Active Externalism, Virtue Reliabilism and Scientific Knowledge

(Forthcoming in *Synthese*)

ACTIVE EXTERNALISM, VIRTUE RELIABILISM AND SCIENTIFIC KNOWLEDGE

University of Edinburgh

S. Orestis Palermos

Abstract: Combining active externalism in the form of the extended and distributed cognition hypotheses with virtue reliabilism can provide the long sought after link between mainstream epistemology and philosophy of science. Specifically, by reading virtue reliabilism along the lines suggested by the hypothesis of extended cognition, we can account for scientific knowledge produced on the basis of both hardware and software scientific artifacts (i.e., scientific instruments and theories). Additionally, by bringing the distributed cognition hypothesis within the picture, we can introduce the notion of epistemic group agents, in order to further account for collective knowledge produced on the basis of scientific research teams.

1. INTRODUCTION

Given that knowledge is a cognitive phenomenon, we should expect that philosophy of science, epistemology and cognitive science be interrelated in some way. In fact, in a series of recent papers, Ronald Giere (Giere 2002*a*; 2002*b*; 2006*;* 2007; Giere & Moffat 2003) has attempted to bring cognitive science and philosophy of science together. As far as the field of epistemology is concerned, however, Giere makes the following remark: "Philosophy of science illuminates the problems of epistemology but not much the other way around" (Giere 2002*b*). This may be too strong, but the truth remains; philosophy of science and mainstream epistemology have so far been at odds—an awkward situation owing to the fact that the latter discipline has traditionally been individualistic whereas the former has been for the most part socially oriented.[1]

[1] Apart from some very few exceptions pointing towards the opposite direction (Goldman 2004, 2010; Fuller 2007, 2012), for the most part, even the emerging field of *social epistemology* has so far focused on the individual epistemic subject, by studying only how *individualistic* knowledge is affected by social factors. For a general discussion of the field of social epistemology, the debate over its methodology, and its future see (Palermos & Pritchard 2013). For an extended discussion on the compatibility of contemporary approaches within cognitive science with contemporary accounts of knowledge and justification within mainstream and social epistemology, see (Carter et al. 2014). Non-mainstream approaches to social epistemology (e.g., Science and Technology Studies and the Sociology of Scientific Knowledge (Barnes, Bloor, & Henry, 1996; Bloor, 1991; Latour, 1999, 2007; Latour & Woolgar, 1986)) can, no doubt, also prove relevant to bridging the gap between epistemology and philosophy of science and cognitive science. In what, follows, however, we will need to bracket such approaches in order to save space for the discussion of mainstream epistemology, which is the primary focus of the present paper.

Why this methodological tension? In conceiving of knowledge—typically the primary focus of epistemology—as a cognitive (i.e., mental) phenomenon, mainstream epistemology has traditionally focused on the individual cognitive agent and the cognitive processes that lie under her skin. Cognition after all—it is largely held—rests within the individual's head. Accordingly, to account for knowledge, one should focus on the cognitive/epistemic properties of the individual agent. To the contrary, far from being an individualistic enterprise that rests solely on the internal mental powers of individual scientists, the process of science relies heavily on scientific instruments and social structures and institutions (hardly anyone can deny the social and artificially scaffolded nature of the scientific process, especially after the publication of Kuhn's *The Structure of Scientific Revolutions*, in 1962).

Indeed, it may even be argued that the gap between scientific knowledge (which heavily relies on scientific instruments and collaborative research teams) and the rest of the processes of knowledge-acquisition that mainstream epistemology has traditionally focused on (e.g., perception, memory, testimony and so on) is so big that no single, unified epistemological approach could account for them all—maybe, it could be further argued, there are several kinds of propositional knowledge and some of them require special treatment. However compelling this may sound to some, here we will pursue the exact opposite line of thought; i.e., our epistemic intuitions pick out only one concept of knowledge that operates in every acknowledgement of a true belief as an instance of knowledge, and this is so, despite the fact that knowledge can be attained via disparate processes whose (physical) implementation may be entirely unrelated. Granted, in most cases, knowledge supervenes on flesh and grey matter alone; but other times, it arises out of the interaction of our bodies with epistemic artifacts, and occasionally it may even be the product of several organisms and their tools operating in tandem. Despite the wild disparity of the underlying processes, however, all of these cases result in a single type of knowledge and epistemology must find a way to abstract away from irrelevant considerations of physical implementation in order to capture what all these cases have in common.

In effect, given the dissimilitude between individualistic and scientific knowledge, the ability to capture the latter can be thought of as a critical test for the adequacy of any mainstream account of knowledge. As we shall see, however, the trick for providing such a widely encompassing account of knowledge might not come from epistemology itself, but from doing epistemology with a different approach to cognition in mind—one that allows for cognition to extend or even be distributed to the social domain.

We can bring the above points together by considering Giere again. Giere's main goal in the aforementioned list of publications is to demonstrate how certain forms of externalist philosophy of mind, known as active externalism (and especially the extended and distributed cognition hypotheses), can make cognitive science pertinent to philosophy of

science (despite the fact that the former, just like traditional epistemology, has focused for the most part on the individual). Here, we will focus on the potential impact of externalist philosophy of mind on philosophy of science once again. The difference of the present approach, however, is that, this time, active externalism will act as a link between mainstream epistemology and philosophy of science, in order to provide a *mainstream epistemological analysis* of certain important aspects of the scientific process. Specifically, it will be argued that if we read virtue reliabilism—one of the most promising mainstream accounts of knowledge—along the lines suggested by the extended and distributed cognition hypotheses, we can reduce the gap between mainstream epistemology and philosophy of science in at least two ways.

In some more detail, in section 2, we will introduce active externalism and virtue reliabilism, and we will explain how we can combine the two. In section 3, we will take advantage of this move to demonstrate how virtue reliabilism can account for knowledge produced on the basis of both hardware and software scientific artifacts (i.e., scientific instruments and theories), and we will discuss how this approach can reveal the hidden, social nature of certain instances of scientific knowledge that are normally thought to be solely down to the individual scientist. Finally, section 4 will introduce the idea of epistemic group agents in order to account for collective knowledge produced on the basis of scientific research teams.

2. VIRTUE RELIABILISM AND ACTIVE EXTERNSALISM

2.1 Active Externalism

As a general approach to the nature of mind, active externalism (Clark and Chalmers 1998; Clark 2007, 2008; Hutchins 1995; Theiner 2011; Wheeler 2005; Menary 2006, 2007; Rowlands 1999; Wilson 2000, 2004) is standardly contrasted with Putnam (1975) and Burge's (1986) meaning, or passive externalism, as it concentrates on the aspects of the environment that *drive* one's cognitive loops in an ongoing way. Focusing on the specifics, active externalism has appeared in the literature under several labels and formulations—e.g., the extended mind thesis (Clark and Chalmers 1998), cognitive integration (Menary 2007), environmentalism (Rowlands 1999), locational externalism (Wilson 2000, 2004), the hypothesis of extended cognition (Clark and Chalmers 1998), the hypothesis of distributed cognition (Hutchins 1999) and so on. Here, however, we will only concentrate on the latter two.

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, sec. 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 (we will return to this example in §3), or perceiving a chair through a tactile visual substitution system.[2] 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 hypothesis of distributed cognition (Hutchins, 1995; Theiner et al., 2010; Theiner and O'Connor, 2010; Sutton et al., 2008; Wilson, 2005; Heylighen et al., 2007) may sound more challenging still. According to this form of active 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 more 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.

With respect to argumentative lines, active externalism, especially in the form of the extended mind thesis, has been traditionally associated with common-sense functionalism (Braddon-Mitchell & Jackson, 2006). It has been recently argued (Chemero 2009, Palermos 2014*a,* Carter et al. 2014), however, that contrary to the extended mind thesis, the focus of the extended and distributed cognition hypotheses is not on mental states (such as beliefs and desires, understood in common-sense functionalist terms), but on extended (and distributed) *dynamical* cognitive processes and the overall cognitive *systems* these processes give rise to. Accordingly, the extended and distributed cognition hypotheses do not need to rely for their support on common-sense functionalism; instead, they can be motivated on the basis of Dynamical Systems Theory (DST)—perhaps, the most powerful, if not the only, mathematical framework for studying the behavior of dynamical systems, in general.[3]

According to this conceptual framework, in order to claim that two (or more) systems give rise to some extended or distributed process and, thereby, to an overall extended or distributed system (either way, to a *coupled* system, in DST terms), what is required is the existence of non-linear relations that arise out of *continuous reciprocal interactions* between the contributing parts (Chemero, 2009; Froese et al., 2013; Sutton et al., 2008; Theiner et al., 2010; Theiner & O'Connor, 2010; Wegner et al., 1985; Tollefsen & Dale, 2011; Palermos 2014*a*). This is because the aforementioned non-linear relations give rise to an overall non-

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

[3] See also (Shani, 2013), whose view—*viz.,* moderate active externalism—is similar to what we here call the hypothesis of extended cognition (though note that Shani's arguments do not so heavily rely on DST, and his view is stronger than the hypothesis of extended cognition in that it denies—instead of remaining silent on the matter—the extension of (common-sense functionalist) mental states. For more details on why common-sense functionalism is necessary for the extended mind thesis, but not the extended and distributed cognition hypotheses, see (Palermos 2014*a*). Again though, note that the hypotheses of extended and distributed cognition are neither incompatible with common-sense functionalism, nor anti-functionalist on the whole. In so far as a cognitive process is a function, these two hypotheses *are* compatible with functionalism.

decomposable system that consists of all the contributing subcomponents. In some more detail, two reasons for postulating the overall system is that the aforementioned non-linear interactions (1) give rise to new systemic properties that belong only to the overall system and to none of the contributing systems alone (therefore one *has to* postulate the overall extended or distributed system) and (2) prevent one from decomposing the two systems in terms of distinct inputs and outputs from the one subsystem to the other (therefore one *cannot but* postulate the overall system) (Palermos 2014*a*).[4] Accordingly, on the basis of dynamical systems theory, we can claim that in order to have an extended or even distributed cognitive system—as opposed to a cognitive system that is merely embedded in the sense of being dependent on, but not constituted by, certain environmental aspects (cf. Adams & Aizawa, 2001, 2010; Rupert, 2004, 2009)—all we need is that the contributing parts (i.e., the relevant cognitive agents and their artifacts) interact continuously and reciprocally with each other.[5]

2.2 Virtue Reliabilism

To introduce active externalism within contemporary epistemology we need an account of knowledge that places in its center the notion of cognitive ability, but in a way that is neutral as to whether cognitive abilities are supposed to be realized within the agent's organismic boundaries or not. As luck would have it, however, there is already such an account on offer, *viz.*, virtue reliabilism (see Greco 1999, 2004, 2007, 2010; Pritchard 2010*b;* Palermos & Pritchard 2013; Palermos 2011; Palermos 2014*b*; Palermos *forthcoming*).[6]

According to virtue reliabilism, knowledge is creditable true belief, which is *creditable because it is true in virtue of the manifestation of cognitive ability*. On this view, cognitive ability is understood as a reliable belief-forming process that has been appropriately integrated into the agent's cognitive character, where the agent's cognitive character mainly consists of the agent's cognitive faculties of the brain/central nervous system (CNS), including her natural perceptual faculties, her memory, and the overall doxastic system. In addition, however, it can also consist of "acquired skills of perception and acquired methods of inquiry including those

[4]Note that the argument of this approach moves from the existence of extended/distributed cognitive processes to the existence of extended/distributed cognitive systems. There are some alternative approaches to active externalism available in the literature, however, that seem to focus either on the relevant extended systems (e.g., (Gelder, 1995; Haugeland, 1993, 2000) or on the relevant extended processes (Menary, 2013), alone.

[5] To preempt a possible worry, here, the relevant reciprocal interactions need only be continuous during the operation of the relevant coupled cognitive system and the unfolding of any processes related to it. For example, if, as part of her job and during normal working hours, individual *S* participates in distributed cognitive system *X*, *S* does not need to continuously interact with the other members of *X,* when she is at home. However, whenever *X* is in operation, *S* must continuously and reciprocally interact with the rest of the *X*-members. For more details on how dynamical systems theory, in general, can help us clearly distinguish between the hypothesis of extended cognition and the hypothesis of embedded cognition as well as avoid several other worries with respect to the hypothesis of extended cognition (e.g., the 'cognitive bloat' worry and the 'causal-constitution' fallacy), see (Palermos 2014*a*).

[6] There are several other proponents of virtue reliabilism—most famously Sosa (1988; 1993; 2007). The reason why only the above references have been included in the main text is to indicate a specific lineage of virtue reliabilism that is particularly apt for our present purposes. In the beginning of the line, however, is Greco, who has, himself, been heavily influenced by Sosa's alternative.

involving highly specialized training or even advanced technology" (Greco 1999, 287).[7] Here is a relatively weak formulation of virtue reliabilism we can work with:

> *$COGA_{weak}$*
>
> 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 2010*b*, 136-7). [8]

The reason why virtue reliabilists turn to an account of knowledge that stresses the creditable nature of the cognitive success (i.e., believing the truth) as well as its origin in the agent's cognitive ability has to do with knowledge-undermining epistemic luck involved in Gettier cases. 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 a defective justificatory process, *just happens* to be true for reasons that are extraneous to one's justification: In a lucky turn of events, one's belief, which would otherwise be false (given it is produced in a defective way), turns out to be true. Contrast this with cases of success through the manifestation of 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 true belief is the product of the manifestation of one's ability then believing the truth cannot have been lucky; of course, one may still be lucky to believe anything at all (because, say, one could have easily been killed), but believing the *truth* is not lucky itself. Accordingly, and since credit is normally attributed in cases of success through ability, virtue reliabilists hold that when some agent knows, his belief must be *true because of cognitive ability*, such that the success be creditable to him.

In other words, virtue reliabilists want to accentuate the importance of the way one arrives at one's *true* belief, i.e., the *process of getting things right*. It is not enough that one forms one's belief on the basis of virtue (i.e., ability) *and* that one's belief be true: The mere

---

[7] It should be noted that the notion of one's 'cognitive character' as employed by virtue reliabilists is a technical notion, whose meaning may seem counterintuitive to our common-sense understanding of one's character as consisting of general traits, such as open-mindedness, wisdom, courage, honesty, understanding, empathy and the like. In fact, within mainstream epistemology, there are two main ways in which one can understand intellectual virtues and character: 1) in terms of general 'trait-virtues', such as open-mindedness, honesty, understanding and so on and 2) in terms of specialized 'faculty-virtues', such as memory, perception, reasoning and so on. (Zagzebski, 1996) is a typical example of the former approach and (Greco, 1999, 2010) of the latter. For the distinction between these two approaches to intellectual virtues and thereby character, see (Baehr, 2006) and (Greco & Zagzebski, 2000). Since it is the second approach to virtues—in terms of faculties and cognitive abilities—that informs the virtue reliabilist's sense of one's 'cognitive character', in what follows, the agent's 'cognitive character' refers to the agent's set of cognitive faculties (as opposed to general, 'trait-virtues').

[8] This is a weak formulation of virtue reliabilism for two reasons. First, because it is only a necessary condition on knowledge (several epistemologists hold that virtue reliabilism is a necessary component, but to have an adequate theory of knowledge, they argue, it must be further supplemented by either the safety or the sensitivity principle. (Pritchard 2010*a*)). Second, because, in order to also accommodate testimonial knowledge (Pritchard 2010*b*), it requires that one's cognitive success be significantly, as opposed to primarily, creditable to one's cognitive agency. Accordingly, '$COGA_{weak}$' stands for 'weak COGnitive Agency' to indicate that this is an account of knowledge, which requires that one's cognitive success be creditable to one's cognitive agency only to a significant (as opposed to primary) degree. For more details, see (Pritchard 2010*b*).

conjunction of these two conditions does not preclude Gettier cases from counting as knowledge. Virtue reliabilists, instead, focus on the *relation* between these two conditions. In order to know, getting to the truth of the matter must be creditable to one and for that to be the case, one must believe the truth *because of* one's cognitive ability. Put another way, it is only when one's true believing reveals the manifestation of one's ability that one's *true* belief can be creditable to one (and thereby constitute knowledge). Therefore, it should be no surprise that—and we should mark this to better appreciate the arguments to follow—virtue reliabilists put particular weight on the *process via which one arrives at the truth* (as opposed to merely believing something that also is, or happens to be, true).

Now, as we mentioned before, according to virtue reliabilism, in order for a belief-forming process to count as a cognitive ability it must be part of the agent's cognitive character. So what could it be required in order for a process to be so integrated? As far as common-sense intuitions are concerned, Greco (1999, 2010) has noted that the relevant belief-forming process must be neither strange nor fleeting (i.e., it must be a normal, dispositional cognitive process). Despite such broad intuitions, however, Greco has noted in later work (2010) that in order for a process to be appropriately integrated within one's cognitive character it must interact cooperatively with it. Specifically he writes: "cognitive integration is a function of cooperation and interaction, or cooperative interaction with other aspects of the cognitive system" (2010, 152).

The reason why Greco spells out 'cognitive integration' and 'cognitive character' in this way has to do with a minimal notion of epistemic responsibility/subjective justification. Specifically, Greco is after a notion of subjective justification, which is inline with epistemic externalism in that it denies that in order to be subjectively justified/epistemically responsible one needs to have access to the reasons for which one's beliefs are reliable. Unfortunately, going into the details of how the integrated nature of one's cognitive character can allow one to be justified in absence of any positive reasons for one's belief is beyond the scope of the present paper, but the main idea is this:[9] If one's belief-forming process cooperatively interacts with other aspects of one's cognitive system, then it can be continuously monitored in the background such that *if* there is something wrong with it, *then* the agent will be able to notice this and respond appropriately. Otherwise—if the agent has no negative beliefs about his/her belief-forming process—he/she can be subjectively justified in employing the relevant process *by default*, even if he/she has absolutely no positive beliefs as to whether or why it might be reliable. In other words, on virtue reliabilism, provided that one's belief-forming process is integrated to one's cognitive character such that one would be in a position to be

[9] For a detailed overview of the epistemic internalism/externalism debate and how it maps onto the internalism/externalism debate within philosophy of mind see [(Carter et al. 2014). For a detailed analysis of the above minimal yet epistemically adequate notion of subjective justification/epistemic responsibility and its relation to cognitive integration see (Palermos 2014*b*).

responsive *were there* something wrong (with the process), one can be subjectively justified in holding the resulting beliefs merely by *lacking* any negative reasons against them.

2.3 Virtue Reliabilism and Active Externalism

It has been previously argued that reading virtue reliabilism along the lines suggested by the extended cognition hypothesis is not only an available option (see Pritchard 2010*b;* Palermos & Pritchard 2013), but actually necessary for accounting for many instances of knowledge acquired via the employment of epistemic artifacts (Palermos 2011; Palermos 2014*b*). Here we will only make a few remarks about the strong compatibility between the two views, as the present goal is to demonstrate how their combination can be turned to our advantage, especially with respect to understanding scientific knowledge acquired on the basis of epistemic instruments.

To start with, first notice that there is nothing in the formulation of $COGA_{weak}$ or in the concepts involved thereof that restricts knowledge-conducive cognitive abilities to processes within the agent's head. To the contrary, the idea of a cognitive character that may consist of "acquired methods of inquiry including those involving highly specialized training or even advanced technology" seems to be compatible with, or even prefigure, the hypothesis of extended cognition.

If we focus on the details of the two theories, however, we can make a much stronger claim. Specifically, both theories put forward the same condition in order for a process to count as part of the agent's cognitive system/character (and, thereby, by the lights of virtue reliabilism, as knowledge-conducive): Just as proponents of extended cognition claim that a cognitive system is integrated when its contributing parts engage in reciprocal interactions (independently of *where* these parts may be located), so Greco claims that cognitive integration of a belief-forming process (be it internal or external) is a matter of cooperative interaction with other parts of the cognitive system.[10]

We see, then, that both in epistemology and philosophy of mind and cognitive science, satisfaction of the same criterion (cooperative interaction with other aspects of the agent's cognitive system) is required for a process to be integrated into an agent's cognitive system and thereby count as knowledge-conducive. Accordingly, there is no principled theoretical bar disallowing extended belief-forming processes from counting as knowledge-conducive. An agent may extend his cognitive character by incorporating epistemic artifacts to it.

[10] Elsewhere (Palermos 2011; 2014*b*), it has been argued that both theories put also forward the same broad, common sense functionalist intuitions on what is required from a process to count as a cognitive ability. Briefly, both views state that the process must be (a) normal and reliable, (b) one of the agent's habits/dispositions and (c) integrated into the rest of the agent's cognitive character/system.

So, for example, in this way, we can explain how a subject might come to know the position of a satellite on the basis of a telescope, while holding fast to the idea that knowledge is belief that is true in virtue of *cognitive ability*.[11] Even though the belief-forming process in virtue of which the subject believes the truth is for the most part external to his organismic cognitive agency, it still counts as one of his cognitive abilities, as it has been appropriately integrated into his cognitive character.[12] Moreover, the subject satisfies $\text{COGA}_{\text{weak}}$, since his believing the truth is significantly creditable to his cognitive agency (i.e., his organismic cognitive apparatus): It is the subject's organismic cognitive faculties that are first and foremost responsible for the recruitment, sustaining, and monitoring of the extended belief-forming process (i.e., telescopic observation), in virtue of which the truth with respect to the satellite's position is eventually arrived at.

In cases like this, therefore, even though it is the external component that accounts (at least in big part) for the truth-status of the agent's belief, the agent's cognitive agency—i.e., his organismic cognitive faculties—is still significantly creditable for integrating and sustaining the relevant external component into his cognitive system. In other words, in accordance with the demands of $\text{COGA}_{\text{weak}}$, even though believing the truth is the product of some extended cognitive process, the agent's cognitive success is still significantly creditable to his organismic cognitive faculties: Just as Clark suggest, "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, sec. 9).

3. THE EXTEDED COGNITION HYPOTHESIS AND SCIENTIFIC KNOWLEDGE

To recap, an epistemic agent's cognitive character—i.e., her cognitive system—may extend beyond her organismic cognitive faculties that make up her cognitive agency, by incorporating belief-forming processes, which rely for the most part on environmental elements. The way to test whether such external processes have been integrated into the agent's cognitive character is to check whether by employing them the agent engages in continuous reciprocal interactions with them—whether, that is, he delivers outputs on their basis, which recycled as inputs drive her cognitive loops along.

---

[11] Making observations through a telescope clearly qualifies as a case of cognitive extension as it is a dynamical process that involves ongoing reciprocal interactions between the agent and the artifact. Moving the telescope around, while adjusting the lenses, generates certain effects (e.g., shapes on the lens of the telescope), whose feedback *drives* the ongoing cognitive loops along. Eventually, as the process unfolds, the coupled system of *the agent and his telescope* is able to identify—that is, see—the target satellite.

[12] On the basis of the feedback loops between the agent and his artifact (see the footnote above).

So depending on whether the above criterion is met, the employment of *hardware* external elements such as calculators, microscopes, telescopes, TVSS and so on can occasionally qualify as genuine hardware extensions of one's cognitive character. It should be interesting, however, to ask whether something analogous applies to what would count as *software* extensions.

3.1 Languages as Software Extensions

From the point of view of the hypothesis of extended cognition, the development of language might have been, to a certain degree, the outcome of the humans' need to externalize their thoughts to the public space so that they can more easily manipulate them.

Drawing on Vygotsky's (1986; 1978) ideas as vindicated by recent bodies of developmental research,[13] Clark suspects that self-directed speech (be it vocal or silent inner rehearsal) is a crucial cognitive tool that allows us to highlight the most puzzling features of new situations, and to direct and control our own problem-solving actions (Clark 1998, 164). Of course, as he further notes, the effect of language on human thought needs not be restricted to speech, since written language may have similar, and possibly more powerful, results.[14] For example, as I write down this paper, Clark would note,

> 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 [paper] 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).

Briefly, the main idea is that language in general, and words in particular, enable us to capture abstract ideas and rich experiences in memory. This has the direct effect of allowing thoughts to become objects of further attention and reflection, opening them up to a range of further mental operations. This *feedback of one's thoughts to one's own cognitive system* gives rise to the distinctively human capacity of meta-cognition, or, as Clark calls it, "second order cognitive dynamics" (1998, 177).

Moreover, this capacity to externalize one's thoughts in *recyclable* linguistic representations could be far more active and transformative than one may initially think, since

---

[13] See (Berk & Garvin 1984). See also (Bamberger & Brofsky, 1979; Olson, 1996a,1996b; Tomasello, 2009; Wertsch, 1988)

[14] Following Olson (2002; 1996a) (who draws on (Goody 1986, 1987) and (Shankweiler & Liberman, 1972)), and contrary to (De Saussure, 1916/1983) and Bloomfield (1933), we should here note that writing may not be the mere transcription of speech, but a language in its own right. For ideas similar to Olson's see (Harris 1986) and Linnell (2005). For a discussion of how they fit into the current debate over active externalism, see (Theiner 2013) and (Theiner 2011, ch. 4).

the particular linguistic abilities one possesses may guide or restrain one's ongoing trains of thought in a profound way. Take the construction of a poem for example:

> We do not simply use the words to express thoughts. Rather, it is often the properties of the words (their structure and their cadence), which determine the thoughts that the poem comes to express. A similar partial reversal can occur during the construction of complex texts and arguments. By writing down our ideas we generate a trace in a format that opens up a range of new possibilities. We can then inspect and re-inspect the same ideas, coming at them from many different angles and in many different frames of mind (Clark 1998, 176).

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. Similarly, Olson (2002, 160) notes that the invention of a notational system involves "the creation of a new conceptual scheme with new possibilities for thinking." Musical scores, for example, allow one to think about musical structure in a new way and, mathematical notations may constitute important building blocks at the foundations of mathematics (Chemla, 2012). Language in other words, does not only facilitate existing mental capacities but creates the possibility for new ones. In Olson's words (*ibid.*, 161), "inventing a new writing system and learning to deal with a writing system is not just a matter of improved storage and communication of information but a new form of representation, thought and consciousness."

Much of the true power of language, in other words, "lies in the underappreciated capacity to re-shape the computational spaces which confront intelligent agents" (Clark 1998). In what ways does language do this? Some of its distinctively *transformative* effects on our biological cognitive systems, as Clark (1998, 169-173) enumerates them, are the following ones: memory augmentation, attention and resource allocation, and manipulation and representation of data.

Interestingly, in a somewhat similar vein, but drawing inspiration from complex systems and chaos theory, Logan (2003; 2006; 2008) presents the controversial yet promising idea that speech is the first proper language embedded in an evolutionary series of languages, preceded by pre-verbal proto-languages (tool making, social intelligence, and mimetic communication), and followed by more task or domain specific languages, such as written language, mathematics and science. According to this picture, each new language emerged from the previous forms of language as a bifurcation to a new level of order in response to an information overload that the previous set of languages couldn't handle. And similarly to Clark, Logan accentuates the transformative effects of words (in the form of both utterances and signs) on our biological faculties:

> A concept in the form of a word links many percepts of an individual and, hence, extends the brain's capacity to remember. Words as concepts are a form of *artificial memory* which creates *artificial connections*. Words bring order to a chaotic mind filled with memories of a myriad of experiences. Language is an emergent order (Logan 2006, 153).

So just like Clark, Logan, too, holds that language serves *two* and not just one fundamental function; obviously, it is a form of (i) communication, but it is a form of (ii) information-processing, as well.

Overall, then, we get the following picture with respect to language: Language allows agents to not only communicate with each other but also think about their own thoughts and transform their minds, on the basis of mutual interactions with linguistic elements (e.g., symbols in the form of utterances or written signs). These self-generated cognitive loops allow agents to extend their cognitive characters beyond the first-order, percept based (at least on Logan's view),[15] cognitive dynamics that their organismic cognitive capacities initially provide them with. But is language or, more accurately, public language and text the only software external artifacts that cognitive agents use so as to extend their cognitive characters in such a way?

3.2 Scientific Theories as Software Extensions

As we briefly mentioned before, Logan claims that

> [S]peech, 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 (2003, 3).

Here we will only concentrate on the interesting case of scientific theories.[16] One way to interpret the above quote is by taking Logan to refer to the (social) *practice* of speech, writing, math and science as languages. On further reflection, however, his remark about the necessity for unique semantics and syntax in order for a system to qualify as a language in itself indicates that it is specific verbal and written languages, as well as particular scientific theories,

[15] In his (2008), Logan claims that our primitive, biological cognitive capacities are all percept-based and that it is only with the advent of language that we acquired concepts. On the face of this it would seem that his approach is a form of concept empiricism. But this may turn out to not stand upon further reflection: First, the debate over empiricism versus rationalism can be construed in several ways ((Markie, 2013), see also (Weiskopf, 2007) for a recent defense of several weak versions of empiricism, as informed by neuroscientific evidence)); second, and against the background of all these dialectical possibilities, Logan stresses the deeply transformative effects of language on thought, indicating that as thought evolves in response to the transformative effects of language, it may develop into something that is more than just its perceptual origins.

[16] In what follows, I do not draw a clear distinction between theories and their models. The two most prevailing views about the relation between models and theory are 1) the syntactic view (associated with the logical positivists) and 2) the semantic view (Suppes 1961; 1967; Suppe, 1977; van Fraasen,1980; Giere, 1988). Both of these views understand the relation between models and theory as a particularly close one. According to the syntactic view, models are dictated by theories, whereas on the semantic view, theories just are families of models. Recently however, Morgan and Morrison (1999) have put forward a third alternative according to which models are 'autonomous agents' in the sense that they are not entirely reliant either on theories or on the world they are meant to represent. For more details, see (Morgan and Morrison 1999, ch. 2).

that he actually refers to as languages. And if Logan is right, such that we can indeed view specific scientific theories as languages and thereby as software artifacts too, we can take this to be a first indication that, according to the hypothesis of extended cognition and virtue reliabilism, scientific theories could also count as belief-forming processes that can extend the epistemic agent's cognitive character beyond his organismic cognitive faculties. In the next section we will start exploring this possibility by considering a further hint on why scientific theories should be seen as software cognitive artifacts that can extend our organismic cognitive capacities. Then we will conclude the argument by considering some specific examples in §3.2.2.

3.2.1 A Hint: Observations are theory-laden

A hint that scientific theories can constitute software cognitive extensions along the lines suggested above comes from the old problem of the theory-ladenness of observations. Briefly speaking, the validity of scientific theories depends on their accordance with empirical observations. It has been claimed, however, that observation involves perception as well as other underlying cognitive processes. As Kuhn claims, "something like a paradigm is a prerequisite to perception itself. What a man sees depends both upon what he looks at and what his previous visual-conceptual experience has taught him to see" (1962, 113). Observations heavily depend on some underlying understanding (which stems from the already existing scientific theories and commonsensical habits of thought) of the way in which the world functions; and that understanding influences what is perceived, noticed, or deemed trustworthy of consideration. Therefore, the argument goes, since empirical observations presuppose a theoretical understanding, they cannot be the final arbiters of the validity of scientific theories.

Historically, the issue first emerged between Hempel (1966; 1970), who defended the distinction between observational and theoretical terms, and Hanson (1961; 1969) who maintained the theory-laden thesis of observation. Specifically, according to Hanson, not only are the observational sentences theory-laden but the observations themselves are theory-laden (1969):

> In short we usually "see" through spectacles made of our past experience, our knowledge, and tinted and molted by the logical forms of our special languages and notations. Seeing is what I shall call a "theory-laden" operation, about which I shall have increasingly more to say.

Famously, the debate also has a counterpart in the philosophy of mind, as it was taken up by Churchland (1979; 1988; 1989) and Fodor (1984; 1988). Fodor, by appealing to illusions such as the Muller-Lyer experiment whereby the subjects' knowledge of the illusion does not alter their defective impressions, thinks that perceptual processes are modular (i.e.,

independent, closed, domain-specific processing modules). So, by definition, bodies of theory that are inaccessible to the modules do not affect the way the perceiver sees the world. Churchland, on the other hand, relying on studies such as those utilizing the ambiguous pictures of rabbit/duck and young/old woman, argues that higher cognitive processes can have an impact on visual processes. Specifically, higher order theories provide the agents with internal representations, which pick out important distinctions and structures in the external world. When the input to the agent's perceptual processes is variegated, or noisy, and thereby not clearly represented, these representations allow the agent to "respond to those inputs in a fashion that systematically reduces the error messages to a tickle. These I need hardly remind, are the functions typically ascribed to "theories"" (1989, 177).

Considerations such as those of Hanson and Churchland have been widely thought to produce a relativistic picture of science—and possibly epistemology as well—whose most prominent proponents are thought to be Feyerabend (1975) and Kuhn (1962) (the latter quite possibly unjustly though). As one of Kuhn's most infamous passages goes: "In so far as [the scientists'] recourse to that world is through what they see and do, we may want to say that after a revolution scientists respond to a different world" (1962, 110). Fortunately, however, modern cognitive psychology points away from relativism, at least as far as the theory-ladenness of observations is concerned—even though the phenomenon is not altogether denied.

In particular, Anna Estany (2001, 208) holds that

> The beliefs of the higher or more fundamental level influence how perceptual units are interpreted by the lower levels [...] Humans use both types of processes in perception because each have characteristic advantages and disadvantages. Thanks to top-down processes we can recognize patterns with incomplete or degraded information. Moreover, top-down processes make perception faster, but they can induce us to make mistakes in a perception by relying on previous knowledge.

Accordingly, our perceptual systems *do* get guidance from higher order expectations. As it has been further pointed out, however, when attention is caused by mismatches between expectation and reality the inputs from the arousal system constitute a "reset wave" making it possible not to fall into arbitrary, relativistic errors of perception (Estany 2008, 213). Similarly, Brewer and Lambert (2001), exploring the literature on relevant experiments, concede that "perception *is* determined by the *interaction* of top-down theory information and bottom-up sensory information" (178, emphasis added):

> However, note that in all of the above cases the stimuli were either ambiguous, degraded, or required a difficult perceptual judgment. In these cases the weak bottom-up information allowed the top-down influences to have a strong impact on perceptual experience. It seems likely that strong bottom-up information will override top-down information. [...] Thus, the top-down/bottom-up analysis allows one to have cases of theory-laden perception, but does not necessarily lead down the slippery slope of relativism.

Arguably, this sounds like a promising response to the question as to whether the theory-laden nature of observations leads to relativism. For the present purposes, however, the solution to this problem is less important than its very existence. For it shows that theories in general, and hence scientific theories too, can take the form of cognitive dispositions whose employment actively drives the agents' cognitive character on the basis of reciprocal (top-down-bottom-up) interactions between the two. This process of mutual interactions creates outputs (observations), which recycled as inputs drive the agent's overall cognitive system along, producing in effect additional observations that will themselves be the basis of new (and hopefully, still empirically testable, as Estany, Brewer and Lambert argue) assumptions and so on. This theory-laden nature of the scientific process may sometimes lead to mistakes, but for the most part, when the input is ambiguous, noisy, or variegated it is an important facilitatory effect that boosts the scientists' performance in several ways. In particular, Brewer and Lambert (*ibid.*) note that background theories affect not only the scientists' perceptual processes but they also play a significant role in other aspects of the scientist's cognitive and epistemic life including attention, data evaluation and interpretation, data production, memory, and communication. Interestingly, but not at all surprisingly on the present view (which takes scientific theories to be akin to languages if not languages themselves) these transformative effects of theories are strikingly similar to the effects that language has on biological cognition, according to Clark (see also §3.1); namely, memory augmentation, attention and resource allocation, and data manipulation and representation (Clark 1998, 169-173).

3.2.2 Extended Scientific Problem-Solving

The discussion of the previous section was only intended as a hint that scientific theories may be seen as software artifacts that extend one's cognitive character. Here is why: The main idea was that once one becomes fluent with a scientific theory, the way one perceives the world (in qualitative terms), the aspects of the world one attends to and considers worthy of consideration, and the way one stores and communicates one's experiences will be fundamentally altered. In other words, scientific theories have a very strong impact on one's point of view of the external world, allowing one to observationally interact with it in ways that would be otherwise unavailable and which, in turn, affect back the way one develops one's theories (and so on). Scientific theories, moreover, are external in the sense that no one is born with them inscribed in one's neural apparatus. Theories, instead, are acquired through a long period of training and practice during which scientists interact with teachers, professors, textbooks, scientific equipment, and so on. And once scientists become masters of such externally derived theories, the cognitive operations (including the observations) they are able to perform are qualitatively altered and significantly enhanced. Hence, scientific theories

may be seen as external software epistemic artifacts that extend one's cognitive character. Here is the wrinkle, however, and why the above can be nothing more than a hint: In order to resist this picture, the opponent of cognitive extension does not need to deny that once such external scientific theories are appropriately internalized will have dramatic effects on the epistemic agent's cognitive loops. Crucially, however, he will further claim that all processing, including making theory-laden observations, will be exclusively performed within the scientist's head. Why, then, should this count as a case of cognitive extension?

Clark, of course, is aware of this line of arguing against languages as software extensions, and this is why he does not restrict himself in claiming that all languages do is facilitate or enhance one's inner processes of thought and reason. Instead, the examples Clark uses in order to illustrate his point involve agents who physically manipulate external linguistic symbols and representations so as to achieve cognitive tasks that would otherwise be infeasible. So are there any examples from the scientific domain, which could motivate the view that scientific theories can count as software artifacts that extend our cognitive capacities in a similar way? [17]

In fact there are plenty, and soon we will start with one from mathematics. Before moving on to the actual example, however, a few preparatory remarks are in order: Drawing on McClelland et al. (1986), Giere and Moffat (2003) claim that human brain networks have evolved for and are best at completing and recognizing patterns in input provided by the environment (i.e., humans are excellent pattern-matchers). But if that's correct, they ask, "how does man do the kind of linear symbol processing required for activities such as using language and doing mathematics"? (Giere and Moffat 2003, 302). "The answer given by McClelland et al. was that man does the kind of cognitive processing required for these activities by creating and manipulating external representations. These latter tasks can be done by a complex pattern-matcher" (ibid.).

So, with the above in mind, here is an example from mathematics: Think about a complex, say, 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 process by using pen and paper to externalize the problem in symbols.

---

[17] As a referee also points out, several authors (e.g.,Goody, 1977; Latour & Woolgar, 1986; Olson, 1996a; Tufte, 2001) have in the past claimed that symbolic representations play a crucial role in the development and visualization of scientific knowledge. These authors, however, have not advanced the strong claim that cognition is literally externalized, rather than *heavily dependent* on external tools and aids, along the lines suggested by the alternative hypothesis of *embedded cognition* (Adams& Aizawa 2001, 2010; Rupert, 2004, 2009). As noted in §2.1, however, the hypothesis of extended cognition should be clearly distinguished from the hypothesis of embedded cognition, which is a conservative view: On the present approach, scientific theories do not act as mere scaffolds of the epistemic agent's internal cognitive capacities, but they are literally constitutive of the cognitive repertoire that makes up the agent's cognitive character. Briefly, the reason, according to dynamical systems theory, is that when the dependence between the cognitive agent and the world is one-way, cognition is indeed merely embedded. When a cognitive system interacts reciprocally with some aspect of the world, however, then the two constitute a coupled system that consists of both of them. It is for this reason that, in what follows, the focus will be on cases whereby the epistemic agent *mutually* interacts in an ongoing way with external scientific symbols.

Then we can serially proceed to its solution by performing simpler multiplications, starting with 9 times 7, and externally storing the results of the process for use in later stages. The process involves eye-hand motor coordination and is not simply performed within the head of the person reciting the times tables. It involves intricate, continuous interactions between brain, hand, pen and paper, all the while it is being transparently regulated by the normative aspects of the notational/representational system involved—for instance, that we cannot multiply by infinity, that we must write the next digit under the second to last digit of the number above, what operation we must perform next and so on.[18]

Accordingly, mathematics (again, a language according to Logan's view) can be seen as a software artifact that allows epistemic agents to literally extend their cognitive characters to the world and thereby transcend their limited organismic cognitive capacities. Of course, it may be objected that mathematics is not really a scientific theory, but the point is that if the above analysis is correct, very similar descriptions can be provided for any solution of some scientific problem, which involves the physical manipulation of external scientific symbols, formulas, or even graphs.

Take, for example, the use of chemical formulas in organic chemistry as introduced by Berzelius and later refined by Dumas, in the early nineteenth century:

> Assuming that the basic constituents in reactions are conserved, one can represent chemical reactions by equations in which the numbers of all constituents are the same on both sides of the equation. That is, the equation must balance. One can literally do theoretical chemistry by manipulating these symbols in the following example: (Giere & Moffatt 2003, 304)
>
> $CH_4 + 2O_2 = CO_2 + 2H_2O$ (The burning of methane)

Such formulas are clearly external representations that form part of an extended cognitive system that allows scientists to explore possible reactions in organic chemistry. "That is, the cognitive process of balancing an equation does not take place solely in the head of some person, but consists of interactions between a person and physical, external representations" (*ibid.*).

Moreover, in relation to the above, recall the point we made in the previous subsection about how the scientific theories (or models) one already has in mind may affect the way one will further develop one's theories and so on: Klein (1999) explains how Dumas, by modifying and using Berzelius' chemical formulas as 'paper tools', came up with the notion of 'substitution', later to be developed into a new theory about the unitary structure of organic compounds. Specifically, she argues that it was *the physical manipulations* of these formulas and formula equations that led to the conception of the notion of substitution. And eventually, it

[18] For the importance of the normative aspects of the external representational systems in explaining cognition see (Menary 2007).

was through these formulas that the concept of substitution allowed Dumas to see the link between the theory of proportion and the notions of compound and reaction.[19]

So to return to the discussion of the transformative effects of science, the above is a clear example of a scientific theory that does not only alter the agent's inner cognitive processes, but also allows him to externalize his problems in symbols, whose physical manipulation enables him to come up with solutions that, arguably, would otherwise be unavailable. This example, however, is only one out of a surprisingly large number of similar cases. In fact anyone who solved problems in mathematics, physics, chemistry, logic, geometry or even biology (recall the Mendelian inheritance trees) at school can come up with one's own examples.

By actively driving (and constraining) the ongoing cognitive loops, scientific theories—like public language and text—can then be seen as software artifacts that allow the scientist's cognitive character to literally extend beyond her natural cognitive capacities. As Lakatos once wrote, the "methodological falsificationist *uses our most successful theories as extensions of our senses*" (1970, 107, emphasis in the original).[20]

This all should be interesting in its own. In addition, however, and before closing this section, we should also note how such an understanding of scientific theories and artifacts is well-positioned to capture and reveal the inherently social nature of scientific knowledge, even when, apparently, this is solely down to the individual scientist. Previously, we noted that when an agent gains knowledge on the basis of a telescope, her cognitive success is significantly creditable to her cognitive agency on account of having appropriately integrated the artifact within her cognitive character. What about the rest of the credit, however? This is a fair worry, for, in such cases, the prevailing factor in the causal explanation of the agent's cognitive success is the *integrated* extended belief-forming process that consists of both one's cognitive agency and the epistemic artifact, operating in tandem. So, even though *a* (or even if *the* most) significant part of the credit has to be attributed to the agent's internal cognitive capacities (i.e., his cognitive agency), at least *some* credit must also be attributed to the external aspects of the overall process.

Should we then attribute credit to telescopes, microscopes, calculators, languages, scientific theories and so on? It seems that the answer can't be positive. To see why, consider that even though Greco (2004) holds that credit attributions are very much akin to causal explanations, attributions of responsibility, praise, or merely neutral action (i.e., attributions of positive, negative, or merely neutral credit, respectively) have been traditionally associated

[19] An anonymous referee suggests that a similar example with an equally important impact on the development of further scientific theories is the case of Feynman diagrams. Similarly, Kaiser (2005, p. 156) accentuates their role by noting that "Feynman diagrams have revolutionized nearly every aspect of theoretical physics."
[20] According to Lakatos, science and all scientists proceed on the basis of 'sophisticated methodological falsificationism', which is the methodological background against which he presented his rational reconstruction of science.

with *intentional* agents. Accordingly, the remaining credit should be attributed not to the artifacts themselves but to the individuals that *intentionally* brought the relevant extended belief-forming processes about. Notice further, however, that, frequently, we will not be able to attribute the rest of the credit to only one single individual, because, in most cases, a (potentially very large) number of individuals contributes to the development of such reliable belief-forming processes, by means of providing even more belief-forming (sub-) processes or data produced on their basis. Accordingly, many times, the remaining credit, i.e., the credit that is associated with the external portion of the epistemic agent's extended cognitive ability, will be dispersed among a potentially large part of the agent's epistemic community.[21] Overall, then, the view of scientific knowledge we get from the combination of $\text{COGA}_{\text{weak}}$ with the extended cognition hypothesis, is one, whereby to become a scientist, the individual needs a scientific community, able to supply him with the necessary reliable-belief forming processes that he can then integrate within his cognitive character so as to come to know the truth of some scientific proposition *p*.[22]

4. THE DISTRIBUTTED COGNITION HYPOTHESIS AND SCIENTIFIC KNOWLEDGE

So far, we have been focusing on the combination of virtue reliabilism with the extended cognition hypothesis. The advantages of introducing active externalism within virtue epistemology may not stop here, however. For example, we may further make the additional claim that there can 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.

As mentioned before, the hypothesis of distributed cognition has been developed in parallel with the hypothesis of extended cognition (Hutchins 1995; Theiner et al., 2010;

---

[21] One possible worry here is to ask whether the above distribution of credit suggests that we should also posit distributed cognitive systems that would be vastly extended not only in space but also back in time. This would be a rather awkward claim to make, but we can clearly answer in the negative: As noted in §2.1, in order to have an extended or distributed cognitive system we need that all the contributing members reciprocally (i.e., non-linearly) interact with each other. The dependency relations that give rise to the distribution of credit I refer to above, however, are all, one-way, linear ones. Accordingly, given the dynamical systems theory approach to extended and distributed cognition we here draw upon, we have no reason to think that the above distribution of credit requires to also posit distributed cognitive systems that may even extend back in time. Thanks to an anonymous referee for requesting me to clarify this potentially confusing point.

[22] Think about Newton's infamous remark in a letter to his rival, Robert Hooke:

> What Descartes did was a good step. You have added much several ways and especially in taking the colours of thin plates into philosophical consideration. If I have seen further is by standing on the shoulders of Giants.

Even if, as is often assumed, the above is in fact intended as a sarcastic remark (due to Hooke's short build), the ambivalence that Newton takes advantage of derives from the truthfulness of the expression within the context it is uttered—that no matter the genius and the individual efforts of a scientist, his or her achievements will always rest upon and derive from the achievements of other scientists and in general the scientific community he or she is a part of.

Theiner & O'Cnnor, 2010; Sutton et al., 2008; Wilson, 2005; Heylighen et al., 2007) and differs from the latter position only in that, this time, the cognitive system extends to include epistemic artifacts as well as other agents. Moreover, most proponents of the view (Sutton et al. 2008; Theiner et al. 2010; Tollefsen and Dale 2011; Heylighen et al., 2007) point out that, again, it is the existence of *non-linear, reciprocal interactions* between the contributing members and their artifacts that is the criterion by which we can judge whether we have an integrated distributed cognitive system. Accordingly, by the lights of virtue reliabilism there could be knowledge-conducive cognitive characters/systems, which may nevertheless be distributed.

This is an interesting possibility, because it can allow us 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 at least one 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 then use the *individualistic* approach of virtue reliabilism in order to account for knowledge that is collectively produced and which is, thereby, distinctively social.

For example, we can use $\text{COGA}_{\text{weak}}$ to explain how a research team gains knowledge on the basis of an experiment. Even though the knowledge-conducive belief-forming process consists of several experts *and* their experimental devices engaging in reciprocal (socio-epistemic) interactions, the *collective cognitive success* of believing the *truth* of some (scientific) proposition will still be significantly creditable to the group's cognitive agency—i.e., the assembly of the organismic cognitive faculties of its individual members: It is the assembly of these organismic cognitive faculties that is first and foremost responsible for the emergence and efficient sustaining of the collective's belief-forming process. To paraphrase Clark, *cognition is organism centered even when it is distributed*. Crucially, however, given that any cognitive success that is collectively produced in this way will only be creditable to the *collection* of the members' cognitive agencies *as a whole* and to none of the individual members alone, it won't be known by any individual alone, but by the group agent as a whole. In other words, by combining an individualistic condition on knowledge, such as $\text{COGA}_{\text{weak}}$, with the hypothesis of distributed cognition, we can make sense of the claim that p is known by G (the group agent), even though it is not known by any individual alone.

How is this possible and what does it mean, exactly? First, we must make clear what it *doesn't mean*. To claim that a proposition p is known by the epistemic group agent as a whole, in the sense presented here, is not to claim that the relevant proposition is collectively known, because it is collectively *believed* (or 'accepted'). This is an alternative approach to collective knowledge (see for example (Gilbert, 2007a, 2007b, 2007c, 2010, Rolin 2008, List 2011,

Tuomela 2004)) that is not necessary to the present approach, and should be clearly distinguished from it. Of course, group knowledge, just as any other type of knowledge, will always involve belief in the proposition known, and the relevant belief must also, on some appropriate construal, qualify as the belief of the group. Nevertheless, whether the relevant belief (or acceptance) counts as the belief of the group, because it is of the summative or non-summative type, the belief of some operative members of the group or merely the belief of a single representative (for an overview, see Tollefesen 2004) is an issue we do not here need to take a clear stance on—in point of fact, on the present account, any of these possibilities with respect to group belief may give rise to collective knowledge.

To see why we do not here need to delve into the details of group belief in order to make the case for collective knowledge, remember that, as we noted in §3.1, virtue reliabilists accentuate the importance of the *process via which one gets to the truth of the matter*: One's true belief is creditable to one, such that it can thereby constitute knowledge, only if one arrives at the truth in virtue of the belief-forming process (i.e., cognitive ability) one employed to form one's belief. Accordingly, on the basis of virtue reliabilism, which accentuates the importance of the cognitive process via which one arrives at the truth, we can motivate collective knowledge on the basis of cases where arriving at the truth of some matter is the product of a collective belief-forming process. In other words, on the present, *virtue reliabilist* approach to collective knowledge, claiming that a group, as a whole, can have knowledge of a proposition *p* is not because the relevant proposition is collectively and irreducibly believed, but because getting to the truth of the matter as to whether *p* (or *not-p*) could only be *collectively achieved* and is thereby creditable only to the group as a whole.[23]

With all that said, however, we must further disambiguate the following subtle, yet important point: To claim that a proposition *p* constitutes, say, *scientific knowledge* that can only be had by a research team as a whole does not mean that *p* cannot also come to be known by individuals. For instance, claiming that the existence of the Higgs-boson can only be scientifically known by the ATLAS group at CERN does not mean that individual laypeople cannot come to know that the Higgs-boson exists, say by watching the news. The answer to how this is possible lies, again, in the virtue reliabilist's emphasis on the process via which one gets to the truth of the matter. Given that one can truly believe a proposition on the basis of a multitude of different knowledge-conducive processes, *it is possible that two individuals can know*

[23] Similarly to fn. 22, one may worry that the present account overgeneralizes what may count as collective knowledge such that any case of interpersonal interaction may count as a case of collective knowledge. In response, each time, we need to ask the following questions: First, judging by the criterion of continuous reciprocal interactivity, is the relevant process a collective belief forming process? And, second, is the output of the relevant process a true belief? So, for example, asking for directions from a stranger and acquiring a true belief on that basis won't count as collective knowledge, but merely as individual, testimonial knowledge, because the relevant process is merely a one-way, linear interpersonal interaction. In contrast, as I have previously noted elsewhere (Carter et al. 2014) the products of transactive memory processes (Wegner et. al, 1985; Wegner, 1986)—provided that they are true—are typical cases of collective knowledge.

that p all the while their knowledge of p is not the same; for instance one may have perceptual knowledge of *p,* whereas another person's knowledge of *p* may be testimonial. To put the same point in terms of our previous example, it is possible that both the ATLAS group and a reader *of Nature* can know that the Higgs particle exists, and still it may be the case that this piece of knowledge is different in each case. The reader of *Nature* can only come to truly believe the relevant proposition on the basis of a testimonial chain at the beginning of which lies the scientific knowledge of ATLAS. Accordingly, her knowledge can only count as *testimonial knowledge* of a scientifically derived proposition; in order for such a scientific proposition to be known at all, however, such that it can then be testimonially transmitted, getting to its truth for the first time can only be the product of, and thereby creditable to, ATLAS (or a similar scientific research-team) as a whole.[24] In other words, even though scientifically derived propositions such as "the Higgs particle exists" may be individually known on the basis of testimony, they may only be *scientifically* known by a research team as a whole.[25]

Actually, philosophers and anthropologists of science have already attempted to analyze knowledge produced on the basis of scientific experiments along lines very similar to the above. Think for example Cetina's (1999) ethnographic study of high-energy physics experiments in CERN, or Giere's descriptions of the Indiana University Cyclotron facility experiments (2002*b;* 2006) and observations made with the Hubble Telescope (2007).

It should be noted, however, that even though Giere wants to accentuate the distributed nature of the *process* that produces knowledge in such cases as well as the accompanying spread of epistemic responsibility in the form of credit (and why not, we may add, of blame as well?), he still wants to *deny* the existence of a distributed agent that knows. According to Giere, the reason for this contrast with what we claimed above—i.e., that there *can indeed* be cases where the subject of knowledge is an epistemic group agent or mind—is simple: "[C]ognitive agency is tightly bound up with related concepts such as that of intention, responsibility, consciousness, and in general having a mind in what is (sometimes derisively) called the 'folk-psychological' sense of 'mind'" (2006, 715). And since clearly, as Giere further argues, such concepts do not apply to scientific experiments (though cf. Cetina 1999), we should resist the idea of 'distributed knowing' or of a distributed epistemic agent in general. Instead, Giere suggests, we should opt for an "epistemology without a knowing subject" ((Giere 2007), quoting (Popper 1968)), according to which, in cases like the above, it is only correct to claim that "it has been scientifically established that *p*" (Giere 2002*b*) or that 'it is scientifically known that *p*'.

[24] Unless the overall belief-forming process of the research team can somehow be technologically emulated.
[25] Thanks to two anonymous referees for *Synthese*, whose insightful comments helped clarify the points laid out in the above two paragraphs.

Following the account on offer, however, may allow us to resist this impersonal image of scientific knowledge. Indeed, collaboratively produced scientific knowledge does not belong to *any* particular *individual* subject, *S*. Recall, however, that, on the present account, even such knowledge is still knowledge of a subject, *G*—the *epistemic group* agent: A result that is quite crucial from the point of view of mainstream epistemology, whose spirit has traditionally been methodologically individualistic (Goldman 2010, p.3), and which has always assumed that knowledge is knowledge *of* a subject. To see how the present approach can help mainstream epistemology resist the attempt to promote 'epistemology without a knowing subject' it is instructive to examine Giere's appeal to common-sense psychology a bit closer.

To begin with, everyday discourse should actually make it clear that the idea of epistemic group agency does not really run against common sense. Claims such as 'FBI knows that *p*' (Goldman 2004) or that 'the ATLAS research team knows that *p*' are ubiquitously uttered in modern society.[26] Moreover, even though it is true that not every aspect of a scientific experiment, or a group, in general, may have consciousness, intentions, or is responsible, these are not sufficient grounds for denying them the status of a group mind. After all, not every aspect of our minds is responsible, conscious or has intentions. Rather, only *some* parts of our cognitive systems (if any at all) bear these properties, and yet this is sufficient for the entirety of our cognitive systems to qualify as minds. Why then should the situation be any different in the case of groups? Courtesy the relevant aspects of their individual members, such collective systems will surely bear these properties too.

Of course, it may be objected that even so, what is still missing, but is necessary for groups to qualify as minds in themselves, is that groups possess these properties not at the 'sub- (individual) level but at the group level itself. Nevertheless, if we leave collective responsibility to the side for the moment—we will be returning to it shortly—closer inspection demonstrates that the objection cannot run very far either with the property of intentionality or the property of consciousness. With respect to the former—and remember that we do not here need to take a clear stance on the topic—'collective intentionality' has received considerable support by several prominent philosophers (Gilbert, 2007a, 2007b, 2007c, 2010; List, 2011; Pettit, 2002; List & Pettit, 2006; Tollefsen, 2002*a*; Tollefsen, 2002*b*; Tuomela & Miller, 1988; Tuomela, 2004). Accordingly, epistemic group agents are in fact good candidates for qualifying as bearers of (collective) intentional states.

[26] Consider Goldman's (2004) attempt to explain Sandy Berger's (former US national security adviser) mysterious dictum that "the F.B.I. didn't know what it did know" about the 9/11 attack. Goldman claims that Berger refers to two different conceptions of the bureau. Under a summative conception of the agency as an aggregate of individual members, FBI *did* have knowledge of the attack, because certain agents individually possessed sufficient evidence, such that *were* this evidence to be put together, it *would* lead to a successful prediction of the attack. These individual agents, however, failed to appropriately combine their pieces of evidence in the right way meaning that the FBI, taken as a corporate or group entity, failed to know that the attack would take place.

Similarly, turning to the topic of consciousness, objecting to group agency on its basis cannot be substantial enough. Consciousness is perhaps the most widely debated topic within philosophy of mind and cognitive science (see (Van Gulick, 2004) for an overview) and, as the problem of the 'explanatory gap' (Levine, 1983; McGinn, 1991; McGinn, 1989; Chalmers 1996) suggests, most of it has proceeded on the basis of *a priori* conceptual analysis. In other words, consciousness is a topic that is far from understood, severely compromising any attempt to settle further debates on its basis. Not only that, but 'consciousness', as Block (2002, 206) has put it, is actually a "mongrel concept" that can refer to several different types of things (e.g., 'phenomenal consciousness', 'access consciousness', 'self-consciousness', to name but a few), none of which can be guaranteed to be incompatible with the idea of group minds. In fact, not even *phenomenal consciousness*—i.e., the least likely type of consciousness to be possessed by groups—can provide a definitive verdict with respect to whether groups can have a mind of their own. Especially not, when prominent philosophers of mind (e.g., Block, 2002; Chalmers 1996) insist on the conceptual possibility of philosophical zombies: i.e., entities with minds exactly like our own, but who lack phenomenal consciousness altogether. As Tollefsen (2006) summarizes the point:

> It is not clear [...] that one can settle what a mind is by a priori means or by reflecting on our experience as human minds. It is clear that human minds "feel" (at least some of the time) and that the zombies of philosophical thought do not. But it is not clear that minds without phenomenological experience are conceptually impossible. The fact that human minds feel and collective systems (and zombies) do not should not lead us to skepticism about the possibility of collective minds. Rather, we should conclude, as we do in the case of animal minds, that there are different sorts of minds.[27]

Of course, such a conclusion could hardly come as a surprise for many who work within philosophy of mind and cognitive science, where "minds that stop short of having the full range of properties that our minds have are commonplace. Newborn human infants, non-human animals, and certain kinds of [sci-fi] machines are recognized as possessing such mind, manifesting only some of the psychological states or abilities characteristic of the minds of normal adult human beings" (Theiner & Wilson, 2013).

Characteristic of this trend is Wilson's suggestion (2001*b,* p. 267) to make sense of mindedness ascriptions in such cases—as well as in cases where philosophers and cognitive scientists are willing to ascribe agency to groups (e.g., Theiner et al., 2010; Theiner & O'Connor, 2010; Theiner and Wilson, 2013; Barnier et al., 2008; Tollefsen, 2006; Wegner, 1985; Wegner; 1986; Sutton, 2008)—on the basis of what he calls '*minimal mindedness*':

[27] It is important to note, however, that, usually, philosophical zombies are not invoked to demonstrate that there can be minds that lack phenomenal consciousness. Their main role in the literature has been to demonstrate the shortcomings of physicalism: that is, that there can be creatures that are physically and behaviorally identical to us, but which nevertheless lack consciousness. Since, however, the standard description is that zombies "behave just like us, and some even spend a lot of time discussing consciousness" (Kirk, 2011), the assumption seems to be that zombies may well qualify as mindful, even though they are entirely void of consciousness.

> *X* has a minimal mind, just in case *X* engages in at least one psychological process or has at least one psychological ability.[28]

On the basis of this, however, and since groups can qualify as having the psychological ability to engage in group-justificatory processes, it transpires that groups, after all, can also qualify as possessing at least a minimal form of mind.

Therefore, we see that Giere's common-sense psychological attempt to disallow groups from having minds in themselves is, to say the least, inconclusive. This should be welcome for mainstream epistemology, which now seems that it can safely ascribe knowledge to groups. Additionally, though, this may count as a positive result for *anyone* interested in group knowledge in general, for the alternative idea of 'epistemology without a knowing subject' appears to be a rather metaphysically unstable view: Since the existence of any given property is normally tightly associated with the existence of a corresponding system this property belongs to,[29] how can we follow Giere in accepting the existence of a collective, epistemic, cognitive property—such as a collective belief-forming process/collective justification—in the absence of some epistemic, distributed cognitive system—i.e., an epistemic group agent—this property belongs to?[30]

---

[28] One possible worry with the idea of minimal mindedness is that it may run the risk of being too liberal, depending on what may count as a psychological process or ability. At this point, employing a rather common tactic, Wilson responds without offering a definition. Instead, he attempts to 'fix our ideas' on the basis of the following incomplete but suggestive list of what may count as a psychological process or ability: "perception, memory, imagination (classical Faculties); attention, motivation, consciousness, decision-making, problem-solving (processes or abilities that are the focus of much contemporary work in the cognitive sciences); and believing, desiring, intending, trying, willing, fearing, and hoping (common, folk psychological states)" (Wilson 2001*b*, 266). Of course, it might still be objected that this is not an entirely successful approach to clarifying what may count as minimally mindful, as some of the most basic psychological processes and abilities listed above can be plausibly ascribed to systems whose mentality is rather dubious (for example, problem-solving or perceiving can be plausibly, even if metaphorically, ascribed to certain computers). Note, however, that this worry gradually fades away as we focus on more complicated psychological processes and abilities, such as imagining, decision-making, believing and, as in the present case, being justified. Surely, any system that posses such complicated psychological traits should qualify as at least minimally mindful.

[29] Even though this claim may sound as presupposing substance metaphysics, the present paper draws on dynamical systems theory (see section 2.1), which should be a good indication that the present approach is rather sympathetic to the spirit and methodology of process philosophy (for an overview of the debate between substance metaphysics and process philosophy, see Seibt 2012). According to dynamical systems theory, however, properties cannot be conceptualized in the absence of the systems they belong to: *Properties are behavioral regularities that arise out of processes of interactions between the components of a system*. Accordingly, the present approach seems to be orthogonal to the debate between substance metaphysics and process philosophy as it does not ascribe metaphysical primacy to either substances or processes—an adequate description of reality requires both.

[30] Perhaps one way to do so is to follow Wilson (2001a; 2001*b*; 2004; 2005) who has attempted to propose such a deflated approach to social properties on the basis of what he calls the 'social manifestation' thesis: Collective psychological properties, whatever they are, are properties of individuals, no matter they can only be manifested insofar the relevant individuals constitute part of a social group. The problem, however, is that Wilson's insistence on the bearers of such collective properties being exclusively individuals is a form of favoritism towards individualism. Specifically, Wilson derives his social manifestation thesis from what he calls (a) wide and (b) radically wide realization: Respectively, the ideas that the total (in the case of (a)) and core (in the case of (b)) realization of certain properties is at least partly located outside the individual who possesses the relevant property (2001*a*). However, when Wilson is pressed to explain why the bearers of such properties are not themselves wide but are, instead, exclusively individuals, all he offers by way of an explanation are a few remarks to the effect that either the core realizers of the relevant property are realized in large part—*even if not wholly*—by the activity of the individual, or that individuals, in general, must stand out in our explanations, because they "are spatio-temporally bounded, relatively cohesive, unified entities that are continuous across space and time" (Wilson, 2001*a*, p. 24). Since, however, in the case of distributed cognition the relevant collective properties are radically wide, realized by

Of course, in return, this point may generate the objection that even if we accept, on the present view, that there is an overall cognitive system such a collective belief-forming process belongs to, it is not clear whether such a system can qualify as an *epistemic group agent* in itself. Given the preceding discussion with respect to common-sense psychology, however, there is no promising rationale to motivate the denial of *cognitive agency* to such a distributed cognitive system. This leaves the ascription of *epistemic* group agency, in particular, as being the only remaining cause for concern. Again, however, there is no principled reason to deny the relevant collective cognitive system the status of an *epistemic* group agent in itself. As noted in §3.2, according to virtue reliabilism, epistemic agency is a rather weak notion that manifests itself in the actions of initiating, sustaining, and monitoring the relevant belief-forming process. Specifically, according to virtue reliabilism, epistemic agency is manifested in the following weak, (epistemically) externalist sense of *epistemic responsibility*: If there is something wrong with the relevant belief-forming process then the agent will be able to spot this and respond appropriately, otherwise—if there is nothing wrong—the agent can be *by default* responsible (i.e., subjectively justified (Palermos 2014*b*)) in employing the relevant belief-forming process and its resulting beliefs without even being aware that he does so or that the process is reliable. Accordingly, it is not at all obvious why one should deny *epistemic* agency to a collective cognitive system either: After all, it is the assembly of the individual members of the group *as a whole* that initiates and sustains the relevant collective belief-forming process and it is the same assembly operating *as a whole* that is responsible for it: It is the participating members' reciprocal interactions—which bind them together into a unified whole—that allow their cognitive ensemble to effectively be in a position to respond appropriately in cases where there might be something wrong with some part of the overall process.

One possible worry, however, is that the above notion of collective epistemic responsibility (and agency) may be too weak. Specifically, it may seem too liberal, because it may allow individuals who would normally not count as parts of a research team to actually qualify as proper parts of the overall group that is collectively responsible and creditable for the final scientific findings. Imagine, for example, a group that is made up of both lab technicians and scientists conducting a scientific experiment. Before and during the experiment the technicians causally interact continuously and reciprocally with the scientists in ways that are directly relevant to the experiment. They attend to the operation and maintenance of the equipment, teach the scientists how to use it, and continually monitor the operation of the machinery. Accordingly, given dynamical systems theory and virtue

---

the activity of *all the contributing individuals operating in tandem*, and since distributed cognitive systems are spatio-temporally bounded, cohesive, unified wholes in themselves, none of these explanations is satisfactory. In fact, even Wilson himself seems to admit as much: "In at least some cases of wide realization, particularly those of radically wide realization, there is [no] non-arbitrary way to single out individuals as the subjects or 'owners' of the corresponding mental properties. If we have wide realizations of mental states, and thus wide mental states, so too we should have 'wide subjects' of those states" (*ibid.*, p. 24).

reliabilism, the technicians count as proper parts of the distributed cognitive system/epistemic group agent that conducts the experiment.

Nevertheless—as the objection may further go—it may seem counterintuitive to claim that the technicians form proper parts of the same epistemically responsible network that includes the scientists of the team: The technicians do not fully understand or endorse the scientific findings (even though they have played a significant causal role in their production); and even if they are duly mentioned in the final paper, they do not assume any of the epistemic rights, duties and responsibilities that we normally associate with the scientists (e.g., to defend the collectively produced scientific findings if they are appropriately challenged).

In response, one practical reason to resist the above intuitions may be the thought experiment itself. It is rather implausible that scientific groups are organized in the way presented above: Lab technicians usually perform only mundane tasks and work under the supervision of senior scientists. Moreover, it is doubtful that scientists do not know how their equipment works; many times they have to modify or even assemble their equipment themselves and, in any case, it is implausible that they may rely on apparatus that they do not know how to operate or when it is likely to malfunction. Despite these practicalities, however, an important point that the present approach can bring to light is that if there indeed are experiments to which lab technicians contribute in the way specified by the thought experiment above, then perhaps the relevant lab technicians *should not* only be 'duly mentioned' in the final publication. After all, no one individually understands or is able to defend—not even the principal scientific investigators themselves—the scientific findings that have been collectively produced by the research team: The justification for the final result was produced by the group (including the technicians) as a whole and no individual (either a technician or a scientist) could recreate it on his/her own. The point of distributed cognition and epistemic group agency is—precisely—to accentuate the distributed nature of the cognitive achievement and the credit that comes with it: Whoever has *constitutively* contributed to the production of the final result should count as part of the team that is epistemically creditable for it, independently of title, position or any other social status. This may run against current scientific practice, but it is a claim that seems quite plausible in itself, and which perhaps should be taken into serious consideration when shaping future science policies.[31]

[31] Alternatively, if one does not agree with this claim, it is possible to explain why lab technicians fail to form proper parts of the epistemic subject that deserves epistemic credit for the final piece of knowledge—despite qualifying as proper parts of the cognitive systems that gave rise to it—by complementing the above virtue reliabilist approach with the collective intentionality approach to group knowledge (Gilbert, 2007a, 2007b, 2010; List & Pettit, 2006; List, 2011; Rolin, 2008): The lab technicians are not jointly committed to accepting or endorsing the result of the experiment and so they are not epistemically responsible or creditable for it. It should be noted, however, that, at least on certain formulations (e.g., Rolin 2008), the collective intentions approach to collective knowledge gives rise to collective epistemic responsibility in the robust sense of having the capacity to monitor and adjust one's epistemic states in accordance with norms of rationality. In other words, at least certain

In closing this section then, we see that if we remain open-minded with respect to the widely debated topic of what properties are required for a system to qualify as a mind—as well as whether such properties can have collective counterparts—or, simply, follow philosophers of mind and cognitive scientists who are regularly willing to ascribe at least a minimal degree of mindedness to groups (just as they may do with infants, animals, or even certain kinds of futuristic machines), we are in a position to provide a plausible metaphysical and epistemological explanation of what the subjects of group knowledge might be. Otherwise, the cost to be incurred is that mainstream epistemology will remain at a loss about how to account for knowledge that is collectively produced, with the only possible alternative being 'epistemology without a knowing subject', whose proponents are yet to provide it with both an epistemological and metaphysical support.[32] However formidable such an exercise may turn out to be, on the present approach we can simply claim that insofar as there is a collective belief-forming process, then it will belong to a group cognitive agent, who can qualify as an *epistemically responsible* agent in itself; and if believing the truth can only be attributed to the *collection* of the individual members of this epistemic group agent as a whole, then there is group knowledge too.

5. CONCLUSION

In §2, we went through the reasons why virtue reliabilism is particularly apt for an interpretation along the lines suggested by active externalism, and especially by the extended and distributed cognition hypotheses. §3 explored how scientific theories can be viewed as software artifacts that allow the individual scientist to extend her cognitive character beyond her organismic belief-forming processes, and §4 was dedicated to how research teams can be viewed as epistemic group agents that exist and gain knowledge on the basis of non-reducible, distributed cognitive abilities.

The upshot is that, if correct, the above analysis provides an account of knowledge that seems applicable to a disparate variety of cases, despite the fact that knowledge can be attained in a multitude of fundamentally different ways. That is, by suggesting that knowledge-conducive belief-forming processes can take the form of either software or hardware, individual or social, cognitive artifacts of some corresponding epistemic individual or group agent allows for abstracting sufficiently away from irrelevant considerations of

---

formulations of the collective intentionality approach to group knowledge may be epistemically internalist in spirit and as such they may run counter the spirit of virtue reliabilism, which is an epistemically externalist approach to both individual and group knowledge. Thanks to an anonymous referee for suggesting this alternative as well as for proposing the thought experiment and objection discussed in the above three paragraphs.

[32] The collective intentionality approach to collective knowledge (Gilbert, 2007a, 2007b, 2007c, 2010, Rolin 2008, List 2011, Wray 2007; Tuomela 2004) is not here considered as an alternative, since it does not focus on knowledge that is collectively produced, but only on knowledge being collectively *possessed*. But, even if there were a story for such an approach to tell with respect to knowledge that is collectively *produced*, there would still be way more to tell about how such a story would fit within contemporary epistemology.

physical implementation. In this way, we can group vision, reasoning, memory, telescopic observation, scientific theories and even research teams together, providing in effect a unified theory of knowledge that seems able to account for many different (all?) aspects of our individual *and* social epistemic nature.[33]

REFERENCES

Adams, F., & Aizawa, K. (2010). *The Bounds of Cognition* (1 edition.). Malden, MA: Wiley-Blackwell.

Adams, F., & Aizawa, K. (2001). 'The bounds of cognition'. *Philosophical Psychology*, *14*(1), 43–64. doi:10.1080/09515080120033571

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.

Baehr, J. (2006). 'Character, Reliability and Virtue Epistemology'. *The Philosophical Quarterly*, *56*(223), 193–212. doi:10.1111/j.1467-9213.2006.00437.x

Bamberger, J. S., & Brofsky, H. (1979). *The art of listening: developing musical perception*. Harper & Row.

Barnes, B., Bloor, D., & Henry, J. (1996). *Scientific Knowledge: A Sociological Analysis*. A&C Black.

Barnier, A. J., Sutton, J., Harris, C. B., & Wilson, R. A. (2008). 'A conceptual and empirical framework for the social distribution of cognition: The case of memory.' *Cognitive Systems Research*, 9(1–2), 33–51.

Berk, L. & Garvin, R. (1984). 'Development of Private Speech among Low-Income Appalachian Children'. *Developmental Psychology 20*(2): 271-286.

Block, N. (2002). 'Concepts of Consciousness'. In D. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 206–219).

Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart and Winston.

Bloor, D. (1991). *Knowledge and Social Imagery*. University of Chicago Press.

Braddon-Mitchell, D., & Jackson, F. (2006). *Philosophy of Mind and Cognition: An Introduction* (2nd Edition edition.). Malden, MA: Wiley-Blackwell.

Brewer, F. W. & Lambert, B. L. (2001). 'The Theory Ladeness of Observation and the Theory-Ladeness of the Rest of the Scientific Process'. *Philosophy of Science,* Vol. 68, No. 3, Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part I: Contributed Papers, pp. S176-S186.

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

[33] I am thankful to Adam Carter, John Greco and two anonymous referees for *Synthese* for their insightful comments on previous drafts. 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.

Carter, J. A, Kallestrup, J., Palermos, S. O., Pritchard, D. (2014). 'Varieties of Externalism', *Philosophical Issues*, Vol 24 (1): 63-109.

Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

Chemero, A. (2009). *Radical Embodied Cognitive Science.* MIT press.

Chemla, P. K. (2012). *The History of Mathematical Proof in Ancient Traditions*. Cambridge: Cambridge University Press.

Cheon, H. (2014). 'In What Sense Is Scientific Knowledge Collective Knowledge?', *Philosophy of the Social Sciences*, *44*(4), 407–423. doi:10.1177/0048393113486523

Churchland, P. M. (1979). *Scientific Realism and the Plasticity of the Mind.* Cambridge: Cambridge University Press.

—— (1988). 'Perceptual Plasticity and Theoretical Neutrality: A Reply to Jerry Fodor', *Philosophy of Science* 55: 167-187.

—— (1989). *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science.* Cambridge: The MIT Press.

Clark, A. (1998). 'Magic Words, How Language Augments Human Computation'. In *Language and Thought: Interdisciplinary Themes*. (1998). P. Carruthers and J. Boucher (Eds). Cambridge University Press: Cambridge.

—— (2007). 'Curing Cognitive Hiccups: A Defense of the Extended Mind', *The Journal of Philosophy*, 104: 163-192.

—— (2008). *Supersizing The Mind*. Oxford University Press.

Clark, A., & Chalmers, D. (1998). 'The Extended Mind'. *Ananlysis* 58, no. 1: 7-19.

Saussure, F.(1916/1983). *Course in general Linguistics*. Duckworth. (Translated by Roy Harris)

Estany, A. (2001). 'The Thesis of Theory-Laden Observation in the Light of Cognitive Psychology'. *Philosophy of Science,* Vol. 68, No.2 (Jun ., 2001), pp. 203-217.

Feyerabend, P. K. (1975). *Against Method: Outline of an anarchistic theory of knowledge.*

Fodor, J. (1984). 'Observation Reconsidered', Philosophy of Science, 51: 23-43.

—— (1988). 'A Reply to Churchland's 'Perceptual Plasticity and Theoretical Neutrality'', *Philosophy of Science* 55: 188-198.

Fraassen, B. van (1980). *The Scientific Image*. Oxford: Clarendon Press.

Froese, T., Gershenson, C., & Rosenblueth, D., A. (2013). 'The Dynamically Extended Mind', available at: http://arxiv.org/abs/1305.1958.

Fuller, S. (2007). *The knowledge book: Key concepts in philosophy, science and culture.* Durham: Acumen.

—— (2012). 'Social epistemology: A quarter-century itinerary.' *Social Epistemology* 26 (3-4): 267-83.

Gelder, T. van. (1995). 'What Might Cognition Be If Not Computation?', *Journal of Philosophy*, *92*(7), 345–81.

Giere, R. (1988). *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

—— (2002*a*). 'Discussion Note: Distributed Cognition in Epistemic Cultures'. *Philosophy of Science*, 69.

—— (2002*b*). 'Scientific Cognition as Distributed Cognition'. In *Cognitive Bases of Science*, eds. Peter Carruthers, Stephen Stitch and Michael Siegal, Cambridge: Cambridge University Press, 2002.

—— (2006). 'The Role of Agency in Distributed Cognitive Systems'. *Philosophy of Science,* 73, pp. 710-719.

—— (2007). 'Distributed Cognition without Distributed Knowing'. *Social Epistemology*. Vol. 21, No. 3, pp. 313-320.

Giere, R. & Moffat, B. (2003). 'Distributed Cognition: Where the Cognitive and the Social Merge'. *Social Studies of Science*. 33/2, pp. 1-10.

Gilbert, M. (2009). Shared intention and personal intentions, *Philosophical Studies, 144*(1), pp. 167–187. doi:10.1007/s11098-009-9372-z

—— (2007a). 'Collective Epistemology'. *Episteme*. Vol. 1 No. 2, pp. 95—107. doi:10.3366/epi.2004.1.2.95

—— (2007b). 'Modeling Collective Belief'. *Synthese*, Vol. 73, pp. 185-204,

—— (2007c). 'Remarks on Collective Belief'. *Socializing Epistemology: The Social Dimensions of Knowledge 1994* . Available at SSRN: http://ssrn.com/abstract=1052361

—— (2010). 'Belief and Acceptance as Features of Groups'. *Protosociology: An International Journal of Interdisciplinary Research,* Vol. 16, pp. 35-69.

Goldman, A. (2004). 'Group Knowledge Versus Group Rationality: Two Approaches to Social Epistemology'. *Episteme.* Volume 1, Issue 01, June, pp 11-22

—— (2010). 'Why social epistemology is real epistemology'. In *Social Epistemology*, edited by Adrian Haddock, Alan Millar and Duncan Pritchard, 1-28. Oxford: Oxford University Press.

Goody, J. (1977). *The Domestication of the Savage Mind*. Cambridge University Press.

—— (1986). *The logic of writing and the organization of society*. Cambridge, UK: Cambridge University Press.

—— (1987). *The interface between the oral and the written*. Cambridge, UK: Cambridge University Press.

Greco, J. (1999). 'Agent Reliabilism', in *Philosophical Perspectives 13:* Epistemology (1999). James Tomberlin (ed.), Atascadero, CA: Ridgeview Press, pp. 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, pp. 57- 69.

—— (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press.

Greco, J., & Zagzebski, L. (2000). 'Two Kinds of Intellectual Virtue'. *Philosophy and Phenomenological Research*, *60*(1), 179. doi:10.2307/2653438

Hanson, N. R. (1961). *Patterns of Discovery*. Cambridge: Cambridge University Press.

—— (1969). *Perception and Discovery; An Introduction to Scientific Inquiry*. San Francisco: Freeman, Cooper.

Harris, R. 1989. 'How does writing restructure thought?', *Language & Communication*, 9(2–3): 99–106.

Haugeland, J. (1993). 'Mind Embodied and Embedded'. In Y.-H. H. Houng & J. Ho (Eds.), *Mind and Cognition: 1993 International Symposium* (pp. 233–267). Academica Sinica.

—— (2000). *Having Thought: Essays in the Metaphysics of Mind* (New Ed edition.). Cambridge: Harvard University Press.

Hempel, C. (1966). *Philosophy of Natural Science.* Englewood Cliffs, NJ: Prentice Hall.

—— (1970). *Aspects of Scientific Explanation.* New York: The Free Press.

Heylighen, F., Heath, M., Van Overwalle, F. (2007). 'The Emergence of Distributed Cognition: A Conceptual Framework'. In Proceedings of collective intentionality IV (2004), Volume: IV, Publisher: University of Siena.

Hutchins, E. (1995). *Cognition in the Wild.* Cambridge: MIT Press.

Kaiser, D. (2005). 'Physics and Feynman's Diagrams'. *American Scientist*, *93*(2), 156.

Klein, U. (1999). 'Techniques of modeling and paper-tools in classical chemistry. In Morgan, M. & Morrison, M. (eds.), *Models as Mediators: Perspectives on the Natural and Social Science*. Cambridge: Cambridge University Press.

Knorr-Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press.

Kirk, R. (2011). 'Zombies'. *The Stanford Encyclopedia of Philosophy.*

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: The University of Chicago Press.

Lakatos, I. (1970). 'Falsification and the Methodology of Scientific Research Programmes'. In *Criticism and the Growth of Knowledge.* Imre Lakatos Alan Musgrave (eds.). Cambridge University Press, 1970.

Latour, B. (1999). *Pandora's Hope: An Essay on the Reality of Science Studies*. Cambridge, Mass: Harvard University Press.

—— (2007). *Reassembling the Social: An Introduction to Actor-Network-Theory* (New Ed edition.). Oxford; New York: OUP Oxford.

Latour, B., & Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press.

Levine, J. (1983). "Materialism and qualia: the explanatory gap". *Pacific Philosophical Quarterly*, 64: 354–361.

Linell, P. (2005). *The Written Language Bias in Linguistics*. London: Routledge.

List, C. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford ; New York: OUP Oxford.

List, C., & Pettit, P. (2006). 'Group Agency and Supervenience'. *The Southern Journal of Philosophy*, *44*(S1), 85–105. doi:10.1111/j.2041-6962.2006.tb00032.x

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 Mulvihil (eds). Also available at http://www.upscale.utoronto.ca/PVB/Logan/Extended/Extended.html

—— (2006). 'The Extended Mind Model of the Origin of Language and Culture', in *Evolutionary Epistemology and Culture,* N. Gontier et al. (eds), Printed in Netherlands. 149-167.

—— (2008). *The Extended Mind: The Emergence of Language, the Human Mind and Culture.* University of Toronto Press.

Markie, P. (2013). 'Rationalism vs. Empiricism'. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2013.). Retrieved from http://plato.stanford.edu/archives/sum2013/entries/rationalism-empiricism/

McClelland et al. (1986). McClelland, J.L., Rumelhart, D., E., and the PDP Research Group (eds.), *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, vol. 2 .Cambridge, MA: MIT Press.

McGinn, C. (1989). 'Can we solve the mind-body problem?', *Mind*, 98: 349–66

—— (1991). *The Problem of Consciousness*. Oxford: Blackwell.

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 McMillan.

—— (2013). Cognitive integration, enculturated cognition and the socially extended mind. *Cognitive Systems Research*, *25–26*, 26–34. doi:10.1016/j.cogsys.2013.05.002

Morgan, M. & Morrison, M. (1999). *Models as Mediators: Perspectives on the Natural and Social Science*. Cambridge: Cambridge University Press.

Olson, D. R. (1996a). 'Language and Literacy: what writing does to Language and Mind'. *Annual Review of Applied Linguistics*, *16*, 3–13.

—— (1996b). *The World on Paper: The Conceptual and Cognitive Implications of Writing and Reading*. Cambridge University Press.

—— (2002). 'What writing does to the mind'. In *Language, Literacy, and Cognitive Development: The Development and Consequences of Symbolic Communication*. Amsel, E., & Byrnes, J. P. (eds.). Psychology Press.

Palermos, S. O. (*forthcoming*). 'Could Reliability Naturally Imply Safety?', *European Journal of Philosophy. DOI: 10.1111/ejop.12046*

—— (2011). 'Belief-Forming Processes, Extended', *Review of Philosophy and Psychology*, Vol 2 (4): 741-765.

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

—— (2014*b*) 'Knowledge and Cognitive integration', (2014), *Synthese*, Vol 191 (8): 1931-1951.

Palermos, S. O. & Pritchard, D. (2013). 'Extended Knowledge and Social Epistemology'. *Social Epistemology Review and Reply Collective* 2 (8): 105-120.

Pettit, P. (2002). 'Collective Reasons and Powers'. *Legal Theory*, *8*(04), 443–470.

Pritchard , D. (2010*a*). 'Knowledge and Understanding', in A. Haddock, A. Millar & D. H. Pritchard, *The Nature and Value of Knowledge: Three Investigations*, Oxford: Oxford University Press.

—— (2010*b*). 'Cognitive Ability and the Extended Cognition Thesis'. *Synthese.*

Putnam, H. (1975). 'The Meaning of "Meaning"'. In *Language, Mind and Knowledge.* K. Gunderson (ed.). Minneapolis: University of Minnesota Press.

Rolin, K. (2008). 'Science as collective knowledge'. *Cognitive Systems Research*, *9*(1–2), 115–124. doi:10.1016/j.cogsys.2007.07.007

Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes*. New York: Cambridge University Press.

Rupert, R. D. (2004). 'Challenges to the Hypothesis of Extended Cognition'. *Journal of Philosophy*, *101*(8), 389–428.

—— (2009). *Cognitive Systems and the Extended Mind* (First Edition edition.). Oxford ; New York: OUP USA.

Shani, I. (2013). 'Making it mental: in search for the golden mean of the extended cognition controversy'. *Phenomenology and the Cognitive Sciences*, *12*(1), 1–26.

Shankweiler, D., & Liberman, I. (1972). 'Misreading: A search for causes'. In J. Kavanagh & I. Mattingly (Eds.), *Language by ear and by eye; the relationships between speech and reading* (pp. 293–317). Cambridge, MA: MIT Press.

Sosa, E. (1988).'Beyond Skepticism, to the Best of our Knowledge'. *Mind,* New Series, vol. 97, No.386, pp. 153-188

—— (1993). 'Proper Functionalism and Virtue Epistemology'. *Nous,* Vol. 27, No. 1, 51-65.

——— (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge,* Oxford: Clarendon Press.

Suppe, F. (1977). *The Structure of Scientific Theories*. Chicago: University of Illinois Press.

Suppes, O. (1961). 'A Comparison of the Meaning and Use of Models in the Mathematical and Empirical Sciences', pp. 163-77 in H. Freudenthal (ed.), *The Concept and Role of the Model in Mathematics and Natural and Social Sciences.* Dordrecht: Reidel.

——— 'What is a Scientific Theory?' pp. 55-67 in S. Morgenbesser (ed.), *Philosophy of Science Today.* New York: Basic Books.

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,* Issues 1-2, pp. 33–51.

Sutton, J. (2008). 'Between Individual and Collective Memory: Coordination, Interaction, Distribution'. *Social Research*, 75 (1), pp. 23-48.

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*, Vol. 11, Issue 4, pp. 378-395.

Theiner, G. & O'Connor, T. (2010). 'The Emergence of Group Cognition', In A. Corradini & T. O'Connor (eds.), *Emergence in Science and Philosophy*. Routledge. 6—78.

Theiner, G. & Wilson, R. (2013). 'Group Mind'. *Encyclopedia of Philosophy and the Social Sciences*.

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

Tollefsen, D. (2002a). 'Organizations as True Believers'. *Journal of Social Philosophy*, *33*(3), 395–410.

——— (2002b). 'Collective Intentionality and the Social Sciences'. *Philosophy of the Social Sciences*, *32*(1), 25–50. doi:10.1177/004839310203200102

——— (2006). 'From Extended Mind to Collective Mind'. *Cognitive Systems Research*, 7(2-3), pp. 140-150.

——— (2004). 'Collective Intentionality'. *Internet Encyclopedia of Philosophy*

Tomasello, M. (2009). *The Cultural Origins of Human Cognition*. Harvard University Press.

Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2nd edition edition.). Cheshire, Conn: Graphics Press USA.

Tuomela, R. (2004). 'Group Knowledge Analyzed'. *Episteme,* 1 (2), pp. 109-127.

Tuomela, R., & Miller, K. (1988). 'We-intentions'. *Philosophical Studies*, *53*(3), 367–389. doi:10.1007/BF00353512

Van Gulick, R. (2004). 'Consciousness'. *Stanford Encyclopedia of Philosophy.*

Vygotsky, L. (1978). *Mind in Society: Development of Higher Psychological Processes* (New edition edition.). Cambridge: Harvard University Press.

—— (1986). *Thought and Language* (2nd Revised edition edition.). Cambridge, Mass: MIT Press.

Wegner, D., Giuliano, T., Hertel, P. (1985). 'Cognitive interdependence in close relationships'. In W. J. Ickes (Ed.), *Compatible and incompatible relationships* (pp. 253–276). New York: Springer-Verlag.

Wegner, D. (1986). 'Transactive Memory: A Contemporary Analysis of the Group Mind'. In *Theories of Group Behavior*. Eds. B. Mullen and G. R. Goethals. New York: Springer-Verlag.

Weiskopf, D. A. (2007). 'Concept Empiricism and the Vehicles of Thought'. *Journal of Consciousness Studies*, *14*(s 9-10), 156–183.

Wertsch, J. V. (1988). *Vygotsky and the Social Formation of Mind* (Reprint edition.). Cambridge, Mass.: Harvard University Press.

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

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

—— (2005). 'Collective Memory, Group Minds, and the Extended Mind Thesis'. *Cognitive Processing*, Vol. 6, Issue 4, pp. 227-236.

—— (2001*a*). 'Two Views of Realization'. *Philosophical Studies,* 104, pp. 1-31.

—— (2001*b*). Group-Level Cognition. *Philosophy of Science,* 68 (3), pp. 262-273.

Wray, K. B. (2007). 'Who has Scientific Knowledge?', *Social Epistemology*, *21*(3), 337–347. doi:10.1080/02691720701674288

Zagzebski, L. T. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. New York, NY, USA: Cambridge University Press.