A Defence of the Hypothesis of Extended Cognition by Appeal to *Interactor Cognitive Systems*.

8405684

MSc Mind, Language and Embodied Cognition

The University of Edinburgh

2012

Contents

1. Introduction

In their 1998 paper 'The Extended Mind', Andy Clark and David Chalmers put forward the idea that cognition "just ain't in the head" (Clark & Chalmers, 1998, p.7) and can literally extend into environmental surroundings. This idea has become known as the *extended mind thesis*, or the *hypothesis of extended cognition* (henceforth, HEC), and it allows for the possibility that cognitive systems may include non-neural features of the environment. It can be seen as a radical formulation of the more established notion of the mind being *embodied*,[1] i.e. the way in which the mind is intricately inter-locked with the body and world in which it is situated and leans heavily on body/world contributions.

Since its conception, HEC has received a vast amount of attention,[2] proving to be controversial at both commonsense and academic levels. From a commonsense perspective, the 'mind'[3] and the brain are generally treated as one and the same, and so the notion of the mind literally extending outside of the skull is quite alarming (not to mention a messy prospect!). From an academic perspective, the dominant approach in philosophy, psychology, and artificial intelligence research has been a *cognitivist* (or, *internalist*) one. In other words, researchers have focused on constructing a systematic science of the mind by appeal to notions of representation, causation and law within the internal structure of the brain (Descombes, 2001). For present purposes we need not delve into a more detailed description of cognitivism than this; the key point is that cognitivist research concerns itself with the *internal* workings of the brain. Thus, in claiming that certain cognitive states[4] may be "constituted partly by features of the environment" (Clark & Chalmers, 1998, p. 12), HEC is seemingly proposing a radical change

[1] The term *embodied* has only become widespread in literature in recent decades, and the idea of the mind as embodied is by no means the prevalent view of the mind. However, unlike HEC, it can be said to have a considerable backlog of historical support. See Clark (1997), Heidegger (1962), Hendriks-Jansen (1996), Husserl (1931), Johnson (1990), Merleau-Ponty (1962), Varela, Thompson & Rosch (1991), Yuasa (1987)

[2] See Adams & Aizawa (2001, 2008), Block (2005), Clark (2001, 2004, 2007, 2008), Dennett (1996), Grush (2003), Hurley (1998, 2003), Menary (2010), Rowlands (1999), Rupert (2004, 2008, 2009, 2010), Segal (1997), Sprevak (2009, 2010), Sterelny (2004), Wheeler (2005), Wilson (2002, 2004).

[3] I consider 'mind' interchangeable with 'cognitive system'. I take for granted that any talk of cognitive systems/states/processes/mechanisms automatically insinuates the existence of a mind (Rupert, 2009). However, neither 'mind' nor 'cognitive system' should be considered as a synonym for 'brain'. It will become clear later in the paper that I consider 'mind' to be an abstract definition of collections of cognitive processes across a range of situations, not limited to any one cognitive structure or unique to autonomous humans.

[4] 'Cognitive state' is used in a broad sense, i.e. the specific psychological conditions behind reasoning or thinking (e.g. beliefs, thoughts, desires). Talk of cognitive states automatically entails 'cognitive mechanisms' and 'cognitive processes', i.e. the psychological machinations behind the formation and maintenance of cognitive states.

of research direction.

Consequently HEC seems to have faced an uphill struggle since its nascence. However, it has proved difficult to discard. Many people have found its central claim compelling and continue to defend it, none more so than Clark (2001, 2004, 2007, 2008a, 2008b, 2010a, 2010b, 2010c). The motivation of this paper is to present yet another line of defence for HEC, building on recent work by Deborah Tollefsen (2006). Whereas Clark focuses on cognition extending from an organism to an artefact, Tollefsen proposes that the best way to argue for HEC is by appealing to cognitive systems in which cognitive processes (reciprocally) extend from one human individual to another (what I will call an *interactor cognitive system*). I intend to make a case for this proposition more forcibly than Tollefsen originally does and using novel arguments against some of the more recent HEC criticisms. On the back of these arguments I also intend to make some more general comments about the problem with defining and discussing 'minds'.

The set-up of this paper can be broadly outlined into two interrelated parts:

(1). Part One will focus on consolidating the claim that cognitive states do indeed extend from one individual to another.
(2). Part Two will then use (1) to rebuff some of the major criticisms levelled at HEC.

Going into more detail, the argument will run as follows: In the next section I will outline HEC in its original guise, including the paradigmatic case of Otto and his notebook (Clark & Chalmers, 1998). I will then give support for Tollefsen's (2006) claim that if we accept the case of Otto and his notebook, we can logically entertain the possibility that cognition can (reciprocally) extend from one individual to another (forming an *interactor cognitive system*). I will swiftly reject some initial objections to this approach as well as showing it has support from literature on social cognition and group psychology. Section 4 will focus on three of the major (interrelated) criticisms HEC has faced - the distinctive nature of internal[5] cognition (Adams & Aizawa, 2001, 2008), the problem of cognitive integration (Weiskopf, 2008; Spaulding, 2011), and the need for a 'persisting locus of cognition' (Butler, 1998; Rupert, 2004, 2010) - and how

[5] Throughout this paper 'internal' cognition is taken to mean *intra-cranial* cognition, i.e. neural unfoldings 'inside the head'. Conversely, 'external' cognition is taken to mean *trans-cranial* cognition, i.e. those cognitive states that may include extra-cranial media as constitutive parts of their processing.

these can be avoided by appealing to an *interactor cognitive system*. I will conclude with a short discussion of the problem with trying to define 'minds'.

PART ONE

2. HEC and Otto's Infamous Notebook

2.1. Introducing HEC

HEC is a kind of *externalism*, specifically *active externalism*. Externalists hold that a being having a specific mental state, X, may depend on them being situated in a specific environment and holding a special kind of relationship with this environment (Lau & Deutsch, 2010). Most generally, externalism focuses on semantic content: the meaning of words depends on the precise environment in which a being is located, and not just on "what is in the head" (Davidson, 1987, p. 443). To give a trivial example, my knowledge of it being sunny outside depends on it actually being sunny; rather than some trick of artificial light, or my thinking 'sunny' means 'cold and wet'. An example such as this, in which the *content* of my belief (that it is sunny) can be said to be dependent on the distal features of my environment (i.e. the weather) is an illustration of *content externalism* (Putnam, 1975; Burge, 1979).

The *active externalism* of Clark and Chalmers (1998) takes this idea one step further, suggesting that the environment not only determines the content of our cognitive states, but can play an *active role* in forming our cognitive states. Instead of making a claim about cognitive-state-content, active externalism makes a claim about the 'vehicles'[6] of cognitive states. "Active role" is the essential phrase; the assertion is not merely that the mind is inter-locked with its environment and deeply reliant on it (this is the more conservative claim of *embodied* or *embedded cognition*), but that the environment can contribute to cognition in such a way that it is *literally a part* of the cognitive processing. Environmental resources do not make a merely *causal* contribution to cognition, but a *constitutive*[7] one, meaning that the cognizing organism

[6] The media, or resources, that constitute cognitive states.

[7] This claim about the potentially *constitutive* properties of external resources has received considerable attention. Adams and Aizawa (2001; 2008), in particular, have argued that just because x and y are tightly *causally*-coupled, this does not mean that they are *constitutively*-coupled, i.e. that x is part of y or vice-versa. HEC supporters deny this, claiming that causal 'loops' between individual organisms and external resources are so intricately inter-locked that the external resources come to partly constitute the same system as the organism. I accept this second view. As

and the environmental resource establish a *coupled system* in which all systemic components play an active role (Clark & Chalmers, 1998). Such a coupled system is a cognitive system in its own right, with the cognitive mechanisms being *jointly realised* by both environmental and neural activity, and there being no relevant difference between the environmental and neural parts (Sprevak, 2010).

Active externalism (HEC) is thus defined as the view that:

> Human cognitive processing literally extends into the environment surrounding the organism, and human cognitive states literally comprise – as wholes do their proper parts – elements in that environment.
> (Rupert, 2004, p. 389)

It is important to note that HEC-advocates do not intend to diminish the importance of internal processing. Nor do they intend to suggest that extra-cranial media is *in itself* cognitive apparatus and that our mental content may be running wild in the world. Their aim is to rid us of the notion that cognition is limited to neural machinery and does in fact "promiscuously criss-cross the boundaries of brain, body, and world" (Clark, 2007, p. 164). While the brain performs many cognitive operations, others are delegated to manipulations of external media (Clark & Chalmers, 1998). Using pen and paper to perform long-multiplication (McClelland et al., 1986; Clark, 1989), re-arranging letter tiles to create Scrabble words (Kirsch, 1995), or storing phone numbers and addresses in the iPhone (Chalmers, 2008) are all examples of the environment (partly) performing a procedure that would otherwise have been done in the brain. Rather than reject the fruitful work carried out by cognitivists, HEC (and all externalism) endeavours to expand cognitive scientific research to include cognitive mechanisms 'outside the head' and thereby produce a more fecund discipline.

2.2. Otto and His Notebook

In order to clarify their position, Clark and Chalmers (1998) describe the case of Otto and his notebook:

---

things stand, however, this causal-constitution debate has reached somewhat of a stalemate (Sprevak, 2010). Consequently, I will not enter into a full analysis of the debate here, nor later in the paper, but will assume that external resources can manifest *constitutive* properties.

> Otto suffers from Alzheimer's disease, and like many Alzheimer's patients, he relies on information in the environment to help structure his life. Otto carries a notebook around with him every-where he goes. When he learns new information, he writes it down. When he needs some old information, he looks it up. For Otto, his notebook plays the role usually played by a biological memory.
> (Clark & Chalmers, 1998, p. 12)

By acting as a substitute for his biological memory, Otto's notebook comprises the majority of his beliefs,[8] and these beliefs guide his daily activity, allowing Otto to navigate his physical and social surroundings.

Otto is compared with the case of Inga, who has a normal (biologically-based) memory. Both Otto and Inga wish to go to an exhibition at the Museum of Modern Art: Inga thinks for a moment, recalls that the museum is on 53rd street and goes there; Otto, on the other hand, consults his notebook, reads that the museum is on 53rd street and also goes there. It seems reasonable to say that both Otto and Inga *believe* the museum to be on 53rd street and they both held this belief before consulting their notebook/memory (the belief was simply waiting to be accessed). And this is not a solitary scenario. We can imagine that Otto uses his notebook constantly and as a matter of necessity, allowing it to play the central role in his life that Inga's biological memory plays in hers. What we have, therefore, are two pertinently analogous cases: the information in the notebook *functions* just like Inga's ordinary belief; the notebook-information just happens to lie "beyond the skin" (Clark & Chalmers, 1998, p. 13).

At a cursory glance, it is immediately obvious that Otto's notebook and Inga's biological memory are not *exactly the same*. But this is not what HEC claims. What matters is that there is a kind of *functional equivalence*, i.e. in terms of guiding behaviour, Otto's notebook and Inga's memory are playing the same functional roles. The fact that the notebook is an external resource should not diminish the *active role* it is playing in the cognitive mechanism of Otto's memory. To clarify this notion of functional equivalence, the *parity principle* is introduced, which says:

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

[8] It should be noted that the 'beliefs' in question are *non-occurrent beliefs*, i.e. beliefs that are stored in memory and are not about anything actually *occurring* in time, but are prone to occur under certain circumstances. In the subsequent example, the *belief* of Otto/Inga in the museum's location is not at the 'forefront of their mind', i.e. not *occurrent* (being recalled or looked up in the notebook), until they decide they wish to go to the exhibition.

(Clark & Chalmers, 1998, p. 8)

The parity principle remains a contentious issue. It is often misread as claiming that there is fine-grained similarity between internal and external cognitive processes, but this is not how it is intended to be construed. The idea behind it is that it should act as a "kind of *veil of metabolic ignorance*" (Clark, 2007, p. 167) that would remove any tendency to think that internal processes hold governance on what it takes to count as a cognitive process, or set any kind of benchmark on the details of the causal/constitutive contributions a cognitive resource must make (Wheeler, 2010). It is thus perhaps better regarded as calling for 'fair-treatment' (Sprevak, 2009), rather than parity. Several hypothetical examples have been produced to illustrate this - from intelligent Martians with green slime instead of neurons (ibid) to snake-like accountants who perform calculations by movements atop an advanced touch-screen (Clark, 2008a) - none of which we would discount from 'cognitive' status, purely because their cognitive systems functioned unlike our own. I agree (as most people will) with the principle of asking for fair-treatment amongst cognitive systems, but I disagree with the exact formulation of the original parity principle (the reasons for this are made clear in footnote 19). However, as my defence of HEC does not hinge on the parity principle I will not enter into a full debate here. Functional equivalence and parity are important concepts that will crop up again throughout the paper, but they are not the saving grace of HEC arguments.

Returning to the case of Otto, Clark and Chalmers (1998) provide four criteria that the notebook must meet in order for it to fulfil the functional role of belief in biological memory:

1. "First, the notebook is a constant in Otto's life" (p. 17). It is not a diary that Otto picks up at the end of the day, nor is it split into several volumes. It is a single notebook that is always available to Otto and is "typically invoked" (Clark, 2004, p. 6)
2. "Second, the information in the notebook is directly available without difficulty" (1998, p. 17). This goes hand-in-hand with the idea that Otto always has the notebook with him and he can easily access it.
3. "Third, upon retrieving information from the notebook he automatically endorses it" (p. 17). The information "should be deemed as trustworthy as something retrieved clearly from biological memory" (Clark, 2004, p. 6).
4. "Fourth, the information in the notebook has been consciously endorsed in the past, and indeed is there as a consequence of this endorsement" (1998, p. 17). The point here is

> that it is Otto *himself* who puts information in the notebook, rather than it just appearing or having been endorsed by someone else.

These criteria not only serve to filter out some of the more exotic cases[9] of HEC, but also make more apparent the similarities between Otto's notebook and Inga's memory. Otto and his notebook are creating a *coupled system* in which both internal and external resources are playing an active *constitutive* role, just as all of the internal parts of Inga's biological memory play an active *constitutive* role in her cognitive processing. From a functional perspective, the extra-cranial location of the notebook make it no less a constitutive part (and therefore a 'cognitive' part) of Otto's cognition.

We thus have our paradigmatic case of HEC: Otto's cognitive processing literally extends into his environment (his notebook), and his cognitive states literally comprise - as wholes do their proper parts - the contents of his notebook.

Armed with the example of Otto, HEC supporters have argued for a broadening of the research views undertaken by cognitive scientific disciplines. To continue to confine cognition to the internal would be to greatly deprive 'mind' studies of much that should be of interest. Outside of the directly connected disciplines, a HEC reformulation of the 'mind' would also have knock-on effects for law and ethics.[10] The hot-debate surrounding HEC is thus no trivial matter. Somewhat strangely, however, the idea that cognitive processes may extend from one individual to another has been relatively neglected.

3. Tollefsen and *Interactor Cognitive Systems.*

3.1. Olaf and Inga

One person who has picked up the gauntlet of pursuing the idea that cognitive extension may

[9] Clark and Chalmers (1998) mention the likes of filofaxes and the Internet, which (in the majority of scenarios) can swiftly be dismissed as HEC examples due to their failure to meet all the criteria. However any external resource *can* constitute a cognitive resource under the right circumstances.

[10] The concerns for law and ethics would revolve largely around issues of the 'self', e.g. if the possibility of the mind extending is accepted, what of the self? Can the self extend? And if so, it is presumably more susceptible to interference and invasive approaches.

take place between individuals, as well as between an individual and an environmental artefact, is Deborah Tollefsen (2006). Tollefsen sees her claim as a natural progression from the initial work on HEC: if cognition can extend to external resources, why can't the external resource in question be another organism? Indeed, it is perhaps surprising that cognitive extension from one individual to another was not picked up on sooner, given that humans are fundamentally social creatures and we adapt to one another more effectively than we adapt to our environmental 'tools' (Theiner, Allen & Goldstone, 2010). At the end of their paper Clark and Chalmers (1998) mention the possibility of an unusually interdependent couple in which one partner's beliefs play the same sort of role for the other as Otto's notebook plays for him. In a way, Tollefsen (2006) is thus picking up where Clark and Chalmers (1998) left off.

She begins, quite logically, by noting that the major resistance to the notion of 'collective' cognitive states is motivated by the internalist view that cognitive states are located in minds and minds are inside our heads (Tollefsen, 2006). If we accept HEC, however, then we no longer have reason to fear minds being unable to permeate the organismic skin/skull boundary. In other words, with HEC in tow, the mind is free to extend to whatever and *whoever* it likes. All that is needed is the establishment of an indiscriminate functional cognitive system comprised of both external and internal resources.

A system such as this is presented in the form of Olaf and Inga:

> Consider Olaf who is married to Inga. They have been married for 30 years. Olaf does not suffer from alzheimer disease. He is, however, a philosopher. He often gets lost in his work and has difficulty remembering his appointments, phone numbers, addresses, and so on. Inga, however, has a sharp mind and because they spend a great deal of time together Inga provides Olaf with all of the information that he needs in order to get through his day. Indeed, Inga seems to serve just the same purpose that Otto's notebook serves him. She is his external memory.
> (Tollefsen, 2006, p. 143)

The details of this scenario are not crucial, we can imagine that Olaf *does* have Alzheimer's if we prefer or some other memory deficit. What matters is that in this case Inga is playing the role of Otto's notebook, which plays the role of (Inga's) biological memory in the first example.

For some, this step from a person-artifact cognitive system to a person-person cognitive system will seem entirely legitimate. Others, however, will undoubtedly feel uneasy about the

clear differences between a person and an inanimate notebook. Perhaps wary of this, Tollefsen (2006) consolidates the case of Olaf and Inga by explaining how Inga satisfies the four criteria (Clark & Chalmers, 1998) that an external resource must meet in order for it to fulfil the functional role of a cognitive state (in this case, beliefs):

1. "Inga is readily available to Olaf and Olaf typically invokes Inga on a variety of daily details" (Tollefsen, 2006, p. 143). This is achieved by Olaf frequently asking questions of Inga, and Inga being a constantly available presence in Olaf's life.
2. "Inga is easy for Olaf to access" (2006, p. 143). This again stems from her constant availability and the fact that Inga does not refuse to answer Olaf; she is a willing participant in Olaf's cognitive life.
3. "The information that Inga provides Olaf is more or less automatically endorsed" (2006, p. 143). Olaf does not doubt Inga, he trusts her as he would his biological memory. Conversely, Inga is as reliable as biological memory.
4. "The information that is contained in Inga is information that Olaf previously endorsed at some time or another" (2006, p. 143). Olaf supplies Inga with the necessary information; Inga does not fabricate new (un-endorsed) information for Otto or (unjustifiably) alter previously endorsed information.

It seems therefore that according to Clark and Chalmers' (1998) criteria, Olaf and Inga is as much a case of HEC as Otto and his notebook. As long as we accept Otto's notebook as a genuine case of HEC we have no reason to reject (part of) Inga as a cognitive extension of Olaf's mind. Olaf and Inga form what Tollefsen calls a "collective mind" (2006, p. 148).

3.2. Initial Objections

3.2.1. Is Inga merely a cognitive tool?

So far this all seems well and good. The sceptical reader, however, will no doubt require responses to a few preliminary matters. Firstly, there is the issue of agency. We can consider the notebook to be nothing other than an external store for Otto's beliefs, as though some of his cognitive states have been transliterated to paper; it is merely a 'cognitive tool' for Otto to use.

But the exact same cannot be said of Inga. Inga, considered as an individual subject or agent, is presumably still her *own* person; she can act as an external store for some of Olaf's memory states but she also has her own memory states, not to mention her own desires, intentions, meta-thoughts, subdoxastic states, behavioural dispositions etc. Inga cannot be incorporated into Olaf's cognition in the same way the notebook is incorporated into Otto's. In other words, although some of my cognitive states may extend to other people, "these agents are not my resources alone, and they resist individualisation" (Sterelny, 2010, p. 476).

Where, then, does this instance of HEC leave us? Do the beliefs in question belong to Olaf or Inga? Is Inga merely an extension of Olaf?

These are perfectly legitimate worries that the sceptical reader would be right to raise. It is, however, a concern we can breeze past. HEC clearly entails serious consequences for issues of agency and personal identity, but these issues do not supervene on the isolated topic of cognition or the *functional* job that Inga performs. True, we cannot have cognitive states floating free in the world, they must belong to something; but the exact agential state of this 'something' is not immediately crucial to the matter at hand. On top of this it is not Inga, considered as an individual subject with her own mind, that should be brought into question. It is only *part* of Inga that constitutes *part* of Olaf's cognition, namely the part of her harbouring those cognitive states that have been previously endorsed by Olaf and are readily provided upon his asking for them. Inga remains her own person, she simply contains information that under specific circumstances can be considered an extension of Olaf's cognitive system. In response to the previous paragraph's questions, we can say that the beliefs in question are, in some sense, *both* Olaf's and Inga's, and in another sense they are neither his nor hers.[11] We can also say that Inga, considered as an individual subject, is not merely an extension of Olaf. The subjective boundaries between them are somewhat murky, but she is not merely absorbed into Olaf's personhood; at the very least they remain separate entities at a physical level.

I recognise that this response will be unsatisfactory to many, but as I have said I do not wish to become bogged down in the issue of agency. The agential status of a system does not have to

[11] The beliefs are Olaf's in that they are encoded and endorsed by him, as well as retrieved when he desires; they are Inga's in that they are (encoded and) stored by her. Reciprocally, however, they belong to neither of them individually: standard memory models involve each of the stages of *encoding, storage, retrieval*; as neither Olaf nor Inga completes all the stages, the beliefs cannot be (fully) either of theirs.

map directly onto the cognitive status of a system. As far as the 'mind' is concerned we can accept for the time being that an extended system is being formed between the relevant parts of both Olaf and Inga, who are tightly coupled by the constitutive interaction between them. Part of Inga is indeed acting as (part of) Olaf's biological memory, just as the notebook does for Otto. Any opinion on what this means for the agential state of Olaf or Inga should not detract from the assertion that a cognitive system has arisen between them.

3.2.2. Aren't the beliefs simply being *transferred* from Inga to Olaf?

A connected worry is whether the Olaf-Inga relationship can be viewed as mere *transference* of information (Tollefsen, 2006). Again, this is a legitimate concern. It is avoided, however, by making the Olaf-Inga interaction *reciprocal*. Imagine Olaf wishes to write to a friend but can only remember his surname. Inga wishes to do the same but can only remember the first name. It is via *reciprocal* interaction that they succeed in remembering their friend's name:

> Through a process of joint deliberation they jointly retrieve, via a process of joint reconstruction, the name of the person. The process of retrieval then is active and not found "inside" the heads of Inga and Olaf. Indeed, it is done via the joint activity of discussion and deliberation that occurs between Inga and Olaf.
> (Tollefsen, 2006, p. 145)

If we maintain that the required information (the friend's first name and surname) has been previously endorsed (by Olaf and Inga respectively), and the availability and reliability of Olaf and Inga is the same as before, then we have a case of Olaf's cognition extending to Inga and Inga's cognition extending to Olaf. The cognitive interaction is now one of "reciprocal causation"[12] (Clark, 2008b, p. 24) and cannot be considered a one-way transference of information. We thus have a 'joint' cognitive system (Tollefsen, 2006). What's more, we will discover in the subsequent sections that reciprocally extended cognitive systems can manifest properties which go beyond the aggregation of the cognitive capacities of individual members within the system (Theiner et al., 2010). If the capacities for cognitive processing belong to the overall system, rather than to any individual, then we have stepped away from the possibility of

[12] In keeping with earlier remarks from the paper, the reader may feel more comfortable referring to this interaction as 'reciprocal *constitution'*.

transference from one cognitive (sub-)system to another. What we have instead is an "integrated system with clear problem-solving virtues not reducible to the virtues of any of its individual parts" (Clark, 2007, p. 183).

3.2.3. Why '*Interactor' Cognitive System?*

This brings us to an issue of terminology. If we recognise that the Olaf-Inga cognitive system is a 'joint' or 'collective' system, then Tollefsen (2006) seems justified in referring to it as a 'collective system of minds'. As I have previously mentioned, however, I am shunning this phrase and opting instead for *interactor cognitive system* (henceforth ICS). My reasons are twofold.

Firstly, the use of 'joint' or 'collective' in cognitive literature is not a novelty. Both come attached to a previous backlog of academic research and discussions. 'Collective' has a long history in psychology (e.g. Halbwachs, 1938; Jung, 1922) and 'joint' has become pervasive in work on intentionality and actions (as has 'shared' e.g. Bratman, 1993; Gilbert, 2009; Tomasello & Carpenter, 2007; Velleman, 1997). Perhaps the most obvious word to use, 'group', suffers the fate of having been lobbied around by various psychological and philosophical theories (e.g. Burke, 1989; Janis, 1983; Le Bon, 1895; Pareto, 1935).

As excellent as I'm sure these works are in their own right, they are not necessarily helpful antecedents to our present topic of cognition extending between individuals. They are in some ways connected to HEC and ICS (Velleman's 'shared intentions' for instance could be considered a case of HEC[13]), but the connections are generally too tenuous to be of any use. The current work on cognition extending between individuals should be viewed as a new theoretical enterprise; it may find its evolutionary roots in previous work on 'collective/joint/shared/group' psychology or philosophy, but it has branched away from these enough to warrant a new terminological definition. The re-application of any of these expressions would not entail any immediate benefits and could easily confuse matters.

[13] Velleman (1997) claims we can have shared (or collective) intentions (i.e. a single token of intention jointly held and literally shared) via an exchange of speech acts. It's possible to interpret these speech acts as cases of HEC if we consider them as an *external* verbalisation of an internally represented mental state. This would render Velleman's hypothesis compatible with the current discussion of ICS.

The second reason for shunning Tollefsen's (2006) terminology, along with the possible phrases 'joint/shared/group cognitive systems', is that they all naturally suggest a 'coming together' of individual cognitive systems and this is a potentially dangerous way to consider the kind of cognitive system that Olaf and Inga create.[14] As I have already mentioned it is not Olaf and Inga, as agential individuals, that form a new cognitive system. It is only the *relevant parts* of their minds. This is why the newly formed cognitive system is *both* Olaf and Inga's and (more importantly) *neither of theirs individually*. The new system is created from the cognitive machinery of Olaf and Inga, but it is not a mere aggregation of their cognitive capacities, nor is it reducible to either of them as individual entities. Using the terms 'collective/joint/shared/group' conjures up the image of two or more separate entities connecting together, like two people holding hands or two pieces of machinery being bolted together. These are the wrong kind of images (albeit metaphorical) to have. When cognition is extending between individuals their cognitive states are quite literally fusing and blending together. The newly formed cognitive states can *only* belong to the *overall* cognitive system. To put this more logically, let us consider Olaf to be 'system A' with properties a, b, c and Inga to be 'system B' with properties d, e, f. When they reciprocally extend some of their cognitive states to one another they do not form 'system C' with properties a, b, c, d, e and f, because this is tantamount to saying they form 'system A+B'. Instead they form 'system C' with its own unique properties x, y, z.

In order to capture this idea of a new system with its own cognitive properties and processing capabilities I suggest borrowing the term 'interactor' from evolutionary biology. An interactor is classically defined as "an entity that directly interacts as a cohesive whole with its environment" (Hull, 1980, p. 318). Thus for most intents and purposes the paradigm of an interactor is quite simply an organism. Sober and Wilson (1998), however, define an interactor as an entity (or "adaptive unit") whose parts share a *common fate*, rather than an entity whose parts are organised into a cohesive whole. With this slightly altered definition, an interactor can just as easily be considered as a group of organisms as it can an individual, as long as the organisms within the group share a common fate and can unanimously interact as an 'adaptive unit'. Using Sober and Wilson's (1998) reasoning we can therefore have interactors at both the individual-level and the group-level, because both individuals and groups can be evolutionarily selected

[14] Conversely, use of the term 'distributed' suggests a 'splitting up' or 'dividing' of a cognitive system, which for similar reasons is not the best way to view a cognitive system.

upon by forming adaptive units whose parts share a common fate. What's more, these individual- and group- level interactors are capable of persisting alongside (and within) one another.[15]

Applying this swift sojourn into evolutionary biology to our current setting, I feel there are two lessons to take away. Firstly, we can at the very least entertain the possibility that two or more individuals may combine in such a manner that (at certain levels of description) they may be considered a single entity. Olaf and Inga are, at a relevant level of inquiry, combining to create a novel unitary cognitive system: an ICS. Secondly, we can concede that in most situations Olaf and Inga will continue to hold their own individual cognitive states, but this does not mean that they cannot also give rise to new cognitive states which belong uniquely to the ICS that they form. Given the right circumstances cognitive states may persist at any 'interactor'-level of cognition, be it a solitary cognitive module, an individual organism, a duo, or a swarm (Kennedy & Eberhart, 2001).

It should be noted that I am not trying to make any sort of claim about the evolutionary adaptive status of cognitive systems, nor that I take everything Sober and Wilson (1998) say about interactors as gospel. The aim of this section is simply to explain my reservations about discussing collective/joint/shared/group cognition, and the beneficial connotations that I believe the term 'interactor' brings to the table.

3.2.4. What of Philosophical Imprudence?

It may have occurred to the reader that it seems "philosophically imprudent" (Tollefsen, 2006, p. 142) to base the theory of cognition extending from one individual to another on an "already controversial argument" (ibid), i.e. the case of Otto and his notebook. If our only confidence in the case of Olaf and Inga stems from the similarities it shares with Otto and his notebook then any rejection of one will entail a rejection of the other.

[15] Sober (1993) has shown that individuals within a (group-level) interactor may exhibit traits which are not shared by other interactor-members, but the interactor (considered as an overall inclusive unit) will still exhibit its own traits which pervade above and beyond the individual ones. The result of this is that individual traits and (group-level) interactor traits can persist alongside one another, even if the traits are contrary to one another (Ibid). A harmonious balance is struck which benefits interactors at both the individual and group-levels of evolutionary selection.

There are three ways in which we can overcome this concern:

(i). As Tollefsen (2006) suggests and as this paper aims to verify and develop, an ICS defence of HEC will withstand many of the standard objections levelled at HEC. Even if one were to reject HEC outright, it is still possible to agree that ICS is the strongest formulation of HEC.
(ii). We have already seen that although the cases of 'Otto and his notebook' and 'Olaf and Inga' are functionally equivalent in many respects, they present some key differences, such as the fact that Inga is not an inanimate object and consequently not a simple cognitive tool, and the relationship between Olaf and Inga can be one of *reciprocal* cognitive extension. These differences will provide a plateau for the reasons why ICS provides a stronger defence of HEC than standard cases of cognition extending from a person to some artefact.
(iii). Whereas standard examples of HEC – Otto and his notebook, counting on fingers, using an iphone - are largely hypothetical or folk-intuitive, there may already be empirical evidence for ICS.

(i), (ii) and (iii) are all clearly interrelated and the latter two could be subsumed into the first, i.e. all three are ways in which to fortify the premise that an ICS is a stronger form of HEC than standard versions. If this is shown to be true then there is no philosophical imprudence in using the standard formulation of HEC (and Otto and his notebook) as a starting point. Both Tollefsen (2006) and myself believe that ICS arguments are the best way to defend HEC and these arguments are notably different and more robust than those generally used by standard HEC supporters. Clark and Chalmers' (1998) initial establishment of HEC provides a sturdy launching platform, but it is ICS that allows extended cognition to take flight!

Before illustrating this in section 4, it seems sensible to expand upon (iii) so that another feather may be added to the ICS cap.

3.3. Empirical Suggestions of ICS

In the example of reciprocally extended cognition between Olaf and Inga in section 3.2.2 we have an instance of two individuals deliberating, retrieving and reconstructing information between them. This is not some exotic hypothetical thought experiment, but something that we

encounter on an almost daily basis. Whether it be an academic discussion, a boardroom conference or friends deciding how to spend a Friday night, the conclusions that we reach are often collaborative efforts involving inputs from various people at different stages. In such scenarios disparate minds are becoming integrated in a social structure. This phenomenon is known as a *transactive memory system* (henceforth TMS): a social network of interconnected minds in which information is *encoded, stored* and *retrieved* just as it is in individual memory (Wegner, Giuliano & Hertel, 1985; Wegner, 1986, 1995). The basic idea of a TMS is that the same processes that are executed at the individual-level of memory are executed by a group of interdependent individuals who form a "knowledge-acquiring, knowledge-holding, and knowledge-using system" (Wegner et al., 1985, p. 256). Using an analogous argument to the one Clark and Chalmers (1998) originally employed in favour of HEC, we can say that TMS should be subject to a '*social parity principle*': if, as we confront some task, a group functions as a process which, were it done in the head, we would have no hesitation in recognizing as a cognitive process, then the *group* is performing that cognitive process (Theiner, 2009). The crucial point is that the processing is very much a *group* activity:

> The transactive memory system in a group involves the operations of the memory systems of the individuals and the processes of communication that occur within the group. Transactive memory is therefore not traceable to any of the individuals alone, nor can it be found somewhere "between" individuals. Rather, it is a [cognitive] property of a group.
> (Wegner, 1986, p. 191)

A TMS, then, displays manifestations of cognitive capacities which go beyond the mere aggregation of individuals' cognitive capacities (Theiner et al., 2010), in such a way that the newly-manifested capacities can only be said to belong to the system as a whole. However, they do not simply emerge whenever one person deliberates with another. If I ask a librarian where I can find a certain book, we have not formed a TMS just because I've retrieved information from her. As Tollefsen (2006) puts it:

> Individuals must participate in the encoding, storage and retrieval of information. They must be involved in the allocation of information to specific experts, for instance, and in the determination of what information will be stored. Further, transactive memory arises in a group when they are engaged in a common goal or share a common perspective[...]. As a rule we can say that transactive memory systems are properties of groups.
> (p. 145)

It seems, therefore, that a TMS requires a kind of holistic involvement from its group-members in order to create something above and beyond their individual capabilities. This is what we have in an ICS: the reciprocal extension of cognitive states means that cognitive properties can emerge at a supra-individual 'interactor'-level. For example, the encoding, storing and retrieving that we find in a TMS may be achieved by the reciprocal extension of cognitive states within an ICS. The individuals within the ICS/TMS will quite literally constitute one another's cognitive states (i.e. the cognitive states of the ICS).

We should be careful, however, not to treat *every* case of TMS as an ICS. A TMS may be considered as a mere "division of labor among cognitive agents" (Theiner et al., 2010, p. 382) and thus it is possible for a TMS to be formed among acquaintances or even strangers. For example, a study conducted by Liang, Moreland & Argote (1995) had three-person, same-sex groups learn (either individually or as a group) how to assemble radios and then collectively recall one week later how the assembly was achieved. Their aim was to show that the experience of group cooperation can stimulate the development of a TMS, and they did indeed find that groups whose members learned how to assemble the radio *together* recalled more steps of the procedure (and with fewer errors) than groups whose members were trained individually (Liang et al., 1995). The improved performance of groups trained together is often taken to show that a TMS was formed between the group-participants, i.e. they divided the cognitive labour of encoding, storing and retrieving the information about how to assemble the radio (Liang et al., 1995; Theiner, 2009). This may well be an excellent example of a TMS. However, in order for it to be an ICS, each participant must satisfy the four criteria (Clark and Chalmers, 1998) that an external resource must meet in order to be considered as part of the other participants' extended cognitive apparatus. That is, they must be readily available, accessible and reliable, and hold previously endorsed information. It seems unlikely that the radio-assemblers of Liang et al.'s (1995) TMS would meet these requirements and they thus fall short of forming an ICS.

The most accurate and successful TMSs are formed by intimate couples (Wegner, 1986), where long-term closeness implicates a strong cognitive interdependence and a mutual understanding of organisation of expertise and responsibility for information (Atkinson & Huston, 1984). For example, it was shown that long-time couples could learn and recall more words than pairs of strangers because they could divide encoding and storage responsibilities according to a tacit

knowledge of each other's memory skills (Wegner, Erber & Raymond, 1991). Indeed, a well-formed TMS may be "both a sign and a foundation of [a] successful relationship itself" (Wegner, 1986, p. 200). If we return once more to the descriptions of Olaf and Inga's enviably devoted relationship, it seems they constitute not only our paradigmatic case of ICS but also the perfect example of a working TMS.

Interestingly, it may be possible to extend even further the principle that a group, or *interactor*, can manifest cognitive properties which are more than the mere sum of the cognitive capacities of its individual members. Crisafi and Gallagher (2009), building on Hegel's (1949) notion of 'objective spirit', discuss the possibility that HEC may be extended to social institutions such as cultural practices and legal systems, which become "pieces of the mind, externalized in their specific time and place, and activated in ways that extend our cognitive processes when we engage with them" (Crisafi & Gallagher, 2009, p. 125). Bonabeau, Dorigo and Theraulaz (1999) and Kennedy and Eberhart (2001) also believe that groups of simple agents can give rise to "emergent collective intelligence" (p. xix), whereby a 'swarm' of organisms develops cognitive properties that are more than the sum of its parts. Indeed, it has already been empirically shown that organisms within hives of honeybees and colonies of ants may combine to create large unitary cognitive systems (Bosse, Jonker, Schut & Treur, 2006). These concepts have the potential to be empirically tested so that before long we may have established the explanatory power of cognition at a supra-individual level.

Once more, however, we should be wary of straying too far from the matter at hand. The importance of the empirical research into TMS is that we seem to have the potential for an empirically testable ICS that may eventually be extended to large organisations and institutions. The dyadic relationship of Olaf and Inga, in which their cognitive states reciprocally extend to one another in the form of constituted beliefs, seems to be a genuine and testable phenomenon. As things stand, I know of no similar research into any case that bears apposite similarity to Otto and his notebook.

3.4. Assembling the Artillery

I have covered a considerable amount of work in the preceding sections in a hasty manner that

perhaps does not do justice to many of the interesting and enlightening concepts that I have mentioned. However, attention should not be diverted from the HEC/ICS fulcrum of this paper.

Before proceeding, it would seem prudent to amass the key points I have attempted to bring to the fore thus far:

- Following the original arguments of Clark and Chalmers' (1998), the case of Olaf and Inga (Tollefsen, 2006) can be considered *functionally equivalent* to both Otto and his notebook and biological memory.
- The cognitive extension between Olaf and Inga can be *reciprocal* so that they form a coupled system which is constituted by parts of both of them, but belongs to neither of them individually. The agential status of such a system should not hinder discussions of the cognitive system itself.
- I call this kind of system an *interactor cognitive system* (ICS): a cognitive system formed by the reciprocal extension of cognitive states to and from two (or more) individuals which manifests its own cognitive properties that are not reducible to any of its individual constituents.
- There is empirical evidence for ICS in the shape of (certain) *transactive memory systems* (TMS).

With this artillery assembled we can advance to 'Part Two' and the arguments in favour of an ICS approach being the most effective form of defence for HEC.

PART TWO

4. The ICS Defence of HEC

4.1. The Criticisms

Three of the major criticisms HEC has faced involve the distinctive nature of internal cognition (Adams & Aizawa, 2001, 2008), the problem of cognitive integration (Weiskopf, 2008; Spaulding, 2011), and the need for a 'persisting locus of cognition' (Butler, 1998; Rupert, 2004, 2010). I will tackle each of these in turn.

4.2. The Distinctive Nature of Internal Cognition

Perhaps the most persistent of criticisms that HEC has faced is the idea that "as a matter of contingent fact, the cognitive processes we find in the real world all happen to be brain bound" (Adams & Aizawa, 2001, p. 46) and any purported cases of HEC are in fact fallacious. According to this view, it is perfectly true that human cognitive processing interacts with the extra-cranial world and HEC is certainly "a logical and a nomological possibility" (ibid, p. 43). However, it is untrue that external media do *as a matter of fact* constitute cognitive states; our current cognitive scientific knowledge renders it "highly unlikely" (ibid, p. 47) that some environmental resource could present the cognitive properties that our brains do.

The thinking that propels this argument is that we should define and delimit cognition according to a 'mark of the cognitive' (henceforth, MoC), and although a full MoC cannot be proposed, we know enough about the psychological and physical processing of the brain to put forward a few necessary criteria. For Adams and Aizawa (2001) the "first essential condition on the cognitive is that cognitive states must involve intrinsic, non-derived content"[16] (p. 48). This means that unlike words in a book or numerals on a screen, which derive their meaning from social conventions or practices, cognitive states only depend on their *intrinsic* properties and do not derive meaning from elsewhere. Let's call this 'condition (A)'. If such a condition is indeed a criteria of MoC then contrary to externalist theories, a being having a cognitive state, X, supervenes solely on the intrinsic properties of X (Lau & Deutsch, 2010). External media may have a *causal* effect on cognitive processing, answering *why* certain processes take place, but they cannot have a *constitutive* effect, answering *how* these processes are carried out.

Coupled with condition (A) is the fact that "the sorts of processes that occur within brains seem to share certain sorts of regularities that they do not share with systems consisting of brains coupled with tools" (Adams & Aizawa, 2001, p. 63). In other words, internal cognitive processes are characterised by specific laws and mechanisms that do not always extend beyond

[16] As Hurley (2010) notes, there should really be a distinction drawn between 'intrinsic' and 'non-derived': "Various accounts of intrinsic content appeal to causal, historical, functional, or other relations, excluding social relations that presuppose intentional mental content. But content is no more intrinsic to brains in virtue of their relations to nonsocial environments than their relations to social environments. Underivedness is not the same as intrinsicness." (p. 129) However, as clarifying this distinction would involve a lengthy discussion, I will continue to use 'intrinsic' for present purposes.

the organismic skin/skull boundary. For example, if we consider the cognitive processes behind normal human memory we may encounter the 'generation effect' which states that information, e.g. the word 'cognition', will be better remembered if it is generated (e.g. constructed from fragments of letters, 'c_gn_ti_n') rather than simply read in its entirety (Jacoby, 1978). An external memory source such as Otto's notebook would not be subject to such an effect, nor the 'law of effect', the 'primacy effect', the 'recency effect', the 'spacing effect' etc. (Adams & Aizawa, 2008). All notebook-information is 'remembered' indiscriminately, whereas internal cognitive processes are subject to quirks and biases which result from the laws and mechanisms that govern them. This gives us 'condition (B)': internal cognition exhibits unique characteristics due to the laws and mechanisms of the brain.

Taking conditions (A) and (B) together, the crux of the argument is that internal cognition has a *distinctive nature* of "non-derived representations governed by idiosyncratic kinds of processes" (Adams & Aizawa, 2008, p. 10). The external resources of HEC-cases cannot replicate these conditions and consequently those processes that merit the 'cognitive' label only occur intra-cranially.

The intuitive appeal of this kind of *internalist* stance is nicely summed up by Sprevak (2010):

> We appear to carry our cognitive mechanisms around with us, and we appear to be capable of deploying our cognitive mechanisms across a range of different environments. Surely then our cognitive mechanisms are internal to us, and their study a matter of specifying internal processes?
> (p. 356)

If this kind of intuitive appeal can be vindicated, as Adams and Aizawa (2001, 2008) believe it can, then the bounds of cognition can be firmly established around those neural processes carried out in our heads.

HEC supporters, however, are not prepared to give in so easily. Firstly, for the most part they have no need to disagree with much of what Adams and Aizawa (2001, 2008) propose. It is true that the laws and mechanisms present in the work of current cognitive disciplines seem to describe internal cognition and it is true that the external cognitive states seem to have derived content. But what does this matter? In response to condition (A), there is no justifiable reason

to expect that all things 'cognitive' must involve intrinsic content. Clark (2010a) gives the example of manipulating Venn diagrams in one's head; the cognitive processing involved in such manipulation initially seems to involve intrinsic content, but there is arguably a sense of deriving content from the *inspection* of the internal representation of the diagrams (Clark, 2010c). This is true of much of our cognitive processing. A person calculating the sum '152 x 67' may internally envisage multiplying the units and tens separately, creating two rows which are then added together, or they may internally envisage the manoeuvring of an abacus. Similarly, a traveller may plan their route by mentally picturing a map, or, if the roads are known to them, by picturing street names that must be followed in order. In each case, the internal representations rely on the manipulation of content (e.g. numerals, abacus, words, map) with humanly-assigned meaning, i.e. the cognitive content is traceable to human social conventions and is therefore, in some sense, 'derived'. Clark's (20010a) example thus raises an interesting point - the fact that for many cognitive processes, it is unclear at what point the cognitive content can be considered 'intrinsic', or if there is ever such a thing as truly 'intrinsic' content (Dennett, 1998; Hutto, 1999). Even Adams and Aizawa (2008) concede that "not every component of every state of every cognitive process must bear non-derived content" (p. 50), and we surely cannot chauvinistically take 'intrinsic' to simply mean 'carried out by the physical substrate of the brain'. If we cannot pinpoint the 'intrinsic-ness' of cognitive content then how can it justifiably be argued that intrinsic content is an aspect of MoC?

This leads us to a second way to reject condition (A). If we cannot assuredly say where intrinsic content lies within a cognitive system, or at the very least we concede (as Adams & Aizawa (2008) do) that not *every* element of a cognitive system must involve intrinsic content, then why should we exempt the Otto-notebook system from cognitive status? The writing in Otto's notebook only bears derived content, but it can be properly integrated into any "larger cognitive economy" (Clark, 2010a, p. 49) so that the *overall system* of the neural and bodily workings of Otto plus his notebook still has (somewhere, perhaps) intrinsic content. If we are to discount Otto's notebook from being a cognitive component of a cognitive system, then any internal machinations which seem to have derived content (such as the Venn diagram manipulations) should also not be considered 'cognitive'. Condition (A) is thus seemingly being excessively restrictive (Menary, 2006). And this is surely not what Adams and Aizawa desire; they want to confine the mind to the brain, but not shrink it to the content of enigmatic

cognitive states within the brain. Intrinsic content, in a clarified form, is central to certain cognitive scientific debates[17] but it need not extend or permeate every aspect of a cognitive system.

With regards to condition (B) the response is simple. As previously mentioned, few would deny that internal cognition exhibits idiosyncratic kinds of processes due to specific laws and mechanisms. But why should we consider the laws and mechanisms of current cognitive science to be definitive of *all* cognition, rather than just internal cognition (Spaulding, 2011)? The very essence of the HEC approach is the desire to expand cognitive science to include the non-neural features of human life which aid and amplify our cognitive capabilities. HEC does not refute the well-established laws and mechanisms of internal cognition, it merely proposes that a more welcoming and expansive methodology may include other (external) laws and mechanisms which will ultimately enhance our knowledge of the human mind. The fact that Otto's notebook will not be subject to the 'generation effect' or the 'primacy effect' does not in itself discount the notebook from constituting a genuine part of the memory system, it simply means that the notebook is not *exactly the same* as purely internal memory. Such a difference is irrelevant to the overall functioning of a cognitive system. What Adams and Aizawa are in fact doing by suggesting that an aspect of MoC is idiosyncratic internal processing is giving "immediate priority to in-the-head processes [by] setting the benchmarks against which the rest are to be judged" (Clark, 2010c, p. 1050).

Personally, I find these HEC responses compelling. Yes, internal cognition has a 'distinctive nature', but HEC can be included into this nature and adds the possibility of building upon it. If unconvinced, however, we can now see how the problems put forward by Adams and Aizawa (2001, 2008) can be dissipated more easily from an ICS perspective.

Firstly, let's return to condition (A). The initial challenge put forward by Clark and Chalmers (1998) was for an opponent to show that "Otto's and Inga's[18] cases differ in some important and relevant respect" (p. 6). Adams and Aizawa (2001, 2008) believe they achieve this by highlighting the intrinsic content of Inga's biological memory, in comparison with the derived

[17] For example, most people agree that intrinsic content is central to work on consciousness.

[18] Here, Inga should be considered as from the initial Clark and Chalmers (1998) example, rather than being reciprocally-dependent with Olaf.

content of Otto's notebook. If intrinsic content is indeed an aspect of MoC (which seems unlikely) then the notebook can no longer constitute part of a cognitive system. But what of Inga in the Olaf-Inga ICS? As mentioned in section 3.2.1 Inga is not merely a 'cognitive tool'; she is a living, cognizing being that seems to transcend the traditional definition of subsystemic cognitive apparatus. When considered individually, Inga, unlike the notebook, has intrinsic content in the same way any individual being does. Inga can think for herself, she has beliefs, desires, intentions, knowledge and consciousness that are all her own. Similarly, Olaf has his own cognitive states with intrinsic content. If we were to break down an ICS into its independent (non-fused) subsystemic components, every component would have intrinsic content (in as far as an individual brain does), whereas breaking the Otto-notebook system into independent components would leave some components with intrinsic content (those in Otto's brain) and others with derived content (those involving the notebook). In other words, both Olaf and Inga, autonomously, have intrinsic content, whereas although Otto does, his notebook is not *on its own* "intrinsically cognitive" (Clark, 2010b, p. 89). The problem is that the Olaf-Inga ICS is not the two of them taken individually, with their own intrinsic content, it is a fusion of the relevant parts of their minds to create a new and unique cognitive system. So the question becomes whether this ICS can *itself* manifest intrinsic content?

Consider the scenario of Olaf and Inga writing to a friend. Olaf can only remember the surname, Inga can only remember the first name. Olaf remembers the friend lives on 'Memory Lane' but cannot remember the house number, Inga knows it's number 36 but cannot recall the street name. A similar pattern follows for the rest of the address. Olaf and Inga can each only remember snippets of the required information and need their spouse to fill in the rest. Each item of information that Olaf retrieves from Inga has been previously endorsed by Olaf, and Inga has previously endorsed all information she retrieves from Olaf. The writing of the address is thus achieved through a process of joint deliberation, retrieval and reconstruction (Tollefsen, 2006); a process which is "active and not found "inside" the heads of Inga and Olaf" (ibid., p. 145), but rather belongs to the overall ICS. If Olaf were to encode, store and retrieve his friend's address individually then we would automatically attribute the remembering of the address to Olaf. Being just, if the encoding, storage and retrieval is carried out by the ICS (and not individually by Olaf or Inga) then the remembering is done *by the ICS*. It is the 'joint-ness' - the fusion of cognitive states to create a system which is not reducible to either Olaf or Inga

individually - which proves crucial. The manifestation of cognitive states within any cognitive system should be indiscriminately viewed through a "*veil of metabolic ignorance*" (Clark, 2007, p. 167), so that the cognitive states belong to the *system itself* and are not chauvinistically attached to a physically-bound individual. What we have with any cognitive system then, is a 'mind' made up of the collective collaboration and commingling of subsystemic cognitive 'parts' (processes and states), the properties of which supervene on the whole system and cannot be reduced to any individual part. This description is equally applicable to an ICS or an individual (single-subject) 'mind'. In as far as Adams and Aizawa (2001) believe "it is not by anyone's convention that a state in a human brain is part of a person's thought that the cat is on the mat" (p. 48), it can similarly be said that it is not by anyone's convention (outside of the ICS or *individually* within it) that a state in an ICS is part of the system's thought that X. Of course, this is not to say that an ICS is an impermeable structure, it is after all an *extended* system and as such any external media can potentially comprise part of its cognitive processing. Nevertheless, at any given point it should be possible to delineate the ICS according to those active parts that constitute it, just as you can for an individual 'mind'. Therefore, if one takes the internalist view that brains have intrinsic content because the thoughts within a brain are *intrinsic* to the cognitive states within the brain and do not derive meaning externally from social conventions and practices, then one should equally accept that an ICS can have intrinsic content because ICS thoughts are *intrinsic* to the cognitive states of the ICS.

It may be argued that the talk of 'fusion' and 'commingling' to form an ICS is simply persuasive jargon and the interaction between two individuals is actually a social interaction involving much derived content, e.g. Olaf is deriving (parts of) many of his cognitive states from Inga and vice-versa. On this view there are no intrinsic content ICS states; there is only intrinsic content within Olaf and Inga individually, and the supposed cognitive system they form is just a compilation of their personal cognitive states. But this view would not only overlook all the initial arguments and evidence in favour of ICS, it would also not be a hugely damaging critique. It seems that there is no relevant difference between Olaf and Inga deriving content from one another (to form ICS-states) and an individual deriving content from inspection of an internal representation, as with Clark's (2010a) Venn diagrams. In both cases, derivations may be present but they actively "participate in processes that invoke intrinsic contents" (Clark, 2010c, p. 89)

Therefore, either (a) if an individual brain is said to bear intrinsic content simply in virtue of being a cohesive system with subsystemic 'parts' whose properties supervene on the whole system, then the same should be said of an ICS, or (b) neither an individual brain nor an ICS has every cognitive component of every state bearing intrinsic content.

Returning to condition (B), the argument is once again more straightforward. The criticism of Otto's notebook is that it is not subject to the idiosyncratic laws and processes that govern internal cognition. This is, quite simply, because it is a notebook: it is a pile of inanimate sheets of paper bound together and it consequently stores all information indiscriminately. Both Olaf and Inga, however, are cognizing beings who *will* be subject to the same laws and processes that govern internal cognition. If Olaf is to use Inga as an external store for some of his beliefs, these beliefs will encounter the same kinds of "reorganizations, interpolations and creative mergers" (Clark, 2004, p. 21) as they would in his own head. No new idiosyncrasies need to be introduced - the information that Inga and Olaf hold for one another will be stored in the same manner as if they received the information (in the same way) individually. We are still primarily dealing with 'brain-law'.

An internalist, however, may reply that the problem here is that the reorganisations, interpolations and creative mergers of information Inga is holding for Olaf will be *Inga's and not Olaf's* (and vice-versa). The worry is that the information will be processed according to Inga's character and in conjunction with her other cognitive states, which may produce different 'storage' to if the information was held in Olaf's individual memory. There are two issues here. Firstly, this once more overlooks the nature of the ICS. The reorganisations, interpolations and creative mergers of the ICS belong *uniquely* to the ICS, considered as a kind of *interactor*, not to any individuals. Secondly, even if we were to accept that Inga idiosyncratically processes Olaf's extended cognitive states and vice-versa, rather than the processes and states belonging to the system as a whole, the problem can still be discarded. Normal memory is not faultless: we often unwittingly misremember (Nessier & Fivush, 1994), confabulate (Della Barba, 1993) and self-deceive (Trivers, 2011). Consequently, information that Olaf stores individually may be subject to different reorganisations, interpolations and creative mergers depending on his current intentions, mood or surroundings, and in no situation is it guaranteed to be retrieved in a wholly reliable manner. If his own memory is inconsistent with its idiosyncratic processing, then why should Inga be considered an unsatisfactory external store? At the very least Olaf and

Inga will subject any extended cognitive states to the same *kinds* of laws and mechanisms as internal cognition, unlike Otto's notebook. If (and it is a big 'if'), "the cognitive must be discriminated on the basis of underlying causal processes" (Adams & Aizawa, 2001, p. 52), then ICS states maintain (currently-conceived) 'cognitive' status while standard HEC states do not.

Summating the arguments of the preceding paragraphs we can see that ICS not only allows for cognitive extension, it also preserves the 'distinctive nature' of current cognitive psychology. It thereby dissolves one of the most persistent criticisms HEC has faced, successfully walking the tightrope between the aspiration of a more expansive cognitive science while safeguarding the currently privileged status of internal cognition.

4.3. Cognitive Integration and Parity

The second HEC-criticism to which I will respond focuses on cognitive integration and parity. The basic idea behind the criticism is that "integration is part of the everyday dynamics of belief" (Weiskopf, 2008, p. 68) and whenever we receive some new information it is incorporated into a pre-existing store of beliefs, all of which are updated accordingly. Whereas internal cognitive processes carry out this kind of 'updating' automatically, external processes are incapable of analogous integration because they remain distinct from the well-defined mechanisms and idiosyncrasies of internal cognition, and this results in a notable imparity. Such an idea can be split into two, and this is the strategy that Spaulding (2011) takes, arguing that the extended cognitive states comprising HEC-beliefs fail to be properly integrated at (i). cognitive psychological, and (ii). folk-psychological levels of interpretation. Before fleshing out these criticisms it is necessary to say a little about the HEC proponent's views on *functional role*.

For Clark (2008b), and his HEC followers, 'functional role' should not depend on fine-grained distinctions. What this means is that as long as an external process plays the same coarse-grained functional role as an internal process, it should be considered a cognitive process. For example, if both a human and a martian are asked to calculate '596 x 39', the fine-grained implementation of their cognitive mechanisms may well be very different (one may involve green slime instead of neurons), but we would not regard the martian's mechanisms as non-

cognitive just because of the fine-grained divergence from our human mechanisms. Similarly in the case of Otto, the fine-grained mechanisms behind his reading a notebook entry or using his internal memory may differ greatly. At a coarse-grained level of interpretation, however, both an internal belief and a notebook-belief are *relevantly equivalent* in terms of how they guide Otto's behaviour and allow him to navigate his social and physical world. Clark (2008b) refers to this as judging processes according to their "functional poise" (p. 88), i.e. the way in which a process guides behaviour and is organised for "a certain kind of use within a specific problem-solving routine" (ibid, p. 87). In these scenarios, the suggestion is that coarse-grained functional similarity is *sufficient* for achieving cognitive status, and we should not exhibit biases due to differing fine-grained implementations of mechanisms.

Spaulding (2011) entitles this suggestion the 'Sufficiency Claim'. She believes such a view is unjustifiable according to a correct interpretation of the Parity Principle and true functional equivalence requires more than mere coarse-grained 'equivalence'. This then acts as a platform for her central claim that supposedly 'cognitive' external processes actually fail to achieve "sufficient parity" (ibid, p. 19) at *either* (i). cognitive-psychological (fine-grained) *or* (ii). folk-psychological (coarse-grained) levels of interpretation. I believe that Spaulding's (2011) construal of the Parity Principle is incorrect,[19] and fine-grained parity is not necessary to achieve

[19] Spaulding (2011) believes the Parity Principle is the "implicit defense" (p. 11) of the 'Sufficiency Claim' and hence much more than a 'veil of metabolic ignorance'. She writes:

> The parity principle[…]says that if we come across a trans-cranial process that would count as cognitive were it an intra-cranial process, then we should count that trans-cranial process as cognitive. But what would be our reason for regarding the trans-cranial process as cognitive if it were in the brain? Surely not simply the fact it occurs in the brain. Presumably the reason would be because the trans-cranial process *functions like an intra-cranial process*, i.e., it has the same functional role as an intra-cranial process.(p. 11)

On this reading, the Parity Principle allows for any external process to be treated as cognitive as long as it "*functions like an intra-cranial process"*, thereby assuming that internal processes are "uncontroversially cognitive" (Spaulding, 2011, p. 11). The problem, as Spaulding (2011) sees it, is that functioning 'like an intra-cranial' process can be interpreted almost ambiguously; it allows the HEC supporter to gloss over significant functional differences between internal and external processes, as long as there is functional similarity *on some level*. The Parity Principle thus acts as the "implicit defense" (ibid, p. 11) of the 'Sufficiency Claim', *either* fine- *or* coarse-grained parity is sufficient to establish cognitive status, while other important distinctions can be ignored.

However, when we encounter some external process, we do not count it as cognitive because it "*functions like an intra-cranial process*" (Spaulding, 2011, p. 11), but because it has the correct "functional poise" (Clark, 2008b, p. 88). To expect any 'cognitive' process to '*function like an intra-cranial process*' is to half-turn one's back on the core functionalist concept of *multiple realisability* (i.e. the potential for cognitive states to be instantiated by various underlying physical realisers (Wilson, 2004)) and to assume that internal processes are 'cognitive' just because they are intra-cranial, rather than subjecting them to psychological inspection.

The fault here does not lie solely with Spaulding (2011). The Parity Principle should not state that we should consider as cognitive any external process which *were it done in the head*, we would automatically count as cognitive (Clark & Chalmers, 1998). It should be read as arguing that we ought to regard all potentially cognitive processes with 'fair-treatment' (Sprevak, 2009), so that they are judged according to their functional poise and resultant behaviour, and not according to their location or physical realisers. *Both* internal and external processes should be subject to functional-poise-questioning, i.e. we cannot assume internal processes are 'cognitive' just because they are intra-cranial and we cannot exclude

cognitive status. For present purposes, however, I will accept her argument in an unadulterated form and show how an ICS approach withstands both (i) and (ii).

At a cognitive-psychological (fine-grained) level of interpretation the worry is that internal cognitive processes naturally undergo a kind of *cognitive integration* that external processes fail to replicate. Spaulding (2011) chooses the mechanism behind internal semantic memory and gives a lengthy description of how it "is intricately integrated with other cognitive functions of the brain" (p. 14). A similar description could be given for all other internal processes; each is not an isolated phenomenon, but a part of an intricate network of closely-associated mechanisms. Encapsulating this idea for the cognitive states of beliefs, Weiskopf (2008) writes:

> Genuine beliefs as they occur in normal biologically embodied systems are informationally integrated with each other, and sensitive to changes in the person's overall system of beliefs. Environmental states, however, fail to satisfy this central feature of the functional role of belief, and hence fail to be genuine mental states.
> (p. 265)

This idea of cognitive-informational-integration stems from Fodor's (1983, 2000) concept of *mental isotropy*, whereby an organism's beliefs and desires are globally interconnected so that everything a person knows is relevant in constraining and determining all his other beliefs. To put this whimsically, all of our botanical beliefs are constrained by our astronomical beliefs (Fodor, 1983), all of our beliefs about solar physics are constrained by beliefs about evolutionary theory (Carruthers, 2006), and all of our philosophical beliefs are constrained by beliefs about football. An individual cognitive state (e.g. belief) cannot be extracted from the overall cognitive system and studied as a singular occurrence; every cognitive state is, in some way, connected to every other cognitive state. I do not wish to discuss the veracity of *mental isotropy*[20] but the concept behind it sheds light on a problem for most HEC cases - that they lack the integrated interconnectedness of the rest of the brain. External resources have no sort of 'updating system' due to their indiscriminate storing of information. The stoic nature of Otto's notebook prevents

external processes from being 'cognitive' just because they are trans-cranial.

[20] If mental isotropy is taken as true then there is the problem that "[t]he more global (e.g., the more isotropic) a cognitive process is, the less anybody understands it" (Fodor, 1983, p. 107), i.e. if 'global interconnectedness' is really true then we lose the possibility of ever being able to genuinely study *singular* cognitive states and so the study of internal or external cognition, as they are currently conceived, becomes a moot point.

it from developing the integrative nature that internal cognitive states inherently share with one another. Even if the notebook were shrunk and placed within a brain, it would still fail to be related to brain mechanisms in the way they are related to one another, because it lacks the fluid and automatic updating mechanism of internal cognition (Spaulding, 2011). If Otto is told that the Museum of Modern Art is knocked down, he may erase his notebook entry saying 'the Museum of Modern Art is on 43rd street', but this will not automatically update his entry saying 'the museum café makes an excellent latte' (which would evidently no longer be true). Conversely, in biologically-based memory such updating and informational integration is standard procedure (Weiskopf, 2008).

The HEC proponent can reply to this criticism with the argument that Otto's externally stored beliefs (in his notebook) *are* in fact subject to belief-updating mechanisms, they are simply slower and implemented differently to internal mechanisms (Spaulding, 2011). For instance, Otto may employ a complex indexing system so that whenever he alters one of his notebook entries he is immediately directed to connected beliefs which may need consequential alteration. This would not be ideal in practice, but it means that all of Otto's notebook-beliefs could theoretically be globally interconnected. Indeed, a slow and methodical updating system such as this may be considered analogous to the internal updating mechanisms of a stubborn fanatic, for whom the revision of certain beliefs may be a long-winded and deliberate process (ibid). Moreover, a more complex HEC scenario may render this Weiskopf/Spaulding (2008/2011) criticism obsolete, e.g. a "hippocampus-in-a-distant-vat [...may be] so well integrated as to unproblematically count as part of my cognitive apparatus" (Wilson & Clark, 2009, p. 20).

Rather than develop these arguments in detail I intend to show that, once again, an ICS perspective provides the most robust defence of HEC. Weiskopf (2008) writes that beliefs "are not just acquired and inertly stored, waiting to be retrieved; even in storage they are subject to processes that modify and sometimes eliminate them" (p. 275). Contrary to the 'inert' storage of Otto's notebook, ICS-stored-beliefs *are* subject to appropriate modifications. As the previous section explained, an ICS (unlike standard HEC cases) deals almost exclusively with the same laws and mechanisms as internal cognition. That is, for the most part it preserves the *distinctive nature* of current cognitive psychology. As a result, any externally stored beliefs in an ICS will be appropriately informationally integrated with the rest of the cognitive system. If Olaf uses Inga

as an external store for his belief that 'the Museum of Modern Art is on 43rd street' then when he informs her that the museum has been knocked down Inga will not only adjust the belief about the museum's location, but all other connected extended beliefs, such as 'the museum café makes an excellent latte'. The success of ICS in this respect is down to the involved individuals being *active* components of the system (Tollefsen, 2006). Through being 'active' both Olaf and Inga have the capability to implement belief-updating mechanisms analogous to those of internal cognition. They are capable of subjecting any extended cognitive state to all of the same "reorganizations, interpolations and creative mergers" (Clark, 2004, p. 21) that they would be subject to if stored internally. The physical realisers (e.g. the substrate giving rise to cognitive processes and states) of an ICS, unlike most standard HEC cases, are the *same* as those of strictly internal cognition, meaning that if we are to examine the fine-grained implementation of cognitive mechanisms we find the same kinds of processes and information-integration in both individual brains and ICSs. For example, any (dispositional) belief will generally encounter the memory stages of *encoding, storage* and *retrieval*. At a fine-grained level, these stages will involve neural activity in the neocortex, hippocampus and left prefrontal cortex, brain areas which are also respectively implicated in the control of voluntary actions, episodic memory and speech production (to name but a few connected processes; Spaulding, 2011). With Otto's notebook, connection to the neural activity of these brain areas is lost; the words in the notebook are not appropriately integrated with the brain. When cognitive states extend from one person to another, however, the same brain areas come in to play and the cognitive machinery is *homogenous* across all components. For example, when Olaf extends a belief to Inga, both of their neocortices will be involved in the encoding of a short-term memory, Inga's hippocampus will consolidate this short-term memory to a semantic memory and then store it, and her left prefrontal cortex will be involved in retrieving it; it will then re-emerge (after communication) as a short-term memory for Olaf (again involving his neocortex).[21] This is clearly an oversimplification of the complex neuroscience behind the formation of a belief, only picking out the three primary stages of neural activity, but the key point is that the same processes - encoding, storage, retrieval - are taking place and the same brain areas - neocortex, hippocampus, left prefrontal cortex - are involved for both individual internal memory and

[21] Descriptions of the mechanisms behind memory that are mentioned in this paragraph can be found in most cognitive psychology textbooks, e.g. Gazzinga, Ivry & Mangun (2002); Patestas & Gartner (2006).

extended memory. This is equally true in the *reciprocal cognitive extension* of an ICS, the back-and-forth dialogue of Olaf and Inga means that certain brain areas will be repeatedly (re-)engaged, but it is still *relevantly* the same brain areas and, crucially, the same kinds of processes that are involved.

The sceptic may respond to this with two connected worries: (a). an internal cognitive state and external ICS state are still not achieving exact fine-grained parity, and (b). this is because of a crucial difference between individual internal processing and ICS external processing: namely, in an ICS, the implementation of the neural activities behind vocal output and auditory input are employed, i.e. *communication* is involved.

With regards to (a), it is not only unclear why exact fine-grained parity with internal processes should be required for an external process to count as 'cognitive', it is also unsustainable that it should be required. Due to the back-and-forth dialogue between Olaf and Inga the neural activity behind the formation, storage and retrieval of a cognitive state may not be *exactly* the same in an ICS as it is in a strictly internal cognitive system, but why should it be? An individual may come to hold a belief via any kind of sensory input, be it visual, auditory, olfactory or otherwise. In each case the fine-grained neural activity involved in receiving and storing the belief will be slightly different; indeed, even if the sensory input is the same, an individual's neural activity may vary due to their age, mood and surroundings (e.g. Ashby, Isen & Turken, 1999; Smit & Rogers, 2000). Why, then, should ICS extended states need to be *exactly* like internal cognitive states when exact parity even within internal cognition is such a rare thing? If we are to accept Spaulding's (2011) stringent interpretation of the Parity Principle it may be that no two processes (internal or external) will achieve the *exact* fine-grained parity that seems to be required. As Tollefsen (2006) says, if it is necessary for functional parity "that the function be performed by the [exact] same mechanism then accurate judgements of functional equivalency, even among and within uncoupled biological systems, will be rare indeed" (p. 144). What an ICS does achieve, that most standard HEC cases do not, is an adequate kind of fine-grained parity with internal cognition, wherein the same cognitive machinery is used, employing the same kinds of cognitive mechanisms and processes which result in the same kind of cognitive

integration.[22] If more than this is required then it seems there will never be strict fine-grained parity, even amongst internal cognitive processes.

With regards to (b), the suggestion is that the prevention of fine-grained parity between internal and external ICS processes is due to the *communication* involved in ICS processing. This issue similarly arises at a coarse-grained level of interpretation and therefore encroaches upon both strands of Spaulding's (2011) argument. Undoubtedly, the presence of communication entails a difference of sorts, but I fail to see how this difference is relevant to cognitive integration and parity at any level.

Recall Clark's (2010a) discussion of the internal manipulations of Venn diagrams. What happens in this example is that the brain is turning itself upon an internal representation that it has itself manifested. This can be described as a kind of internal *inspection* (Clark, 2010c). Along with this example, consider the folk-psychological notion of 'inner dialogue' (a kind of conversation within one's head) that is used in accomplishing daily tasks or thinking through problems. I believe that the *inspection* of an internal representation, or of 'inner dialogue', is relevantly indifferent to the communication within an ICS. To reiterate, an ICS is a cognitive system formed by the reciprocal extension of cognitive states to and from two (or more) individuals which manifests its own cognitive properties that are not reducible to any of its individual constituents. As its cognitive properties are irreducible to subsystemic components, any communication that takes place *within* the ICS is *internal* to it, in much the same way that the inspection of Venn diagrams, or 'inner dialogue', is *internal* to an individual. Both the dialogue of an ICS and an individual's internal inspection can be viewed as a kind of *intra-communication* that is contingently internal to a unitary cognitive system. The fact that vocal and auditory systems are employed in an ICS is indeed a difference of sorts, but it is a negligible difference at both a fine-grained level (where the same primary brain areas and processes are implicated) and a coarse-grained level (where the same functional poise is maintained).

The further coarse-grained difference that Spaulding (2011) espies is that "the standard folk psychological conception of belief holds that we can have degrees of belief" (p. 15). That is,

[22] This is not intended to suggest that there is any such thing as a *cognitive 'kind'*, but that if we accept Spaulding's (2011) view that external processes need to function like "uncontroversially cognitive" (p. 11) internal processes, then ICS processes achieve much greater fine-grained resemblance to internal processes than standard HEC cases do.

certain beliefs are held more strongly than others and are more (or less) steadfast in their willingness to be manipulated and updated according to newly-learned information. Such a concern is again centred round cognitive integration, i.e. normal (internal) beliefs have a level of certainty attached to them and are integrated in an individual's cognitive processes in such a way that they can be globally scaled according to introspective comparisons, whereas external beliefs are not appropriately integrated. Otto's notebook is incapable of having 'degrees' of belief (unless a complex indexing system were introduced) because all notebook-beliefs are impartially stored in the same way. Once more, however, an ICS *is* capable of presenting degrees of belief. As this paper has hopefully made apparent by now, no aspect of an ICS is inert, and consequently beliefs can be actively integrated in a manner analogous to that of internal processing. Inga can scale Olaf's belief that 'the museum café makes an excellent latte' according to the manner in which he endorses it and her active knowledge of Olaf's cognitive states and mannerisms. The Olaf-Inga ICS is an integrated and fluctuating structure of cognitive states and processes, just as their individual brains are.

In sum, an ICS yet again provides the most robust defence against criticisms regarding cognitive integration and parity. Its *active* nature, along with its preservation of the same laws and mechanisms that govern internal cognition, allow an ICS to fully integrate its cognitive states and maintain relevant equivalence with internal cognition at both a (i). cognitive-psychological (fine-grained), and (ii). folk-psychological (coarse-grained) level of examination.

4.4. The Persisting Locus of Cognition

The final critique to which I will respond is that HEC loses the 'persisting locus' of cognition. This worry has been proposed in various guises, most notably by Butler (1998) and Rupert (2004, 2010). Butler (1998), for example, claims:

> There can be no question that the locus of computational and cognitive control resides inside the head of the subject and involves internal processes in a way quite distinct from the way external processes are involved.
> (p. 205)

The idea behind this is that although the external world is causally essential to a cognizing

organism, it only ever plays a supplementary role and 'inside the head' maintains executive authority. Rupert (2004), in a similar vein, claims that although cognitive processes "depend *very* heavily" (p. 393) on external resources, the persisting locus of cognition remains contingently in the organism. Both authors believe that in spite of the importance of environmental structure and "organismically external props" (ibid, p. 393), it is the brain that has the last word and ultimately governs our thoughts and actions. It is our "naked" brains that are responsible for transforming the environment and manipulating external artefacts (Norman, 1991). The concern is that a HEC approach to cognitive science diverts attention from the most important part of cognition - the stuff that goes on in the head. Thus, Rupert (2010) claims:

> The fundamental theoretical construct of virtually all successful cognitive science[…]is that of a persisting architecture interacting with an ever-changing cast of external materials to produce intelligent behavior. If there is any theoretically interesting divide between what is distinctively cognitive and what merely causally contributes to intelligent behavior, it is to be found in the persisting, integrated nature of cognitive architectures.
> (p. 344)

If we proceed to study all external resources which aid and scaffold our internal cognition, then we may rob ourselves of the persisting cognizing organism and thereby commit to a counterintuitive view of 'the self' (Dartnall, 2004). Clark (2010a) summarises these anti-HEC views as proposing that "the brain is the controller and chooser of actions in a way all that external stuff is not, and so the external stuff should not count as part of the *real* cognitive system" (p. 55). Returning to the case of Otto, we know that the notebook is clearly vital to his cognition. But according to Butler (1998) and Rupert (2004, 2010), it is Otto's brain that holds the power: it is the 'central controller' of Otto's cognitive life. The notebook only takes on its 'cognitive' role because of the way Otto's brain engages with it and creates a dense and complex interactive system of which both brain and notebook become constitutive parts.

From a commonsense perspective, a less-radical approach to cognition seems more acceptable than the revolutionary HEC. We use external props to augment our cognitive ability on a regular basis, just as Otto does, but the gut-reaction is that such environmental exploitation does not amount to the external props becoming a part of our minds. Our brains still do all the hard work, they're just making the most of what's in front of them. Interestingly, I'm not sure if this

folk-intuition holds as much sway when it comes to ICS. We are fundamentally social beings; before we had tools or technology, we had each other. We are biologically-wired[23] to depend upon one another and we are rarely, if ever, fully independent from a network of social relationships, group memberships and self-identities (Smith, 2007). In comparison with standard HEC cases, an ICS may seem more 'natural':

> As closely connected as Otto and his notebook may be, the bonds of connectivity between people is far greater because we have adapted, over generations and within our lifetimes, to be sensitive to each others' smell, sight, behaviors, creations, emotions and thoughts.
> (Theiner et al., 2010, p. 380)

It can even be argued that this connectivity between people is increasing over time due to technological advancements (ibid). Folk-intuition may thus be more welcoming to the idea of cognition extending amongst people, rather than between a person and a lifeless artefact.

From a more philosophical perspective, Clark (2010a) argues that placing the persisting locus of cognition within the head is actually more problematic than it seems. For where exactly in our neural matter is this locus meant to reside? Should we discount as 'cognitive' "any neural subsystems that are not the ultimate arbiters of action and choice?" (Clark, 2010a, p. 55). The assumption that the critics seem to make is that the mind contains some kind of authoritative system that governs the rest of it. But there is no scientific suggestion of this. What if no neural subsystem has the "final say" on cognitive behaviour (Dennett, 1997, 1998)? To where, then, should the locus be attributed? The response cannot be that it should be attributed to the brain in its entirety because for a large number of cognitive processes external resources will be playing a more significant role than certain brain areas. The whole brain is not *active* in every cognitive endeavour, even if the whole brain is passively interconnected. The idea of a specific 'locus' of cognitive control thus seems more of a convoluted proposal than the HEC-critics suggest. In spite of this, Clark (2007) has sought to appease them by conceding that although cognition can extend, it remains "organism-centred" (p. 192) and the brain/CNS is "currently the most active element" (ibid) of cognitive processing.

[23] For example, the release of hormones such as oxytocin has been implicated in the development of trust, empathy and bonding (Domes, Heinrichs, Michel, Berger & Herpetz, 2007; Kosfield, Heinrichs, Zak, Fischbacher & Fehr, 2005)

Yet again, I believe an ICS point of view comes up trumps in the face of this criticism and will be even more successful in appeasing HEC-critics. What seems to be motivating Butler (1998) and Rupert's (2004, 2010) concern is that by including external resources in the architecture of a cognitive system, we are relinquishing agential control to extra-organismic artefacts, and this in turn detracts attention from the tried-and-tested target of cognitive science: the brain. I have already mentioned that I do not believe issues of agency supervene on issues of cognition (section 3.2.1), so I will focus on the notion of 'cognitive control', which Butler (1998) and Rupert (2004, 2010) seem to treat as analogous to 'agential control'. Through lacking agency, the inanimate objects involved in standard HEC cases lack cognitive control, i.e. they are not autonomously cognitive and make no independent or integrative contributions to a cognitive system. HEC-systems thus have an apparent core of cognitive control - an individual's brain (or specific neural circuits within it, depending on the process) - and the external resources are just parasitically exploited. In an ICS, however, cognitive control can be *distributed* across the individuals involved (Tollefsen, 2006). Each individual (constituting subsystemic components) *has* cognitive control, and the active cognitive processing is reciprocally exploited in such a way that no one element of the system is being parasitically manipulated. The *overall ICS* has control for a specific process, in the way that the active neural elements of an individual brain have control for a specific process. For example, when Olaf and Inga mutually recall the address of their friend, it is very much the ICS, rather than Olaf or Inga individually, that has control. An ICS is after all to be considered an *interactor,* in the sense of being a "cohesive whole" or "adaptive unit", and as such should not be divided according to any physical or agential boundaries. If there is to be such a thing as a persisting locus of cognition, therefore, it is distributed across an ICS, instead of being 'lost' to inert external props.

The ICS approach has the added bonus of retaining the current taxonomy of cognitive states and processes, i.e. the neural circuits and mechanisms of the brain. This is once more due to the homogeneity of the cognitive realisers at work in the ICS. By retaining the current psychological taxonomy, ICS manages to accommodate cognitive extension and provide "a coherent and fruitful framework within which to place all, or at least a healthy majority of, significant results in cognitive science" (Rupert, 2004, p. 407). The framework that HEC proposes more generally, on the other hand, requires a more expansive taxonomy that will include a large proportion of

non-neural environmental features.

Naturally, ICS (in virtue of being a kind of HEC) *is* ultimately seeking a more expansive taxonomy, but by largely conserving the precious subject (the brain) of current cognitive disciplines, it skilfully bridges the gap between inflexible internalism and radical externalism.

5. Conclusion

In this paper I have provided a defence of extended cognition by appealing to the notion of cognition reciprocally extending amongst individuals, forming an ICS. Unlike the paradigmatic case of Otto and his notebook, an ICS approach will withstand many of the criticisms levelled at HEC. I have discussed three of these here: the distinctive nature of internal cognition (Adams & Aizawa, 2001, 2008), the problem of cognitive integration (Weiskopf, 2008; Spaulding, 2011), and the need for a 'persisting locus of cognition' (Butler, 1998; Rupert, 2004, 2010). These criticisms are interrelated in that they all suppose that internal cognition has, or should have, a special place within cognitive scientific research. For the most part this is not denied. The worry is that by focusing so intently on the machinery and mechanisms of internal cognition, we are in fact turning a blind eye to much that should be of scientific and philosophical interest. What an ICS approach achieves is a robust affirmation of the possibility of cognition extending, alongside the preservation of the psychological taxonomy that has proved so intriguing and provocative thus far. It should thus have an appeal for internalists, embodied cognition proponents and HEC proponents (alternatively, it will appeal to none of them!).

I believe the difficulty with accepting ICS, or HEC more generally, stems from our folk-intuitions about the 'self' and 'mind', which are inextricably connected. The 'self' is considered a physically-bound and intelligent being; this intelligence comes from the 'mind', which commonsense tells us maps directly onto the brain. Mind and self thus intuitively entail one another. Even being aware of the dense interactive connections we form with others and the environment around us, intuition instils a belief in the importance of the physical boundary that our skin provides. This feeling is reinforced by years of literature and everyday parlance:

> We live together, we act on, and react to, one another; but always and in all circumstances we are by ourselves. The martyrs go hand in hand into the arena; they are crucified alone[…]Sensations, feelings, insights, fancies - all these are private and, except through symbols and at second hand, incommunicable. We can pool information about experiences, but never the experiences themselves. From family to nation, every human group is a society of island universes.
> (Huxley, 1954, p. 13)

Such an intuition is not necessarily ungrounded. In the case of everyday communication, for example, there is an obvious benefit in drawing a distinction between 'I' and 'we'. However, from a scientific or philosophical perspective, it may be that we need to shake off the shackles of this intuition in order to embrace a more expansive study of 'mind', one that we can detach from the physically-bound individual. We must accept that "the self or cognizing subject is fundamentally fragmented, divided, or nonunified" (Varela, Thompson & Rosch, 1991) and cognitive systems can manifest themselves at levels above and below that of an individual organism (e.g. Theiner et al., 2010; Kennedy & Eberhart, 2001). 'Minds', quite simply, are shifting coalitions of mechanisms that can manifest themselves across various levels of investigation, and across various physical realisers, to guide the thoughts and actions of individuals, groups and societal structures.

An ICS is just one such example of this kind of mind, but it may be the first that is capable of drawing internalists and HEC-critics from their stubborn stance on cognition. It is undoubtedly a difficult task to instil in academic circles a view of the mind as a multi-levelled and fluctuating phenomenon, but if we can successfully "let a thousand flowers bloom" (Clark, 2007, p. 170), then I believe the blossom will be worth it.

References

- F. Adams & K. Aizawa, 2001, 'The Bounds of Cognition', *Philosophical* Psychology, 14 (1), pp. 43-64
- F. Adams & K. Aizawa, 2008, *The Bounds of Cognition*, Wiley-Blackwell
- F. Ashby, V, Isen & U. Turken, 1999, 'The effects of positive affect and arousal on working memory and executive attention', in *Emotional Cognition: from brain to behaviour* (eds. Moore & Oaksford), John Benjamins, pp. 245-287
- J. Atkinson & T. Huston, 1984, 'Sex role, orientation and division of labor early in marriage', *Journal of Personality and Social Psychology*, 46, pp. 330-45
- E. Bonabeau, M. Dorigo and G. Theraulaz, 1999, *Swarm Intelligence: From Natural to Artificial Systems*, Oxford University Press
- T. Bosse, C. Jonker, M. Schut & J. Treuer, 2006, 'Collective representational content for shared extended mind', *Cognitive Systems Research Journal*, 7 (2), pp. 151-174
- M. Bratman, 1993, 'Shared Intention', *Ethics*, 104, pp. 97-113
- T. Burge, 1979, 'Individualism and the mental', *Midwest Studies in Philosophy*, 4, pp. 73-121
- K. Burke, 1989, *On Symbols and Society*, University of Chicago Press
- K. Butler, 1998, *Internal Affairs: making room for psychosemantic internalism*, Kluwer
- P. Carruthers, 2006, *The Architecture of the Mind: Massive Modularity and the Flexibility of* Thought, Oxford University Press
- D. Chalmers, 2008, 'Foreword', *Supersizing the Mind* (Clark, 2008), Oxford University Press
- A. Clark & D. Chalmers, 1998, 'The Extended Mind', *Analysis*, 58 (1), pp. 7-19
- A. Clark, 1989, *Microcognition*, MIT Press
- A. Clark, 2001, 'Reasons, robots and the extended mind', *Mind and Language*, 16 (2), pp. 121-145
- A. Clark, 2004, '"Author's reply" to symposium on natural-born cyborgs', *Metascience*, 13 (2), pp. 57-69
- A. Clark, 2007, 'Curing cognitive hiccups: a defence of the extended mind', *The Journal of Philosophy*, 104 (4), pp. 163-192
- A. Clark, 2008a, 'Pressing the flesh: a tension in the study of the embodied, embedded mind?', *Philosophy and Phenomenological Research*, 76 (1), pp. 37-59
- A. Clark, 2008b, *Supersizing the Mind: embodiment, action and cognitive extension*, Oxford University Press
- A. Clark, 2010a, '*Memento's* revenge: the extended mind, extended', in *The Extended Mind* (Menary), MIT

Press, pp. 43-66

- A. Clark, 2010b, 'Coupling, constitution and the cognitive kind: a reply to Adams and Aizawa', in *The Extended Mind* (Menary), MIT Press, pp. 81-100
- A. Clark, 2010c, 'Much ado about cognition', *Mind*, 119 (476), pp. 1047-1066
- A. Crisafi & S. Gallagher, 2009, 'Hegel and the extended mind', *AI and Society*, 25 (1), pp. 123-129
- T. Dartnall, 2004, 'We have always been…cyborgs', *Metascience*, 13 (2), pp. 139-273
- D. Davidson, 1987, 'Knowing one's own mind', *Proceedings and Addresses of the American Philosophical Association*, 60 (3), pp. 441-458
- G. Della Barba, 1993, 'Confabulation: knowledge and collective experience', *Cognitive Neuropsychology*, 10, pp. 1-20
- G. Domes, M. Heinrichs, A. Michel, C. Berger & S. Herpertz, 2007, 'Oxytocin improves "mind-reading" in humans', *Biological Psychiatry*, 61, pp. 731-733
- D. Dennett, 1997, *Kinds of Minds: towards an understanding of consciousness*, Basic Books
- D. Dennett, 1998, *Brainchildren: essays on designing minds (representation and minds),* MIT Press
- V. Descombes, 2001, *The Mind's Provisions: a critique of cognitivism,* Princeton University Press
- J. Fodor, 1983, *The Modularity of Mind*, MIT Press
- J. Fodor, 2000, *The Mind Doesn't Work That Way*, MIT Press
- M. Gazzaniga, R. Ivry & G. Mangun, 2002, *Cognitive Neuroscience: the Biology of the Mind*, W. W. Norton & Co
- M. Gilbert, 2009, 'Shared Intention and Personal Intentions', *Philosophical Studies*, 144 (1), pp. 167-187
- M. Halbwachs, 1938, *Morphologie Sociale*, Librairie Armand Colin
- D. Hull, 1980, 'Individuality and selection', *Annual Review of Ecology and Systematics*, 11, pp. 311-332
- S. Hurley, 2010, 'The varieties of externalism', in *The Extended Mind* (Menary), MIT Press, pp. 101-154
- D. Hutto, 1999, 'Cognition without representation?', in *Understanding Representation* (Reigler & Markus), Plenum Press
- A. Huxley, 1954, *The Doors of Perception*, Penguin
- L. Jacoby, 1978, 'On interpreting the effects of repetition: solving a problem versus remembering a solution', *Journal of Verbal Learning and Verbal Behaviour*, 17 (6), pp. 649-668

- I. Janis, 1983, *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*, Houghton Mifflin Harcourt
- C. Jung, 1922, *Collected Papers on Analytical Psychology*, Balliere, Tindall & Cox
- J. Kennedy & R. Eberhart, 2001, *Swarm Intelligence*, Morgan Kaufmann
- D. Kirsch, 1995, 'Complementary strategies: why we use our hands when we think', in *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, (eds. Moore & Lehamn), pp. 212-217
- M. Kosfield, M. Heinrichs, P. Zak, U. Fischbacher & E. Fehr, 2005, 'Oxytocin increase trust in humans', *Nature*, 435 (7042), pp. 673-676
- G. Le Bon, 1895 (1896), *The Crowd: a Study of the Popular Mind*, Transaction Publishers
- D. Liang, R. Moreland & L. Argote, 1995, 'Group versus individual training and group performance: the mediating role of transitive memory', *Personality and Social Psychology Bulletin*, 21 (4), 384-393
- J. Lau & M. Deutsch, 2010, 'Externalism about mental content', [online] http://plato.stanford.edu/entries/content-externalism/, retrieved 27/7/12
- R. Menary, 2006, 'Attacking the bounds of cognition', *Philosophical Psychology*, 19 (3), pp. 329-344
- R. Menary, 2010, *The Extended Mind*, MITPress
- J. McClelland, D. Rinehart & the PDP Research Group, 1986, *Parallel Distributed Processing: explorations in the microstructure of cognition*, MIT Press
- U. Nessier & R. Fivush, 1994, *The Remembering Self: construction and accuracy in the self narrative*, Cambridge University Press
- D. Norman, 1991, 'Cognitive artifacts', in *Designing Interaction: psychology at the human-computer interface* (ed. Carroll), Cambridge University Press, pp. 17-38
- V. Pareto, 1935, 'Trattato di sociologia generale' (1916), in *The Mind and Society* (ed. Livingston), Brace Harcourt
- M. Patestas & L. Gartner, 2006, *A Textbook of Neuroanatomy*, Wiley-Blackwell
- H. Putnam, 1975, 'The meaning of meaning', *Philosophical Papers, Vol II: Mind, Language and Reality*, Cambridge University Press
- R. Rupert, 2004, 'Challenges to the hypothesis of extended cognition', *Journal of Philosophy,* 101 (8), pp. 389-428
- R. Rupert, 2010, 'Extended cognition and the priority of cognitive systems', *Cognitive Systems Research*, 11 (4), pp. 343-356
- H. Smit & P. Rogers, 2000, 'Effects of low doses of caffeine on cognitive performance, mood and thirst in low and higher caffeine consumers', *Psychopharmacology*, 152 (2), pp. 167-173

- S. Spaulding, 2011, 'Overextended cognition', *Philosophical Psychology*, 25 (4), pp. 469-490
- E. Sober, 1993, *The Nature of Selection*, University of Chicago Press
- E. Sober & D. Wilson, 1998, *Unto Others: the Evolution and Psychology of Unselfish Behaviour*, Harvard University Press
- E. Smith, 2007, 'Social relationships and groups: new insights on embodied and distributed cognition', *Cognitive Systems Research*, 9, pp. 24-32
- M. Sprevak, 2009, 'Extended cognition and functionalism', *The Journal of Philosophy*, 106, pp. 503-527
- M. Sprevak, 2010, 'Inference to the hypothesis of extended cognition', *Studies in History and Philosophy of Science*, 41, pp. 260-270
- K. Sterelny, 2010, 'Minds: extended or scaffolded?', *Phenomenology and the Cognitive Sciences*, 9 (4), pp. 465-481
- G. Theiner, 2008, 'Making sense of group cognition: the curious case of transactive memory systems', in *ASC09: Proceedings of the 9th Conference of the Australian Society for Cognitive Science* (eds. Christensen, Schier & Sutton), Macquarie Centre for Cognitive Science
- G. Theiner, C. Allen & R. Goldstone, 2010, 'Recognizing Group Cognition', *Cognitive Systems Research*, 11, pp. 378-395
- D. Tollefsen, 2006, 'From extended mind to collective mind', *Cognitive Systems Research*, 7, pp. 140-150
- M. Tomasello & M. Carpenter, 2007, 'Shared intentionality', *Developmental Science*, 10 (1), pp. 121-125
- R. Trivers, 2011, *The Folly of Fools: the Logic of Deceit and Self-decpetion in Human Life*, Basic Books
- F. Varela, E. Thompson & E. Rosch, 1991, *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press
- J. Velleman, 1997, 'How to share an intention', *Philosophy and Phenomenological Research*, 57 (1), pp. 29-50
- D. Wegner, T. Giuliano & P. Hertel, 1985, 'Cognitive interdependence in close relationships', in *Compatible and Incompatible Relationships* (ed. Ickes), Springer-Verlag, pp. 253-276
- D. Wegner, R. Erber & P. Raymond, 1991, 'Tranasctive memory in close relationships', *Journal of Personality and Social Psychology*, 61 (6), pp. 923-929
- D. Wegner, 1986, 'Transactive memory: a contemporary analysis of the group mind', in *Theories of Group Behavior* (eds. Mullen & Goethals), pp. 185-208
- D. Wegner, 1995, 'A computer network model of human transactive memory' *Social Cognition*, 13, pp. 319-339
- D. Weiskopf, 2008, 'Patrolling the mind's boundaries', *Erkenntnis*, 68 (2), pp. 265-276

- M. Wheeler, 2010, 'In defence of extended functionalism', in *The Extended Mind* (Menary), MIT Press, pp. 245-270

- R. Wilson & A. Clark, 2009, 'How to situate cognition: letting nature take its course', in *The Cambridge Handbook of Situated Cognition* (eds. Aydede & Robbins), pp. 55-77

- R. Wilson, 2004, *The Boundaries of the Mind: the individual in the fragile sciences*, Cambridge University Press