperber (ed.), *Metarepresentations: A Multidisciplinary Perspective*, New York University Press, 31-52.
- (2004). *Boundaries of the Mind: The individual in the Fragile Sciences: Cognition*. New York: Cambridge University Press.