THE DISTIBUTION OF EPISTEMIC AGENCY

Orestis Palermos *&* Duncan Pritchard

University *of* Edinburgh

Abstract. In this volume, Goldberg (Chapter 1) defines his socio-epistemological research programme by noting that "the pursuit of social epistemology is the attempt to come to terms with the epistemic significance of other minds" (Chapter 1, section 1, p. 8)—and especially the '*epistemic sensibility*' they exhibit when they operate in common epistemic environments. Goldberg of course has in mind the epistemic sensibility of *individual* epistemic agents, but he does not want to exclude the possibility of epistemically sensible *collective* epistemic agents either. The problem, however, is that Goldberg seems to systematically place at the hard core of his programme the assumption that epistemic agents are exclusively individuals and this forces him to leave the question of *epistemically sensible collective epistemic agents* unaddressed. The aim of the present chapter is to extend Goldberg's programme to also account for this possibility. To do so, we elaborate on the idea of *extended knowledge* (Pritchard 2010*b;* Palermos & Pritchard 2013; Palermos 2011; Palermos 2014*b*), according to which knowledge-conducive cognitive abilities can be occasionally extended to the artifacts we interact with or they may be even distributed between several individuals at the same time. On the basis of this approach we demonstrate that collective epistemic subjects can qualify as epistemic agents on the basis of being able to *collectively* exhibit an appropriate form of *epistemic sensibility*.

In this volume, Sanford Goldberg (Chapter 1) defines his socio-epistemological research programme by noting that "social epistemology is the systematic study of the epistemic significance of other minds." (section 3, p. 11). But what can those minds be and how do they differ from the world around us?

Goldberg elaborates by noting that relying on others is not quite the same as relying on the natural world for evidence—as we do, for instance, when we come to know that it's cold outside by seeing someone reaching for their parka or when we discover that we have a mouse problem by finding the droppings under the sink. The difference, explains Goldberg, is that others manifest "*the very results of their own epistemic sensibility*" (Chapter 1, section 1, p. 3).

Goldberg does not employ any of the existing accounts of knowledge and justification in order to offer a detailed definition of what he means by 'epistemic sensibility'. He does venture, however, for a general picture, according to which individuals are members of a 'knowledge environment' that is "structured by various social practices regarding the acquisition, processing, transmission, and assessment of information" (Chapter 1, section1,

p.3). These social practices, in turn, give rise to certain epistemic expectations between the members of the relevant community, and these expectations can be most profitably thought of as the *norms* that govern the epistemic transactions of that community. Accordingly, it is one's awareness of being answerable to such epistemic norms and expectations that informs one's *'epistemic sensibility'* and promotes one from a mere *epistemic subject* that is capable to possess knowledge to an *epistemic agent* that can participate in the socio-epistemic practices of the relevant community:

> The norms articulate what we properly normatively expect of the relevant individuals as *epistemic agents*: we expect that these individuals have acquired the evidence properly expected of them, and in general that they have behaved with the sort of epistemic responsibility properly expected of them (Chapter 1, section 2, p. 9-10).

Such norms, Goldberg further notes, may be established explicitly, as a matter of agreement or they may be part of "practices (e.g., information sharing) that emerge over the course of repeated interaction between the parties after the parties mutually (if perhaps only implicitly) acknowledge their mutual reliance on certain aspects of the practice" (Chapter 1, section 1, p. 5).

To illustrate his point, Goldberg gives the following example: Imagine you ask your doctor what's the best treatment for your condition and your doctor replies X. Independently of whether, in order to accept X, you need to merely lack any doubts against the doctor's statement or whether you also need to possess positive reasons in its support, Goldberg notes that part of the story of why you accept X is that, in our society, doctors are "expected to be knowledgeable in certain ways, to be apprised of the best practices, to be responsive to any relevant developments in their specialties, and so forth" (Chapter 1, section1, p. 4). And it "seems patent that hearers do expect speakers to recognize that when an assertion is made the speaker renders herself answerable to the relevant expectation itself" (Chapter 1, section 2, p. 9).

Accordingly, Goldberg summarizes the hard core of his research programme by noting that "the pursuit of social epistemology is the attempt to come to terms with the epistemic significance of other minds" (Chapter 1, section 1, p. 8). The reason for this is that others "bring their own epistemic sensibility to bear in all sorts of ways as we shape and operate within common epistemic environments" (Chapter 1, section 1, p.8), and this, Goldberg further explains, is particularly important, because it enables our community's "division of labor to be as far-reaching and systematic as it is" (Chapter 1, section 2, p. 11).

At this juncture, however, Goldberg notes that "it is an open question whether the solitary epistemic subject is the only unit of analysis at which to conduct epistemic assessment" (Chapter 1, section 1, p. 7). If anything, "the development and evaluation of the case for and

against [*collective epistemic agents*] ought to be on the agenda of social epistemology" (Chapter 1, section 1, p. 8). The problem for Goldberg, however, is that despite his long-standing interest in social epistemology, he seems to systematically place at the hard core of his programme (or at least in the innermost layers of its protective belt) a strongly individualistic assumption. Specifically, it is highly unlikely that Goldberg would be willing to reject Alvin Goldman's (2010, 3) tenet that the "epistemic agents of traditional epistemology are exclusively individuals." Accordingly, Goldberg is unable to address and indeed leaves the question of collective epistemic agency entirely open.

The aim of the present chapter is to extend Goldberg's programme by taking the discussion up from where Goldberg leaves it. To do so, we will elaborate on the idea of *extended knowledge* as we have introduced it in previous work (Pritchard 2010*b;* Palermos & Pritchard 2013; Palermos 2011; Palermos 2014*b*). Extended knowledge is the result of combining virtue reliabilism from mainstream epistemology with active externalism from contemporary philosophy of mind. What is distinctive about this proposal is that it brings into question the assumption that minds in general, and knowledge-conducive abilities in particular, are necessarily restricted within the heads of individuals. Instead, knowledge-conducive cognitive abilities can be occasionally extended to the artifacts we interact with or they may be even distributed between several individuals at the same time. If that's correct, then it can have interesting implications with respect to Goldberg's research programme—especially if it can be shown that collective epistemic subjects can qualify as epistemic agents on the basis of being able to *collectively* exhibit an appropriate form of *epistemic sensibility*.

1. Extended Knowledge

The extended knowledge research programme results from the combination of active externalism from philosophy of mind with virtue reliabilism from contemporary epistemology. Accordingly, before proceeding with our argument for the distribution of epistemic agency, it will be helpful to say a few things about both active externalism and virtue reliabilism, and then explain how the two approaches can be combined. In the process, we will also bring to the fore the way virtue reliabilism understands the notion of 'epistemic sensibility,' which is the cornerstone of Goldberg's framework, and which is going to be central in making the case for epistemic agents that are distributed in nature.

1.1 Active Externalism

As a general approach to the nature of mind, active externalism is standardly contrasted with Hilary Putnam (1975) and Tyler Burge's (1986) meaning/semantic externalism. Whereas the latter is a passive form of externalism, the former is rather *active*, in that it concentrates on the aspects of the environment that *drive* one's cognitive loops in an ongoing way. 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), location externalism (Wilson 2004), the hypothesis of extended cognition (Clark and Chalmers 1998), the hypothesis of distributed cognition (Hutchins 1995, Sutton, 2008; Sutton et. al, 2008; Theiner et al., 2010), and so on. For the present purposes, however, we will here concentrate on the latter two formulations.

Focusing on cognitive processing—i.e., any processing that is constitutively involved in the completion of a cognitive task—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). Similarly, though perhaps more interestingly so, the hypothesis of distributed cognition (Barnier et al., 2008; Hutchins, 1995; Sutton, 2008; Theiner et al. 2010; Tollefsen & Dale, 2011; Tollefsen, 2006) holds that cognitive processing may not just be extended beyond the agent's head or organism to include artifacts but can be even distributed across 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 cognitive processes and the resultant cognitive systems 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).[1] It has been recently argued (Chemero 2009, Palermos 2014*a*; 2014*b*), 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* that 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), which is perhaps the most powerful, if not the only, mathematical framework for studying the behavior of dynamical systems, in general.[2]

According to this conceptual framework, what is required 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) is the existence of mutual (i.e., non-linear) relations—that arise out of *feedback loops*—between the contributing parts (Chemero 2009, Froese et al. 2013, Sutton et al. 2008, Theiner et al. 2010, Wegner et al. 1985, Tollefsen & Dale 2011, Palermos 2014*a*).

The underlying rationale is that the aforementioned non-linear relations give rise to an overall *integrated* system that consists of all the contributing subcomponents. Typical examples

of such systems include two mutually interconnected pendulums, and the watt governor coupled with a rotation engine. In the case of the watt governor and the rotation engine, for instance, their ongoing mutual interaction gives rise to an overall, coupled system that exhibits the distributed property of maintaining a near-constant speed, irrespective of the load or fuel-supply conditions. This is a qualitative property that does not belong to any of the underlying subcomponents but to their ongoing mutual interaction and thus to the overall system as a whole. Accordingly, in order to account for this long-term qualitative behavior, we need to postulate an overall coupled system that consists of both the watt governor and the rotation engine.

In some more detail, in cases like this, there are two main reasons for postulating the overall coupled system: (1) As mentioned just above, the aforementioned mutual interactions give rise to new systemic properties (like the governor-engine capacity to maintain near-constant speed) that belong only to the overall system and to none of the contributing subsystems alone—therefore, in order to account for such emerging properties, one *has to* postulate the overall extended or distributed system; (2) Said interactions also make it impossible to decompose the two systems' long-term behavior in terms of distinct inputs and outputs from the one subsystem to the other—therefore, in order to make sense of the long-term behavior of the components, one *cannot but* postulate the overall system.[3] For example, in the case of the governor-engine system again, one cannot account for all of the governor's long-term behaviors without simultaneously taking into consideration the behavior of the engine, and *vice versa*. Specifically, it is impossible to decompose all of the governor's behavior in terms of reacting to distinct inputs from, and generating distinct outputs to, the engine, because—when the two of them are mutually interconnected—some of the governor's ongoing behavior both determines and is simultaneously determined by the behavior of the engine (and *vice versa*).

All in all, then, the claim, on the basis of dynamical systems theory, is that in order to have an extended or even distributed cognitive system—as opposed to merely an embedded one (cf. Adams & Aizawa, 2001, 2010; Rupert, 2004, 2009)—all that is required is that the contributing members (i.e., the relevant cognitive agents and their artifacts) *interact continuously and reciprocally* (on the basis of feedback loops) with each other.[4]

1.2 Virtue Reliabilism

Active externalism holds that cognitive systems are not always bound within the heads of individuals; cognitive systems, instead, may occasionally extend to the artifacts with which individuals interact, or even be distributed amongst several individuals engaging in collaboration. Therefore, in order to introduce active externalism to contemporary epistemology we need an account of knowledge that places at 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. And in fact, 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*).[5]

According to virtue reliabilism, knowledge is creditable true belief, where this means that the subject's cognitive success (i.e., getting to the truth of the matter) is attributable to her manifestation of relevant 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 involving highly specialized training or even advanced technology." (Greco 1999, 287) Here is a relatively weak formulation of virtue reliabilism we can work with, known as 'COGA$_{\text{WEAK}}$':[6]

> *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)

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 the knowledge-undermining epistemic luck involved in Gettier cases. As such cases demonstrate, 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 that 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. 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 the manifestation of his cognitive ability*, such that the success be creditable to him.

In other words, virtue reliabilism accentuates 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 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's belief must be true because of the manifestation of one's cognitive ability. This is why—and we should mark this to better appreciate the account of collective epistemic agency to follow—virtue reliabilism puts 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).[7]

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, John 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 (2010) has noted in more recent work that in order for a process to be appropriately integrated into 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)

Why does Greco prefer to spell out 'cognitive integration' and 'cognitive character' in this way? The answer has to do with a minimal notion of epistemic responsibility and subjective justification—or, in Goldberg's words, with a weak notion of *'epistemic sensibility'*.[8] Specifically, Greco is after a notion of subjective justification (/epistemic responsibility/epistemic sensibility), which is in accordance with epistemic externalism in that it denies that in order to be subjectively justified/responsible/sensible one needs to have access to the reasons for which one's beliefs are reliable.

Unluckily, going into the details of how the integrated nature of one's cognitive character can allow one to be justified in the 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, according to virtue reliabilism, provided that one's belief-forming process is integrated into one's cognitive character such that one would be in a position to be 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 (Palermos 2014*b*).

But is this a sense of epistemic sensibility that Goldberg would be happy with? Judging from the following comment, indeed, it seems that he would:

> To be sure, we can make efforts to become aware of the various norms that structure our epistemic environment; and we can bring ourselves to reflect self-consciously on the evidence we have for thinking that things in general (or this person or that device in particular) reliably conform(s) to the norms. I surmise that most mature humans do have a good deal of relevant evidence, and that we do on occasion self-consciously reflect in precisely this way. But I submit that we typically do so only when we suspect that the situation doesn't seem right: when the person speaking to us doesn't appear to be fully confident, or is evasive; when one's watch has been making strange sounds recently; when the thermometer reads 20 degrees F, yet we know that it is in the middle of a Chicago summer; and so forth. (Chapter 1, section 2, p. 11)

1.3 Extended Knowledge

Previously we have argued that reading virtue reliabilism along the lines suggested by the extended cognition hypothesis is not only an available option (Pritchard 2010*b;* Palermos & Pritchard 2013), but actually necessary for accounting for many instances of knowledge acquired via interacting with epistemic artifacts (e.g., telescopes, microscopes, pen and paper when trying to solve complex mathematical problems and so on) (Palermos 2011; Palermos 2014*b*). Moreover, given that the present goal is to demonstrate how the combination of virtue reliabilism and active externalism can allow for epistemically sensible, collective epistemic agents—i.e., groups of individuals that give rise to a collective, knowledge-conducive cognitive ability—we will here limit ourselves to only a few remarks about the strong compatibility between the two views.

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. On 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" (Greco 1999, 287) seems to be compatible with, or even prefigure, the hypothesis of extended cognition.

This is a good indication for being optimistic about the compatibility of the two views. 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. It appears therefore that 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.

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. 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. Specifically, 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. Moreover, the individual subject satisfies $COGA_{weak}$, 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 $COGA_{weak}$, 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. As Andy Clark has expressed the point, "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)

2. The Distribution of Epistemic Agency

The combination of virtue reliabilism with active externalism, and specifically the hypothesis of extended cognition, can therefore open up the possibility of extended epistemic agents, whose extended cognitive characters can give rise to what we may call 'extended knowledge'—i.e., true beliefs that have been arrived at on the basis of epistemically sensible, extended cognitive abilities. But is it possible to also have 'group knowledge' on the basis of *distributed epistemic agents*? In other words, can collectives exhibit at least the weak form of epistemic sensibility that virtue reliabilism suggests is possible on the basis of the phenomenon of cognitive integration?

To outline our answer, the starting point is to again combine virtue reliabilism with active externalism—though this time in the form of the hypothesis of *distributed cognition*. Doing so can allow for the existence of distributed cognitive characters consisting of a collective cognitive ability that emerges out of the socio-epistemic interactions of the members of a group. Such a collective cognitive ability is going to be irreducible to the sum of the cognitive abilities possessed by the individual members of the group and accounting for it will require the postulation of an overall *epistemic group agent* consisting of all the participating members at the same time.

Of course, before even considering the details of such an answer, one may worry that accounting for collective epistemic agents lacks the requisite motivation—do we really need to postulate such collective epistemic entities? On closer inspection, however, this skeptical attitude is not quite right. To the contrary, postulating epistemic group agents, on the basis of distributed cognition, can in fact allow mainstream epistemology to account for knowledge that is collectively produced and which is thereby distinctively social. Put another way, the idea of epistemic group agents that can act as epistemic subjects in themselves can allow virtue reliabilism to account for the most provocative claim that any social epistemologist could ever make: Namely, that there can be knowledge, which is not possessed by any individual alone but by a group of individuals *as a whole*.

This may indeed sound as a far-fetched possibility but, in fact, there have already been several attempts to introduce this sort of collective knowledge within the literature. Think for example of true beliefs produced by scientific research teams. As several philosophers and ethnographers of science suggest, employing the framework of distributed cognition is the most promising way to analyze such collaboratively produced scientific knowledge (Giere 2002*a*; 2002*b*; 2006*;* 2007; Knorr-Cetina 1999; Nersessian et al. 2003*a*; 2003*b*; Neresessian 2005; 2006; Thagard 1993; 1994; 1997; Palermos forthcoming*b*).

And this is not the only example. The most thoroughly studied case of group knowledge is the case of what has come to be known within cognitive psychology as *transactive memory*

systems (TMSs) (Wegner et al. 1985; Wegner 1986)—i.e., groups of two or more individuals that collaboratively encode, store and retrieve information. Consider the following example:

> Suppose we are spending an evening with Rudy and Lulu, a couple married for several years. Lulu is in another room for the moment, and we happen to ask Rudy where they got that wonderful staffed Canadian goose on the mantle. He says "we were in British Columbia…," and then bellows, "Lulu! What was the name of that place where we got the goose?" Lulu returns to the room to say that it was near Kelowna or Penticton—somewhere along lake Okanogan. Rudy says, "Yes, in that area with all the fruit stands." Lulu finally makes the identification: Peachland. (Wegner et al. 1985, p. 257)

As Wegner et al. explain, during the discussion between Rudy and Lulu the various ideas they exchange lead them through and elicit their individual memories. "In a process of interactive cueing, they move sequentially toward the retrieval of a memory trace, the existence of which is known to both of them. And it is possible that without each other, neither Rudy nor Lulu could have produced the item." (1985, p. 257).

Moreover, and in line with the previous discussion, Sutton et al. (2007) suggest that such systems always involve skillful interactive simultaneous coordination between their members. Accordingly, such systems are good candidates for epistemic group agents, because they clearly satisfy the criterion of cognitive integration, as suggested above. As Wegner et al. claim, the members' interaction "gives rise to a knowledge-acquiring, knowledge holding and knowledge-using system that is greater than the sum of its individual member systems." (1985, p. 256)

So, to give a detailed example of how the suggested framework can account for such group knowledge, consider how $COGA_{weak}$ can account for they way a research team produces 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. If anything, it is the assembly of these organismic cognitive faculties that is first and foremost responsible for the emergence and efficient sustaining of the resulting 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.

Overall, then, by combining an individualistic condition on knowledge, such as $COGA_{weak}$, 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 exactly does it mean? 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 (1987, 1994, 2002, 2004, 2009) and 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 so the relevant belief must also, on some appropriate construal, qualify as the belief of the group. But exactly how that idea is to be understood—and there are several live proposals in the literature on this score (see Tolefsen 2004 for an overview)—is not something that we need to take a stand on here. Indeed, on the present account *any* of these possibilities with respect to group belief may give rise to collective knowledge.

The reason why we do not here need to delve into the details of group belief in order to make the case for collective knowledge has to do with virtue reliabilism's stress on 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, and to be as clear as possible, on the present *virtue reliabilist* approach to collective knowledge, in claiming that a group can have knowledge that *p* one is not thereby maintaining that *p* is collectively and irreducibly believed by that group. Instead, one is rather primarily claiming that the group's getting to the truth of the matter as to whether *p* could only be *collectively achieved* and is thereby creditable only to the group as a whole.

We therefore see that it is possible to claim that there can be collective knowledge, in the sense that believing the truth can be the product of the cognitive ability of a collective epistemic agent. In order for this claim to also be consistent with Goldberg's research programme (assuming that Goldberg would be willing to allow for epistemic agents that are not strictly individuals), however, we must further clarify the sense in which a collective's cognitive integration allows it to qualify as an *epistemic agent* in its own right. In other words, we need to ask: How can collectives be in a position to exhibit their own epistemic sensibility—at least in the way virtue reliabilism understands the term?

The answer to this final question is that according to virtue reliabilism, epistemic agency is a rather undemanding notion that manifests itself in the initiation, sustenance, and—most importantly—*background* monitoring of the relevant belief-forming process.

Specifically, epistemic agency is manifested in the weak and epistemically externalist sense of *epistemic sensibility* that is exhibited by the following conditional: 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) 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.[11] Accordingly, it is not at all obvious why one should deny *epistemic agency* to a collective cognitive system. 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.

3. The Distribution of Epistemic Agency and Social Epistemology

In conclusion, then, we have seen that according to Goldberg, "social epistemology is the systematic study of the epistemic significance of other minds" (Chapter 1, section 3, p. 11), where the significance of other minds and the way they differ from the rest of the natural world is that they can exhibit "*the very results of their own epistemic sensibility.*" (3) We have also seen that the combination of virtue reliabilism with the hypothesis of distributed cognition provides a way to expand Goldberg's research programme by incorporating the possibility of collective epistemic agents that can generate group knowledge. Such knowledge is distinctively social because it is not produced by any individual alone but by the members of the relevant group as a whole, and it is only that group as a whole that can count as epistemically sensible with respect to the relevant proposition. Of course, whether this may count as an extension of Goldberg's research programme or as a different research programme altogether is contingent on certain methodological decisions. Specifically, it is an open question whether, alongside his definition of social epistemology, Goldberg would also like to include in the hard core of his research programme the assumptions that (1) epistemic agents do not necessarily have to be individual agents and (2) that virtue reliabilism constitutes at least a necessary condition on knowledge. If (1) and (2) are accepted then the present view can indeed align with Goldberg's research programme in the pursuit of social epistemology. Otherwise, it will constitute a rival to Goldberg's proposal, and the choice between the two (as well as any other existing alternatives) can only be decided on the basis of comparing their future theoretical progress.[12,13]

References

Abraham, F. S., Abraham, R. H., & Shaw, C. (1990). *A Visual Introduction to Dynamical Systems Theory for Psychology*. Santa Cruz, CA: Aerial Pr.

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.

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

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

Carter, J. A. and Palermos, S. O. (forthcoming). 'Epistemic Internalism, Content Externalism and the Subjective/Objective Justification Distinction.' *American Philosophical Quarterly.*

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

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

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

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

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

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.

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

—— (2004). 'Collective Epistemology'. *Episteme*. Vol. 1 No. 2, pp. 95—107.

—— (1987). 'Modeling Collective Belief'. *Synthese*, Vol. 73, pp. 185-204,

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

—— (2002). 'Belief and Acceptance as Features of Groups'. *Protosociology: An International Journal of Interdisciplinary Research,* Vol. 16, pp. 35-69.

Goldman, A. (2010). 'Why Social Epistemology Is Real Epistemology', *Social Epistemology*, (eds.) A. Haddock, A. Millar & D. H. Pritchard, 1-28, Oxford: Oxford 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.

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

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

Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbound.* Palgrave McMillan.

Nersessian, N. J., Newstetter, W. C., Kurz-Milcke, E. & Davies, J. (2003). A Mixed-method Approach to Studying Distributed Cognition in Evolving Environments.*Proceeedings of the International Conference on Learning Sciences.* pp. 307 - 314.

Nersessian, N. J., Kurz-Milcke, E., Newstetter, W. C., & Davies, J. (2003).Research laboratories as evolving distributed cognitive systems. *Proceedings of The 25th Annual Conference of the Cognitive Science Society.* pp.857-862.

Nersessian, N. J. (2005). Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions. In *Scientific and Technological Thinking*, M. Gorman, R. Tweney, D. Gooding, & A. Kincannon, eds. (Erlbaum). pp. 17-56.

—— (2006). The Cognitive-Cultural Systems of the Research Laboratory. Organization Studies, 27(1), pp. 125-145

Palermos, S. O. (*forthcominga*). 'Active Externalism, Virtue Reliabilism and Scientific Knowledge'. *Synthese.*

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

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

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

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

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

Pritchard , D. H. (2010*a*). 'Knowledge and Understanding', in D. H. Pritchard, A. Millar & A. Haddock, *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.

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

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.

Thagard, P. (1993). Societies of minds: Science as distributed computing. *Studies in History and Philosophy of Science Part A*, *24*(1), 49–67. doi:10.1016/0039-3681(93)90024-E

—— (1994). Mind, Society, and the Growth of Knowledge. *Philosophy of Science*, *61*(4), 629–645.

—— (1997). Collaborative Knowledge. *Noûs*, *31*(2), 242–261. doi:10.1111/0029-4624.00044

Theiner, G. & Allen, C. & Goldstone, R. (2010). 'Recognizing Group Cognition'. *Cognitive Systems Research*, Vol. 11, Issue 4, pp. 378-395.

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

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

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

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.

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

Notes

[1] Briefly, common-sense functionalism holds that mental states and processes are just those entities, with just those properties, postulated by our everyday, common-sense, folk psychology.

[2] For a general introduction to dynamical systems theory see (Abraham, Abraham, & Shaw, 1990). Briefly, *Dynamical Systems Theory* (DST) is the branch of theoretical mathematics that is concerned with the properties of abstract dynamical systems. The general strategy of DST is to conceptualize systems geometrically, in terms of positions, distances, regions and trajectories within the space of a system's possible states. Overall, it deals with the long-term qualitative behavior of abstract dynamical systems that can represent—and thereby act as models of—concrete dynamical systems such as engines, the atmosphere, our planetary system, our brains and so on.

[3] To preempt a possible worry here, note that 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 a detailed explanation of why the existence of non-linear relations that arise out of reciprocal interactions between agents and their artifacts ensures the existence of extended cognitive systems see Palermos (2014*a*).

[4] For more details on how dynamical systems theory can help 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*).

[5] 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.

[6] COGA$_{weak}$ stands for Weak COGnitive Agency. 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. Second, because, in order to also accommodate testimonial knowledge, it requires that one's cognitive success only 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 that 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 on all of these points, see Pritchard (2010*b*; cf. Pritchard 2010*a*).

[7] There are problems facing the virtue reliabilist treatment of epistemic luck, but we do not need to get into them here. See Pritchard (2010*a*) for further discussion of these issues.

[8] For an account of the difference between subjective and objective justification, see Carter and Palermos (forthcoming).

[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 following minimal yet epistemically adequate notion of subjective justification/epistemic responsibility and its relation to cognitive integration see Palermos (2014*b*).

[10] Elsewhere—see 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.

[11] For more details on this externalist sense of subjective justification see Palermos (2014*b* and forthcoming*a*).

[12] According to Lakatos' (1970, 69) methodology of scientific research programmes, the only rational criterion for choosing between two competing research programmes is that one of them can anticipate more theoretically novel facts in its growth.

[13] This paper was written as part of the AHRC-funded 'Extended Knowledge' project (AH/J011908/1) which is hosted by the University of Edinburgh's *Eidyn* Philosophical Research Centre, and we are grateful to the AHRC for their support of this research.