Does the mind leak?

On Andy Clark's extended cognition hypothesis and its critics

---

A thesis
submitted in fulfilment
of the requirements for the Degree
of
Master of Arts in Philosophy
by

Uwe Peters

University of Canterbury
2009

Table of contents

Abstract

A growing controversy at the interface of philosophy and cognitive science concerns the question of where cognition is located. In the paper "The Extended Mind" (1998), the book *Supersizing the Mind* (2008) and many other publications, Andy Clark contends that cognitive processes do not only occur in the head but also physically extend into the environment. In contrast and critical response to Clark, Adams/Aizawa (2008) and Rupert (forthcoming) hold that cognition is an entirely brain-bound affair. In the present thesis, I will argue that Clark's extended cognition hypothesis as well as Adams/Aizawa's and Rupert's brain-bound accounts should be rejected because they lack plausibility and are cognitive-scientifically gratuitous. However, even though I dismiss Adams/Aizawa's and Rupert's specific brain-bound views, I will reach a conclusion similar to theirs: contra Clark, cognition remains an internal phenomenon.

Introduction

Is the blood the element with which we think, or the air or the fire?
Or perhaps nothing of the kind – but the brain may be the originating power of the perceptions of hearing and sight and smell and memory and opinion may come from them.

(Plato, *Phaedo*, p. 481)

On January 7, 2008, the *New York Times* reported the following incident. About three years ago Michael T. Arnold arrived at the Los Angeles International Airport on a plane from the Philippines. He had a laptop with him. At the customs, an officer took a look at what was on Arnold's computer hard drive, clicked on a number of folders and found child pornography. Consequently, Arnold faced a criminal charge.

In the course of the subsequent trial, the government held that searching Arnold's computer hard drive at the customs was as legitimate as searching his suitcase. However, one judge, Dean D. Pregerson of the Federal District Court in Los Angeles, objected and did not acknowledge the evidence against Arnold:

> "'Electronic storage devices function as an extension of our own memory,' Judge Pregerson wrote, in explaining why the government should not be allowed to inspect them without cause. 'They are capable of storing our thoughts, ranging from the most whimsical to the most profound.'"[1]

Can external resources, such as computer hard drives, really function as literal, physical extensions of someone's mind? Where does the mind stop and the external world begin? These questions played a crucial role in the aforementioned court case. But they have also

[1] Liptak (2008)

provoked a growing controversy among philosophers, cognitive scientists and psychologists.

At first sight, the answer to the question of where the mind or cognition is localised appears quite obvious. As Kosslyn (1986) remarks, "The fact that cognition is something the brain does is so obvious it seems barely worth stating" (p. 231). Cognition – which comprises all forms of knowing and awareness, such as perceiving, conceiving, remembering, reasoning, judging, imagining, and problem solving[2] – is in the skull. Let us call this the brain-bound view.

Even though this view seems obvious indeed, some disagree. Andy Clark argues, and many other philosophers concur or express similar ideas,[3] that cognition sometimes physically extends into non-neural objects of the environment.

To qualify, Clark does not claim that, for instance, one's thoughts or experiences are environmental objects which can somehow literally be touched, seen or, if small enough, carried around in one's pocket. His point is rather that "the *physical machinery* [emphasis added] that realizes some of our cognitive processes and mental states" extends "beyond the bounds of skin and skull" (2008, p. 76). This is Clark's hypothesis of extended cognition (in the following 'HEC').[4] HEC and two of its most recent critiques – one by the co-author duo Frederick Adams and Kenneth Aizawa[5] and the other by Robert Rupert – are at the centre of the present thesis.[6]

[2] See VandenBos (2007), p. 187
[3] See, for instance, Millikan (1993), Dennett (1996), Haugeland (1998), Wilson (2004) or Noë (2007)
[4] The defence and elaboration of this hypothesis pervade Clark's publications on cognition. See, for instance, Clark (1997), (2001), (2007) or (2008)
[5] In the following, I will abbreviate the names of Frederick Adams and Kenneth Aizawa to 'Adams/ Aizawa.'
[6] Other philosophers who are critical of HEC include Segal (1997), Sterelny (2004), Block (2005), Horgan and Kriegel (2008).

But why deal with Clark's HEC and its critics? What does it matter whether cognition is brain-bound or extended? With respect to these questions, Shapiro (2008), who is sympathetic towards Clark's view, holds that much is at stake in the extended cognition debate. He mentions a number of reasons, including the following.

If Clark's extended cognition view is right, Shapiro contends, then "there is much about psychology that is wrong. Cognitive neuroscience, for instance, would have been grounded in the false belief that all cognitive processes emerge somehow from neural processes […]" (2008, p. 2). Furthermore, computational psychologists would have to take into account environmental features to fully understand the implementation of the algorithmic processes that characterise cognition. Also,

> "Studies of psychopathologies could not limit themselves to an investigation of brain disorders. […]
>
> Perhaps more seriously, if minds are extended then our ordinary ways of describing and thinking about human beings must undergo dramatic revision. We might have to learn to make sense of claims like 'Welch accidentally left his memory on the bus,' or 'Dixon stubbed his mind on his way to work this morning.'" (ibid)

In a very similar vein, Rupert (2004), who is sceptical about HEC, writes: "Acceptance of HEC would […] significantly change our conception of persons" (p. 389).

Now, if Clark's HEC has such supposedly 'significant' and 'serious' consequences, it becomes interesting to see whether the view is tenable. In the following thesis, I will, on the one hand, investigate whether Clark's extended cognition view is tenable and, on the other hand, scrutinise Adams/Aizawa's and Rupert's cases for the opposite, seemingly

obvious brain-bound view. As I will argue, both positions on the location of cognition – Clark's extended view and Adams/Aizawa's and Rupert's brain-bound view – are problematic for the following three reasons.

• First, Clark's extended cognition view entails highly counterintuitive instances of cognitive extension.

• Second, the phenomena that inspire Clark's HEC can without explanatory loss be accounted for in terms of a more conservative and ontologically more parsimonious hypothesis obviating any cognitive-scientific need to embrace HEC.

• Third, with respect to Adams/Aizawa's and Rupert's brain-bound views, their proposals on the location of cognition are unacceptably restrictive.

With these three points in mind, I will contend that Clark's extended cognition view as well as Adams/Aizawa's and Rupert's brain-bound accounts should be rejected since they lack plausibility and are cognitive-scientifically gratuitous. Furthermore, I will argue that cognition occurs, contra Clark, only in the head. However, it is not located in the head because Adams/Aizawa's or Rupert's arguments for brain-bound cognition are persuasive but rather because Clark has so far failed to provide sufficient reasons for believing otherwise.

Before starting the discussion, a few words pertaining to the structure of the following thesis will be useful. I will begin by briefly introducing motivations for the brain-bound view before illustrating how fundamental this view is in cognitive science. After that, I will introduce Clark's arguments for HEC (chapter II) and Adams/Aizawa's critique (chapter

III) of it. Then, in chapter IV, I will examine Adams/Aizawa's critique before proposing my own objections to Clark's extended cognition view (chapter V). Having dealt with Clark's view on the location of cognition throughout the chapters II to V, I will in the chapters VI-VIII introduce and take issue with Adams/Aizawa's and Rupert's opposing brain-bound view. Finally, in chapter IX, I will summarise the results of my discussion of the two views on the location of cognition and draw a conclusion.

Chapter I: Why is cognition in the head anyway? – Background

Let us commence by asking a very simple question: why think that cognition is localised in the head? The seemingly natural, brain-bound view receives strong support from cognitive neuroscience and traditional cognitive science. Cognitive neuroscientists hold that, "[d]ifferent parts of the brain are specialized for different functions" (Kandel et al. 1995, p. 8). By studying the way lesions, strokes, injuries or surgery of the brain affect an agent's cognitive abilities and by drawing on PET and fMRI[7] brain-scanning procedures, cognitive neuroscientists can localise cognitive functions in specific brain areas.[8]

For instance, Broca's area and Wernicke's area in the left hemisphere of the brain have been identified as critical for language. In the 19th century, the French neurosurgeon Paul Broca and the German neurologist Carl Wernicke discovered that particular areas of the left frontal and temporal lobes were responsible for the production and comprehension of language. They made their discoveries after they had autopsied and studied the brains of individuals with speech impairments. Broca, for instance, found that all the individuals he autopsied had brain lesions in a particular area of the left inferior frontal lobe. From the neurophysiological damage to this area and the individual's corresponding inability to produce complex linguistic structures, he was led to proclaim: "Nous parlons avec

[7] 'PET' stands for Positron Emission Tomography and 'fMRI' for functional Magnetic Resonance Imaging, see Thatcher et al. (1994) for an interesting and comprehensive introduction to both methods of measuring neuronal activity.

[8] The neuroscientific project of locating particular cognitive functions in sharply demarcated regions of the brain is not uncontroversial. It rests on the assumption that cognitive functions are modular in nature and realised by isolated brain areas. However, many cognitive functions may not be modular (see, for instance, Sterelny, 2003) and brain regions are not isolated; rather they "work together as fully interconnected, distributed neural networks" (Howieson et al, 2004, p. 287). Furthermore, even if lesion studies should reveal that a specific brain region is necessary for a particular cognitive function, this does not prove that this specific brain region is also sufficient for the cognitive function. In a highly interconnected brain, it is unlikely that an individual area which is, for instance, necessary for language is also sufficient for this very cognitive function (see Uttal, 2001, pp. 163).

l'hemisphere gauche"[9] (trans. 'We speak with the left hemisphere [of the brain]').[10] So, neuroscience holds that "[l]anguage and other cognitive functions are localized within the cerebral cortex," that is, inside the head of an individual (Kandel et al, 1995, p. 9).

Traditionally, cognitive science also embraces the assumption that mental phenomena are internal phenomena. The reason for this assumption is intimately linked to the history of cognitive science and its emergence as a critical response to behaviourism. Behaviourism represents the psychological theory that dominated scientific psychology until the 1960s (Grush, 2001, pp. 2). It eschews inner, only subjectively accessible mental states and proposes that a mental state is nothing but a pattern of sensory stimuli and publicly observable behavioural responses which an individual exhibits, or is disposed to exhibit in a particular situation. For instance, the behaviourist contends that if we say that Jack believes it will rain, then we do not mean that there is some mysterious inner mental process of believing which accompanies Jack's actions. We only mean that Jack's actual or potential behaviour follows a particular stimulus-response pattern. This pattern may be that when we ask him whether it will rain, he will respond with 'Yes'; when he sees the raincoat by the door, he will put it on before leaving the house; when offered an umbrella, he will take it with him, and so on. According to behaviourism, when we use mental state terms such as 'belief,' 'desire,' 'intention' etc., we do not pick out inner entities but behavioural dispositions such as the ones exhibited by Jack. As an analogy consider an object's chemical disposition for solubility: "To say that *X* is soluble is not to say that *X* contains some hidden spirit of solubility. It is just to say that if you put *X* in water, *X* would dissolve." (Clark, 2001, p. 166) Similarly, if Jack has a belief about the weather, there is nothing ghostly in his head; he is merely disposed to act in a certain ways.

[9] Kandel et al. (2000), p. 10
[10] Gazzaniga's fascinating split-brain experiments give also support to this claim. See Gazzaniga (2004), pp. 1189

However, the radical rejection of inner entities and the exclusive focus on behavioural disposition to understand mentality turned out to be problematic and precipitated in the 1960s a "cognitive revolution" (Miller, 2003, p. 141).

> "The cognitive revolution in essence was the realization that any adequate theory of human and animal mentality would need to posit representational states between sensory stimulus and behavioral response – at least for a great many domains of behavior. These states would be theoretical, and not simply reducible to constructs of observables." (Grush, 2001, p. 3)

The assumption that an adequate account of human mentality requires the postulation of internal representations was primarily motivated by new psychological findings on human behaviour. For instance, during the Second World War psychologists began to realise that much skilled human behaviour is simply too complex to be explained in terms of behaviouristic stimulus-response mechanisms (Hampson and Morris, 1996, p. 7). They found that intelligent behaviour such as, for example, flying a plane requires one to be able to anticipate possible problems and plan accordingly.[11] However, both concepts – anticipation and planning – were not available in the behaviouristic framework. The reason is that anticipation and planning crucially involve what behaviourism eschewed: internal, mental representations of possible incidents.

The view that problem-solving behaviour involves internal representations between stimulus and response was further substantiated by experiments on animals. For instance, the psychologist Edward C. Tolman conducted maze navigation experiments with rats and

[11] That human cognition involves mental representations and that it can occur decoupled from any immediate and possible interaction with the environment becomes also evident by our abilities to, for instance, imagine fictional characters based solely on linguistic input from others and to think about counterfactual situations to reflect on how events could have turned out if things had happened differently. Imagining fictional characters or thinking about counterfactual situations depends crucially upon mental representations and takes place decoupled from any immediate or even possible interaction with the respective object of thought (see Wilson, 2002, p. 626).

found that the behaviour of the rats could not be explained in terms of stimulus-response models. Rather, to navigate through the maze and find food, rats seemed to draw upon internal representations (or "cognitive maps") of the location of the food (Kolak, 2006, p. 20).[12] The seeming inevitability in postulating internal entities in order to adequately account for animal and human problem-solving performances was a major catalyst for the demise of behaviourism[13] and the emergence of cognitive science.[14]

The assumption of inner mental entities between sensory stimulus and behavioural response became fundamental to cognitive science (Green et al. 1996, p. 7-8). It gave rise to a picture of cognition that has been described as the "sandwich" model: cognition is seen as "wedged between perception (on the input side) and action (on the output side)" (Clark/Wilson, 2009, p. 56). According to the 'sandwich' model, cognition can roughly be described as follows. An organism receives sensory input via its peripheral nervous system. The sensory input is transformed into abstract symbols and sent to the central processor, the brain. The brain "manipulates these symbols together with symbols from stored information data structures, and forms a plan or settles on some solution to a problem." (Grush, 2001, p. 11-12) Once the plan is made, it sends symbols to "output transducers, which control effectors so as to produce some sort of movement or other effect on the body or environment." (ibid) According to traditional cognitive science, cognition consists of the

[12] For more recent evidence that animals remember where they have stored food and may draw upon mental maps to localise it, see the study of Normand et al. (2009) on chimpanzees.

[13] Alongside new psychological findings, there were other important factors that led to the decline of behaviourism and rise of cognitive science. See Grush (2001, pp. 3), Hampson and Morris (1996, pp. 7) or Kolak et al. (2006, pp. 20) for accounts of how, for instance, Noam Chomsky's linguistic arguments and research in artificial intelligence influenced the transition from behaviourism to cognitive science. For philosophical arguments against behaviourism see, for instance, Braddon-Mittchell/Jackson (2007, pp. 41-45)

[14] The terms 'cognition' and 'cognitive science' were introduced by the principal instigators of the cognitive revolution to demarcate their psychological research projects from behaviourism. George A. Miller, who was one of the pivotal figures in the cognitive revolution (see Miller, 2003), said in an interview: "I think they [early cognitive scientists] were just reaching back for common sense. In using the word 'cognition' we were setting ourselves off from behaviourism. We wanted something *mental* – but 'mental psychology' seemed terribly redundant. 'Commonsense psychology' would have suggested some sort of anthropological investigation, and 'folk psychology' would have suggested Wundt's social psychology. What word do you use to label this set of views? We chose 'cognitive.'" Miller in Baars (1986, p. 210)

mental manipulation of symbols inside the just mentioned input-output 'bun' in the brain. Consequently, to understand cognition, the cognitive scientist focuses, unlike the behaviourist, primarily on what goes on internally. The concentration on internal symbol or representation processing led to the adoption of what Jerry Fodor called a "methodological solipsism:" to explain cognition, cognitive science "bracketed off the world beyond the individual" (Clark/Wilson, 2009, p. 56).

Clark's extended cognition view radically breaks with this methodological solipsism and the traditional internalist conception of cognition in cognitive science and cognitive neuroscience.

Clark summarises his hypothesis of extended cognition (HEC) vividly thus:

> "[T]hinking and cognizing […] may (at times) depend directly and non-instrumentally upon the ongoing work of the body and/or the extra-organismic environment. […] [T]he actual local operations that realize certain forms of human cognizing include inextricable tangles of feedback, feed-forward and feed-around loops: loops that promiscuously criss-cross the boundaries of brain, body and world. The local mechanisms of mind, if this is correct, are not all in the head. Cognition leaks out into body and world." (2008, p. xxviii)

Clark claims that with respect to cognitive processing, there is nothing special about the boundary of skin and skull. In some cases, the brain and some environmental features coalesce into a single system that physically instantiates cognitive phenomena. As a result, the traditional cognitive-scientific 'bracketing off the world beyond the individual' cannot yield an adequate understanding of cognition. But why should one adopt HEC? What are Clark's arguments in support of HEC?

Chapter II: Loosening the skullcap – An exposition of Clark's HEC

Having introduced the background against which HEC develops, I will in the present chapter explain Clark's case for HEC. Clark first formulated the main arguments for his hypothesis in the 1998 paper 'The Extended Mind.'[15] The article offers two thought experiments with the first intended to show that *cognitive processes* may sometimes extend into environmental objects and external processes. The second one is devised to illustrate how *mental states* could be seen as extended into the world. I will explain Clark's thought experiments and arguments for extended cognitive process and mental states in turn.

II. 1. Extended cognitive processes

To make his case for the claim that "cognitive processes ain't (all) in the head" (1998, p. 8), Clark asks us to consider the following.

Scenario (1): Imagine someone sitting in front of a computer screen which shows images of different geometric shapes. The person is asked to decide whether the shapes on the screen will fit into depicted sockets. To judge whether they fit, he or she must rotate the shapes in their mind to align them with the matching sockets.

Scenario (2): Next, imagine another person sitting in front of a similar screen. But this time, to align the shapes with the sockets, the person can physically rotate them on the screen by pressing a rotate button.

[15] The paper was co-written by Clark and David Chalmers. However, since Chalmers takes a cautious distance from HEC (see footnote 1998, p. 7) and writes that the "inspiration" behind the article "was all" Clark's (Chalmers, 2008, p. x), I will consider Clark in the following as the main proponent of HEC and only refer to the co-written paper by 'Clark (1998).'

Scenario (3): Finally, imagine someone sitting in front of that very computer screen at some point in the future. Unlike the subjects in the preceding scenarios, this person has not only mental rotation capacities but also a retinal display and a neural implant by which he or she can quickly rotate the images of the geometric shapes on the display on demand. To rotate the image of the shapes, the person sends a thought command from his or her motor cortex to the neural implant which then initiates the rotation and finally provides an answer to whether or not a particular shape on the computer screen fits into a depicted socket.

With these three scenarios in mind, Clark argues that the process of manual rotation in (2) should count as a cognitive process. His reasoning goes as follows. First, (1) is an indisputably cognitive process. Second, (1) and (3) appear to be cognitively on a par: (1) and (3) involve equivalent internal operations. Third, by hypothesis, (2) functions in turn just as (3), (2) is merely an externally instantiated scenario (3). Clark concludes, since (1) and (3) are cognitive processes and scenario (2) is functionally equivalent to (3), the manual rotation of shapes in scenario (2) should count, in the same way as (1) and (3), as a cognitive process. This is Clark's argument. There is however one crucial assumption underlying the argument which yet has to be made explicit: Clark assumes that any process is cognitive if it is functionally equivalent to a cognitive process. Let us, for clarity's sake, put Clark's argument in a more concise form.

1. Mentally rotating shapes is a cognitive process.
2. Any process that is functionally equivalent to a cognitive process is itself a cognitive process.
3. Physically or manually rotating geometric shapes is functionally equivalent to a cognitive process.

4. Thus, from the premises 1, 2 and 3, manually rotating geometric shapes is a cognitive process.
5. Manually rotating geometric shapes is an external, environmental process.
6. Thus, from the premises 4 and 5, "cognitive processes ain't (all) in the head" (1998, p. 8).[16]

Clark holds that there are more common cases of extended cognition than the just mentioned manual rotation. To further elucidate Clark's argument for extended cognitive processes, let us consider another, more common example: the solving of long multiplications by using a pen and a piece of paper (1998, p. 8). Clark and Wilson (2009, p. 60) elaborate on this example. They claim that on the occasions when I, for instance, use pen and paper to solve mathematical problems such as multiplying 386 x 417, my cognitive process of doing arithmetic literally extends into the pen and paper.

Clark and Wilson provide the following reasons for their claim. If I did the calculations required to complete the multiplication without pen and paper, I would have to internally represent the figures and carry all the partial results of the calculation in my head. Doing so constitutes a cognitive task. By writing the intermediate results down, this cognitive task is off-loaded onto the piece of paper. The numbers that would have to be internally represented and stored are externally represented and stored (see Figures 1 and 2 below).

[16] Clark considers and rebuts a number of objections to his argument in his 1998 original paper. I will not repeat them here, see Clark (1998).

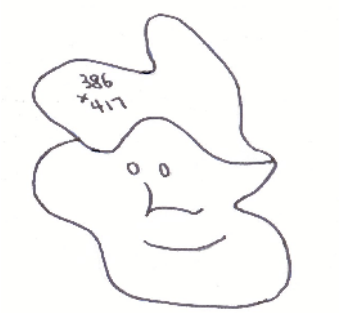

Fig. 1

Multiplying with internal symbols external alone

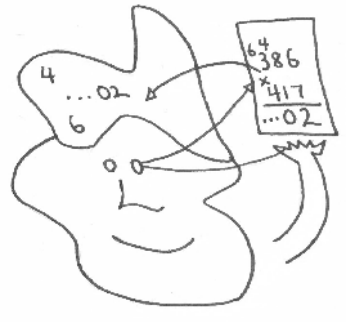

Fig. 2

Multiplying with internal and integrated symbols and resources[17]

Since the pen and the paper support the off-loading and returning of information required for solving the calculation, they fulfil the same cognitive functions that short-term memory would, in the absence of pen and paper, have to perform. Because pen and paper take over the function that a cognitive resource (i.e. short-term memory) would otherwise have to perform, Clark and Wilson hold that my cognitive processing extends into these two external objects.

For the time of the calculation, I, together with pen and paper constitute, what Clark and Wilson call, a

> "transient extended cognitive system (TECS). A TECS is a soft-assembled whole that meshes the problem-solving contributions of the human brain and the nervous system with those of the (rest of the) body and various elements in the local cognitive scaffold [such as pen and paper which support cognitive processing, UP]." (Clark/Wilson, 2009, p. 65)

[17] The illustrations were taken from Clark/Wilson (2009, p. 61).

Clark and Wilson hold that TECS[18] can be distinguished according to "the durability and reliability of the relationship between agent and resource. Thus we might want to distinguish temporary, one-off relationships from those that though transient, are regularly repeated." (ibid) For instance, to

> "solve a new brainteaser, an agent may generate a brand-new, one-off kind of TECS. While a practiced crossword puzzle solver, though perhaps confronted with a new puzzle (and as usual armed with pen and paper) will rapidly generate a well-understood, often repeated, form of TECS." (ibid)

So TECS may vary with respect to their persistence. But Clark and Wilson seem to suggest that no matter how frequently an agent uses pen and paper to do long multiplications and how often these resources are available to her, once they are available and she employs them in performing arithmetic, while they are being used, they automatically belong to her cognitive system. As a result, Clark and Wilson claim that extended cognition is not only an exception to the rule of mostly non-extended cognition but rather that TECS are pervasive: They constitute "quite literally, the soft-assembled circuitry of a great deal of practical human thought and reason." (Clark/Wilson, 2009, p. 73)

If performing arithmetic on paper qualifies as a cognitive process and pen and paper can constitute proper parts of a cognitive system, what other external processes and resources can qualify as cognitive? To determine whether a given external process, such as calculating on paper or manually rotating geometric shapes, counts as an extended cognitive process, and to find out whether a given interaction or coupling between an agent and some external resource constitutes a TECS, Clark proposes the "parity principle" (2008, p. 108).

[18] In the following, I will employ the acronym 'TECS' to denote transient extended systems in the singular as well as plural form.

> "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 in recognizing as a part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process." (1998, p. 8)[19]

The parity principle "appeals to our rough sense of what we might intuitively judge to belong to the domain of cognition (rather than, say, that of digestion), but attempts to do so without the pervasive distractions of skin and skull." (Clark/Wilson, 2009, p. 72) It allows identifying the external parts of a cognitive process and represents Clark's means of determining whether a given interactive coupling between an agent and some external resource constitutes a TECS.

II. 2. Extended mental states

Having introduced Clark's case for extended cognitive processes and TECS, I will now turn to his argument for extended mental states. Clark notes that even if one agrees that sometimes cognitive processes transgress the boundaries of the skull, this is quite compatible with the view that mental states, for instance, beliefs, desires or emotions, occur only in the brain (1998, p. 12). Mental states may after all be entirely skull bound. So, do mental states extend into the world as well?

Clark gives an affirmative answer to this question. However, he does not claim that all mental states transgress the boundary of the skull. His argument concerns standing beliefs alone. 'Standing' or 'dispositional' beliefs can be explained as beliefs which are not entertained in a current situation. In contrast, beliefs that one is currently entertaining are called 'occurrent' beliefs. To illustrate the distinction, suppose you have the belief that, for

[19] The italics in this and all the following quotations are original if not otherwise indicated.

instance, water is H2O. Shortly before reading this sentence, the belief was a standing belief. (I assume you were not entertaining this belief before reading the sentence.) But now that you have read the sentence and are thinking about the fact that water is H2O, the belief is occurrent. Clark does not argue that occurrent beliefs can extend into the world but only makes the case for extended standing beliefs.

He offers the following thought experiment. Scenario (1): Imagine a woman living in New York City. Let us call her Inga. Inga wants to see an exhibition at the Museum of Modern Art (MoMA). She thinks for a moment and recalls that the museum is on 53rd Street, walks to 53rd Street and visits the exhibition. Inga believes that the museum is on 53rd Street, and she did so before she recalled this information from her memory. Her belief that the museum is on 53rd Street was not previously an occurrent belief. Rather, her belief was stored in memory, available on demand. In short, it was a standing belief.

Scenario (2): Enter another New Yorker, Otto. Otto suffers from dementia. Like many dementia patients, he increasingly depends upon information in his surroundings to organise his life. Otto carries a notebook around with him in which he writes down any new information. Whenever he requires some old information, he looks it up. One day Otto decides, just as Inga, to visit an exhibition at the MoMA. He looks up the location of the museum in his notebook, finds that it is on 53rd Street and walks to the museum.

With these two scenarios in mind, Clark argues that:

> "Clearly, Otto walked to 53rd Street because he wanted to go to the museum and he believed the museum was on 53rd Street. And just as Inga had her belief even before she consulted her memory, it seems reasonable to say that Otto believed the museum was on 53rd Street even before consulting his notebook." (1998, p. 13)

Otto's notebook and Inga's memory are, in fact,

> "entirely analogous: the notebook plays for Otto the same role that memory plays for Inga. The information in the notebook functions just like the information constituting an ordinary non-occurrent belief; it just happens that this information lies beyond the skin" (ibid)

Clark maintains that since Otto's notebook entries are functionally on a par with Inga's biological memories, they should, just as Inga's memories, count as standing beliefs. His reasoning and his assumptions could be summarised thus:[20]

1. That which characterises some information as a "belief is the role it plays" (1998, p. 14) in guiding an agent's action and thinking.

2. "The information in the notebook functions just like the information constituting an ordinary" standing belief (1998, p. 13).

3. Thus, from 1 and 2, the entries in Otto's notebook should count as standing beliefs.

Based on 3, Clark's argument continues:

4. Standing beliefs are part of the mind.

5. Otto's standing beliefs are stored in his notebook.

6. The notebook belongs to the environment.

7. Thus, from 4, 5 and 6, Otto's mind extends into the environment.[21]

[20] See Gertler (2007, p. 2)

[21] Again, as with his argument for extended cognitive processes, Clark considers and rebuts a number of objections to his argument for extended mental states in his original paper. I will not reiterate them here, see Clark (1998).

II. 3. Clark's boundaries of mind

Now, if one accepts with Clark that Otto's mind extends into his notebook then the question arises as to what other entities could possibly qualify as environmental features of Otto's or someone else's mind. Which of the objects that an agent interacts with can count as parts of her mind? If an ordinary notebook qualifies as part of an agent's mind, would the books in one's home library also count? If I carry a PDA with wireless internet access and use the internet to retrieve information, does my mind 'spread across' the internet? How can extended minds be identified anyway?

Clark and Wilson hold that "an extended mind is what you get, given the more basic acceptance of the possibility of temporary, soft assembled extended cognitive systems, if and when certain additional coupling conditions are met […]." (2009, p. 67) Such additional coupling conditions, Clark and Wilson continue, are supposed to warrant that the capacities of an extended system are plausibly considered to be the capacities of a single subject, for instance, Otto.

> "For we properly expect our individual agents to be mobile, more-or-less reliable, bundles of stored knowledge and computational, emotional, and inferential capacities, and so we need to be persuaded that the new capacities enabled by the addition of the notebook are likewise sufficiently robust and enduring as to contribute to the persisting cognitive profile we identify as Otto-the-agent." (ibid)

The 'additional coupling conditions,' – Clark calls them also "glue and trust conditions" (forthcoming, p. 5) – that have "to be met by nonbiological candidates for inclusion into an individual's cognitive system" are:

> "1.That the resource be reliably available and typically invoked. (Otto always carries the notebook and won't answer that he 'doesn't know' until after he has consulted it.)
>
> 2. That any information thus retrieved be more or less automatically endorsed. It should not usually be subject to critical scrutiny (e.g., unlike the opinions of other people). It should be deemed about as trustworthy as something retrieved from biological memory.
>
> 3. That information contained in the resource should be easily accessible as and when required.
>
> 4. That the information in the notebook has been consciously endorsed at some point in the past and indeed is there as a consequence of this endorsement." (Clark, 2008, p. 79)

The 'glue and trust' conditions are supposed to prevent just any external resource that provides pieces of information from automatically counting as a proper part of one's mind. For instance, Clark holds that the books in one's home library would not count (2008, p. 80), presumably because they cannot always be carried around and are thus not as reliably available and easily accessible as an internal resource. Furthermore, he maintains that the internet would not count as an extension of an agent's mind because of condition (2) (ibid). He seems to assume that people do not endorse the information from the internet as automatically as they endorse the information retrieved from biological memory.

Even though, for instance, the internet and the library books do not satisfy the conditions, Clark contends that Otto's notebook does manage to meet them. Since the notebook is coupled to Otto in a sufficiently permanent ("Otto always carries the notebook […]", (2008, p. 79)) and reliable way, Otto and his notebook constitute an extended mind or, as I will in

the following call it, to keep this form of cognitive extension separate from TECS, a permanent extended cognitive system ('PECS').[22]

There is one more important thing about the modal character of Clark's claims about extended mental states and PECS.[23] Clark holds that the 'glue and trust' conditions are

> "fairly stringent, and it is unlikely that any actual notebook currently carried by a human agent will meet the demands. In the context of future technology, however, it may be that reliable more permanent forms of personal cognitive augmentation will be relatively commonplace." (Clark/Wilson 2009, p. 67)

In other words, Clark concedes that in contrast to TECS, which are, as mentioned above, supposedly pervasive, presently there may be no external items capable of satisfying the 'glue and trust' conditions. That is, Clark admits that it is unlikely that PECS currently exist. So, his claims about PECS concern the mere possibility rather than the actuality of PECS.

II. 4. TECS and PECS: A summary of Clark's cognitive extensions

It will facilitate the discussion in the following chapters to briefly summarise the preceding sections and clearly identify what is at issue with Clark's HEC.

First of all, Clark's HEC concerns only two aspect of mentality: cognitive processes, such as mental rotation or performing calculations with pen and paper, and particular mental

[22] Even though Clark discusses and makes the case for both PECS and TECS, as I will explain in V. 2., he concedes that only TECS represent actually existing and not merely possible cognitive extensions.
[23] In the following, I will employ the acronym 'PECS' to denote permanent extended systems in the singular as well as plural form.

states, namely standing beliefs. Clark's claim that cognition extends is thus limited. For instance, he does not argue for the extension of consciousness.[24]

Secondly, corresponding to the two aspects of mentality that Clark considers to be potentially extended, he distinguishes between and discusses two different forms of cognitive extensions: transient extended cognitive systems (TECS) and extended minds or permanent extended cognitive systems (PECS).

As previously mentioned, according to Clark, TECS are actually existing, temporary but ubiquitous soft-assembled wholes that mesh "the problem-solving contributions of the human brain and the nervous system with those of the (rest of the) body and various elements in the local cognitive scaffold." (Clark/Wilson, 2009, p. 65)

I mentioned that to identify TECS, Clark proposes the parity principle: any causal coupling between an agent and some external resource qualifies as a TECS if it would be considered a cognitive process were it performed entirely in the head. Since the parity principle allows identifying TECS, it is the first crucial component of Clark's HEC.

The second component of the hypothesis relates to mental states and extended minds or PECS. PECS are systems in which mental states such as standing beliefs extend into the environment. They are, in contrast to TECS, which can be very fleeting, sometimes even only one-off cognitive extensions (Clark/Wilson, 2009, p. 62), characterised by a more permanent, durable and reliable coupling between external resource and agent. In Clark's account, mental states such as standing beliefs do not extend as easily and temporarily as

[24] Other philosophers go further than Clark. For instance, Wilson (2004, see chapter 9) and Noe (2007, pp. 462) make the interesting, more radical claim that perceptual consciousness extends also. But see Block (2005), Clark (2008, chapter III) and Horgan and Kriegel (2008, p. 20) for critical responses.

cognitive processes such as mental rotation or doing calculations on paper. Mental states extend only under certain conditions that TECS do not have to satisfy (Clark/Wilson, 2009, pp. 66-67). [25] These conditions are the 'glue and trust' conditions. Any causal agent-resource coupling that satisfies the conditions qualifies as a PECS or extended mind (Clark/Wilson, 2009, p. 67). Since the 'glue and trust' conditions serve to identify PECS, systems in which mental states extend, they represent the second crucial component of Clark's HEC.

Furthermore, with respect to this second component of Clark's HEC, I noted that Clark himself concedes that no currently available external resource may be able to satisfy the 'glue and trust' conditions. Consequently, in his discussion of extended minds or PECS, Clark does not claim that PECS actually exist but that they could exist.

A final note to prevent possible misunderstandings, I do not mean to suggest that Clark sees a principled difference between TECS and PECS. He writes that there are "many grades along the continuum from the fleeting to the permanent" (Clark/Wilson, 2009, p. 67). Nevertheless, one can usefully distinguish between TECS, short-lived extended cognitive systems, at one end of the continuum and PECS, more permanent extended cognitive systems, at the other end. They can 'usefully' be distinguished because both forms of cognitive extension are, as I have just mentioned, identified by two different methodological tools (the parity principle and the 'glue and trust' conditions). Furthermore, to keep the two forms of cognitive extension and the two different ways of identifying them apart facilitates a critique of Clark's HEC since he makes different modal claims (i.e. TECS are actual, PECS are merely possible). In my critique of Clark's HEC, I will show that both, the parity principle and the 'glue and trust' conditions are problematic. Both of

[25] That TECS do not have to satisfy the conditions will turn out to be problematic for the tenability of HEC as I will illustrate in chapter V.3.

the methodological tools that Clark proposes to identify cognitive extensions will turn out to be unpalatable. Prior to that, however, I will in the next chapter deal with Adams/Aizawa's critique of Clark.

Chapter III: Cognition re-bound – Adams/Aizawa's critique of HEC

Having introduced Clark's case for HEC and expounded the two forms of cognitive extension for which he argues, I will now turn to a recent critique of his hypothesis. The critique in question is by Adams/Aizawa and put forward in their book *The Bounds of Cognition* (2008). In the next sections, I will explain what I take to be Adams/Aizawa's four main charges against Clark's view.

III. 1. No criterion of the cognitive

The first charge against Clark concerns the following point. Adams/Aizawa contend that Clark's case for extended cognition is incomplete because he does not offer a viable criterion of the cognitive.

Adams/Aizawa hold that a theory which asserts that cognitive processing extends into the environment requires an account of what cognitive processing is. It requires a theory of the cognitive. Adams/Aizawa note that without a criterion to circumscribe cognition it remains unclear what is at issue; and any claim about the location of cognition can hardly be assessed (2008, p. 76).

For this reason, they look for a criterion of the cognitive in Clark's account of extended cognition. Adams/Aizawa hold that Clark offers two different answers to the question of what characterises cognitive processes. The first is that cognition consists of information processing (2008, p. 77). The second is that cognition supports intelligent behaviour (2008, p. 80).

Adams/Aizawa argue that both ways of specifying cognition are problematic. First, simply holding that cognitive processing is information processing is too low a standard to identify cognition, since not all information processing represents cognitive processing. CD players, DVD players, digital computers, cell phones etc. are all information processors, but they are hardly cognitive. Adams/Aizawa argue that an adequate theory of cognition should explain the differences between the information processing in electronic devices and a human agent.

With respect to Clark's second answer to what cognition is – his suggestion that cognitive processes are those processes that support intelligent behaviour – Adams/Aizawa note that it immediately raises the further question of the way 'intelligent behaviour' is to be defined. They hold that Clark has not answered this question. Furthermore, they hold that in some respects respiratory or digestive processes support intelligent behaviour as well. But these processes are clearly not cognitive. Thus, Adams/Aizawa contend, "Clark needs to provide a theory of the 'right kind' of support." (2008, p. 81)

In sum, since Clark does not seem to have a viable theory of the cognitive, his extended cognition view is underdetermined and incomplete. This is the first point of Adams/Aizawa's criticism of Clark's HEC.

III. 2. Fallacious reasoning

A further point of Adams/Aizawa's criticism pertains to the logical structure of Clark's argument. When Clark argues that, for instance, a pen and a piece of paper constitute literal parts of an agent's cognitive system because these external resources assist the agent in accomplishing a cognitive task, he commits a "coupling-constitution fallacy" (2008, p. 88).

The fallacy that Adams/Aizawa have in mind comes in two versions: a simple coupling-constitution fallacy, and a system coupling-constitution fallacy

What is meant by the simple coupling-constitution fallacy? According to Adams/Aizawa, Clark commits this type of fallacy when he moves from the uncontroversial claim that internal processes are causally dependent upon environmental features to the claim that cognitive processes constitutionally include these features. Adams/Aizawa see this fallacy committed and articulated in Clark (1998) in passages such as the following.

> "In these cases [for instance, when doing calculations with pen and paper, UP], the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right.
>
> All the components in the system play an active causal role, and they jointly govern behaviour in the same sort of way that cognition usually does. If we remove the external component the system's behavioural competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head." (p. 8)

However, Adams/Aizawa argue against Clark that it simply does not follow from the fact that just because a process or object X "is in some way causally connected to a cognitive process that X is thereby part of that cognitive process." (2008, p. 91)

To illustrate their point, Adams/Aizawa suggest considering the bi-metallic strip in an ordinary thermostat (ibid). The expansion of this strip is causally coupled to the motion of the atoms in the air of the room in which the thermostat is located. Nevertheless, this evidently does not provide any reason to think that the process of the strip's expansion

literally extends into the atoms in the room. The atoms in the air of the room do not constitute literal parts of the bi-metallic strip just because they are causally coupled to the strip. Arguing otherwise is committing the simple coupling-constitution fallacy.

Adams/Aizawa maintain that Clark does argue otherwise and commits the fallacy when holding that because pen and paper causally contribute to an agent's cognitive process of performing long multiplications, the agent's cognitive processing literally extends into these external resources. However, they contend that, just as, for instance, the process of the bi-metallic strip's expansion does not extend into the atoms of the room the cognitive process of doing arithmetic does not extend into the pen and paper.

Adams/Aizawa hold that Clark commits either the simple coupling-constitution fallacy or the system version of it (2008, pp. 89/126). The system version of the fallacy is the inference that Otto's cognitive processes extend into his notebook because he and his notebook constitute a cognitive system. As Adams/Aizawa point out, "the fact that something is an X system does not entail that every component of the system does X." (2008, p. 118).

An ordinary air-conditioning system (ACS) serves as an example to illustrate the fallacy. An ACS can consist of a thermostat, a compressor, an evaporation coil, a fan etc. Each of these components may be involved in different processes within the system. But even though all these parts and processes together constitute an ACS and are necessary for air conditioning, the actual process of air conditioning consists only of the process of cooling of the air as it passes over the evaporation coils (Adams/Aizawa, 2008, p. 117). In other words, even though the thermostat, the compressor, the fan etc. are all constitutive

components of an ACS, the process of air conditioning does not extend beyond the evaporation coil.

By using this example as a foil, Adams/Aizawa argue that even if Clark can make a case for extended cognitive systems, this case would not support his claim that cognitive processes extend. For instance, one may concede that Otto and his notebook constitute an extended cognitive system but still maintain that his cognitive processing is completely bound to one particular part of the extended system, his brain. His cognitive processing may extend from his brain into the notebook as little as the process of air conditioning extends from the evaporation coil into the thermostat, compressor etc.

So Clark's claim that an agent's cognitive processes extend into the environment cannot be reinforced simply by arguing that the agent and parts of her environment constitute a cognitive system. If Clark wants to infer that cognitive processes extend into external resources from the premise that these resources happen to be part of a cognitive system, he is committing the system version of Adams/Aizawa's coupling-constitution fallacy. However, Clark does not only, allegedly, commit either of the two just mentioned fallacies. Adams/Aizawa put forward more challenges to his hypothesis.

III. 3. Insufficient equivalence between external and internal processes

In a third argument, Adams/Aizawa target Clark's assumption of functional equivalence between internal cognitive processes and external processes. They argue that Clark is wrong in holding that the processes which traditionally have been understood as cognitive or mental can occur in a functionally equivalent way in hybrid systems spanning the brain and parts of the environment. Adams/Aizawa point out that, for instance, Inga's

information retrieval from her biological memory is significantly different from Otto's information retrieval from his notebook. There are "numerous psychological differences" between Inga and Otto. These differences, Adams/Aizawa contend, "provide a principled, non-question-begging reason to believe that there are intracranial cognitive processes in Inga that are unlike the causal processes involved in Otto's interactions with his notebook." (2008, p. 136).

To illustrate some of the psychological differences in question, Adams/Aizawa suggest considering how Inga and Otto would fare in a free recall task. In the task, Inga and Otto may hear 20 words from a list before they are asked to recall the words in any order they please. In her recall performance, Inga will display recency and primacy effects. That is, she will have better recall of the words which were read out to her at the beginning of the list (primacy effect) and at the end of it (recency effect). In contrast, when Otto hears the words, since his notebook functions as his memory, he would quickly write down the 20 words in the notebook and, subsequently, recall all words equally easily and correctly. He would not display the recency or primacy effects that are typical of human recall. Adams/Aizawa hold that the absence of these effects in Otto can be used by "psychologists to draw a principled distinction between the processes and mechanisms in Inga and Otto." (2008, p. 137)

To further illustrate, alongside primacy and recency effects, there are other respects in which Inga's and Otto's recall performances differ psychologically. For instance, if Inga is to associate numbers with particular words, say, 4 with 'table,' 6 with 'apple' etc. and is then asked to remember the word that corresponds to a given number, she will with practice be better able to recall words by the numbers. However, Otto's performance, since he simply has to find the right page and correct number in his notebook, will be excellent

from the start. While Inga is very likely to improve her recall of the correct words over time, Otto will not. The effect of practising recall marks another psychological difference that provides "a principled ground for maintaining that Inga has a kind of cognitive processing not found in Otto." (2008, p. 139) Adams/Aizawa hold that if a psychologist were to "apply many of their familiar experimental protocols for testing memory to Otto and his notebook, they would find vast differences between Inga and Otto." (ibid) And since Otto's notebook interactions are thus not functionally equivalent to genuine cognitive processing, they should not be considered to be cognitive processing.

III. 4. Too strong coupling conditions

A fourth point of Adams/Aizawa's criticism of Clark concerns his 'glue and trust' conditions. As mentioned, that a cognitive agent and parts of her environment constitute an extended mind, or, as I called it, a permanent extended system (PECS), depends in Clark's account upon the fulfilment of these conditions (see II. 2. 2). Adams/Aizawa's argument is that the conditions are too strong. The conditions are too strong because even internal, indisputably cognitive resources may in some cases fail to satisfy them (2008, pp. 120).

Adams/Aizawa illustrate their argument by the following little thought experiment. Imagine Dotto, a professor with normal memory capacities who is able to recall most of the phone numbers of his faculty members and friends. One day he bumps his head on a kitchen cupboard. Though he does not suffer any serious injury and all his cognitive capacities remain the same, he suddenly starts thinking that his memory may be impaired and suspects that his memory of phone numbers may not be reliable anymore. As a consequence of his suspicion, whenever he thinks about calling his friends or faculty members and their numbers correctly 'flash' before his mind, Dotto decides not to dial the

numbers. Instead he opts for writing emails to them. He does not invoke his dispositional beliefs about their phone numbers anymore and no longer endorses the numbers that appear before his mind.

Adams/Aizawa argue that according to Clark's conditions, Dotto's memory, even though it is working as well as it did before the accident, no longer constitutes a proper part of his mind. The reason is that number two of the 'glue and trust' conditions rules that any information that a resource provides must "be more or less automatically endorsed" (2008, p. 110) for the resource to qualify as part of an agent's mind.

However, Adams/Aizawa maintain that since Dotto's memory still provides – unbeknownst to him – all the correct information, Clark's condition becomes implausible. Since Clark's conditions have the consequence of even ruling out that well-functioning internal resources count as parts of an agent's mind, Adams/Aizawa conclude that they are too stringent.

However, Adams/Aizawa note that Clark cannot simply dispose of his conditions that a resource has to be automatically endorsed, typically invoked, easily accessible, reliably available, etc., because this would open the floodgate for counterintuitive instances of cognitive extensions. For instance, without the conditions, the internet, library books, etc., may come to qualify as proper parts of an agent's mind. Thus, Adams/Aizawa claim that Clark faces the following dilemma.

> "On the one hand, Clark needs his trust condition(s) to be necessary, so that such things as library books and internet search engines will not count as part of one's cognitive apparatus. One the other hand, Clark needs his trust condition(s) *not* to be necessary, so that such things as the memory and visual processes one does not trust will count as part of one's cognitive apparatus." (2008, p. 124)

Chapter IV: Beware functionalism – An assessment of Adams/Aizawa's critique

Having introduced the four arguments that I consider to be Adams/Aizawa's main charges against Clark, I will in the following chapter take issue with the first three of them. The just mentioned fourth charge, the argument that Clark's 'glue and trust' conditions are too strong, will be dealt with in the next chapter.

IV. 1. No criterion of the cognitive?

In their first argument, Adams/Aizawa hold that Clark does not provide an adequate definition of cognition. They argue that without such a definition it remains unclear how HEC can be substantiated. Clark's two suggestions of what cognition could be, namely, that it involves information processing and that it supports 'genuine intelligent behaviour,' are unsatisfactory. As mentioned, a CD player processes information too; and the digestive system may also be seen as involving processes that support intelligent behaviour.

However, as it turns out, Clark does have a theory of cognition that grounds HEC. Clark argues that the argument for his hypothesis requires "common sense functionalism" (2008, p. 88) about cognitive states. Before spelling out what is meant by this particular version of functionalism, a brief introduction to functionalism in general will facilitate the discussion.

IV. 1. 1. Functionalism underlying HEC

In the philosophy of mind, functionalism is the thesis that what makes something a "mental state of a particular type does not depend on its internal constitution, but rather on the way

it functions, or the role it plays, in the system of which it is a part" (Levine, 2004). In functionalism, a mental state is functionally[26] identified, that is, it can be specified in terms of three kinds of clauses:

> "input clauses that say which conditions typically give rise to which mental states; output clauses that say which mental states typically give rise to which behavioural response; and interaction clauses which say how mental states typically interact." (Braddon-Mitchell/Jackson, 2007, p. 47)

Put differently, a mental state is characterised and defined by its functional or causal relations to (i) sensory input, (ii) behavioural responses and (iii) other mental states. To illustrate, let us consider the mental state pain as an example. With respect to the causal relation between sensory input and mental state, pain is typically caused by damage to body tissue. This damage to body tissue triggers the belief (mental state) that such damage is detrimental to one's well being. It engenders one's wish (mental state) to avoid that uncomfortable state and initiates behavioural responses (crying, anxiety, etc.) to extricate oneself from the state. According to functionalism, the causal connections between sensory stimuli, overt behaviour and other mental states determine the identity of the mental state pain.

Furthermore, the functionalist maintains that any creature which has a state that manifests these causal connections is capable of being in the mental state pain. The idea is that mental states are 'multiply realisable.' That is, they can be realised by creatures with a very different physiology and chemical composition from ours. For instance, as long as silicon chips or beer bottles manage to play the same roles and fulfill the same functions as neurons in a neural network when bringing about the causal patterns that we associate with a particular mental state, then a being which has silicon chips or beer bottles instead of

[26] Specific versions of functionalism propose different interpretations of 'function.' The different interpretations fall into three main families: computational, causal and teleological. See Van Gulick (2009, pp. 132) for a detailed account of the three ways of interpreting the functionalist notion of 'function.'

neurons in its head is capable of realising this particular mental state as well. The underlying point is that substance variations which do not change functional roles are irrelevant mentally.

Though this multiple-realisability intuition is common ground in all functionalist theories, these theories may differ in the way they account for it. The reason is that they may specify the functional, causal roles of a mental state or cognitive process in different terms. There are, broadly speaking, two different families of systematic approaches to characterising the relevant causal roles.

• Empirical functionalism

• Common-sense functionalism[27]

According to empirical functionalism, mental states should be specified in terms of the causal roles that empirical psychological theories have discovered. [28] Empirical psychological theories have discovered that specific causal processes take place in the brain when an agent is in a particular mental state and exhibits a particular behavioral pattern. These 'fine-grained,' only scientifically observable causal processes in the brain in the event of a mental state constitute, according to the empirical functionalist, a mental state. That is, empirical functionalists do understand a mental state in terms of its relation to sensory input, behavioral output and other mental states but then 'zoom in' on what occurs at a micro-physical level when someone experiences a mental state. They typically focus on the inputs to and the outputs from the central nervous system of a human being.[29] For instance, inputs may be characterized in terms of the activation of retinal cones, stimulation of a particular group of neurons, excitation of certain areas of the visual cortex

[27] See Block (1978) and Braddon-Mitchell and Jackson (2007, pp. 48/84) for this distinction
[28] http://homepages.nyu.edu/~jdc372/Coursenotes/Functionalism.htm
[29] Braddon-Mitchell and Jackson (2007, p. 87)

etc. Relevant outputs may be characterized in terms of certain neural signals that are sent to the motor system to initiate a particular set of behavioral responses and so on.

Since empirical functionalism focusses on fine-grained functional or causal processes, that is, processes at a microphysiological, neural level, it runs the risk of being 'chauvinistic:' empirical functionalism is likely to violate the intuition that creatures with a different micro-physiological constitution from ours can still realise our types of mental states.

In contrast to empirical functionalism, common-sense functionalism does not circumscribe mental states in terms of only scientifically observable processes. It rather specifies them in terms of the 'coarse-grain,' that is, externally observable, inputs and outputs which common knowledge typically associates with a particular mental state.

Common-sense functionalism is the ordinary life grasp of mental states that helps us individuate other people's mental states by allowing us to identify relevant input-output patterns in their behaviour. It fixes the object that the empirical functionalist 'zooms in' on. For instance, with respect to the cause of the mental state pain, common sense characterises inputs in terms of bodily damage and outputs in terms of an agent's avoidance behaviour, wincing, crying etc. If we see an agent wincing and rubbing her thumb after hitting it with a hammer, then, without knowing what exactly takes place neurally in her brain, we intuitively interpret this connection of observable inputs and behavioral outputs as indicating that she is in the mental state pain. Our knowledge of "a broad pattern of flexible, informationally sensitive systemic behavior" underwrites "the ascription of some mental state" (Clark, 2008, p. 88)

Since, according to common-sense functionalist theories, the complex causal interactions at a microphysical, neurophysiological level do not matter for the ascription of mental states, these theories are most likely to preserve the abovementioned multiple-realisability

intuition. However, they also face a problem in that they may be too 'liberal.' If the criterion for the possession of mental states is too permissive, and thus too easily satisfied, it may happen that mental states are ascribed to systems or machines which merely skilfully simulate human behavioral patterns. Such generosity in ascribing mental states is intuitively implausible.

Let us, with these considerations of empirical and common-sense functionalism to hand, return to Clark's HEC. As mentioned, he favours common-sense functionalism about mental states. In fact, he holds that common sense, in identifying the coarse-functional, i.e. behaviour-guiding role of a mental state, identifies "what is essential" to a particular mental state or cognitive process (2008, p. 89). So, common-sense knowledge picks out what Clark considers to be the essence of a mental state. It fixes his 'mark of the cognitive.'[30]

But how does common-sense functionalism support HEC? Clark's point is that Otto with his notebook exhibits the behavioural pattern which underwrites our common-sense ascription of dispositional beliefs to an agent. That is to say, a common-sense functionalist would explain Otto's behaviour of walking up 53rd Street towards the MoMA in just the same way as Inga's going there, namely, in terms of standing beliefs. Since the notebook entries guide Otto's actions in the way that standing beliefs guide Inga's actions, they fulfil the same coarse-grained causal or functional role as normal standing beliefs. For Clark, they share the same essential property.

However, there is a problem. Even though common sense would explain both Inga's and Otto's walk to the MoMA in terms of Inga's and Otto's standing beliefs, Clark's claim that Otto's notebook literally constitutes a physical instantiation of Otto's mind violates

[30] Even though Clark crucially relies upon common-sense knowledge about mental states, as far as I can tell, he has not yet provided a case for this very account of cognition. Clark's hypothesis requires a convincing argument for adopting common-sense functionalism, if it is to be based upon this account of cognition.

common-sense intutions about the location of mentality. Clark's HEC, even though it relies upon common sense to fix what counts as cognitive, flies in the face of the ordinary grasp of mental states.

For this reason, Clark requires, as he puts it, "a kind of extended functionalism" (forthcoming, p. 22) 'on top' of the common-sense account of mental states to make the case for HEC. Clark's point is the following. According to common sense, Otto's notebook entries function in explanations of Otto's behaviour in very much the same way as beliefs function in explanations of Inga's behaviour. If this is the case, then Otto's notebook entries should be classified as beliefs, regardless of whether common sense classifies them in this way. The reason is that "it is a familiar philosophical move to argue that if a state shares the most important explanatory features of a belief, then it really is a belief." (Chalmers, 2008, p. 13) So Clark extends common-sense functionalism by rehearsing the already familiar functionalist intuition that a mental state is identified by its function and not by the way it is physically realised. For Clark, the functionalist intuition about mental states thus trumps common-sense knowledge.

Let us sum up what might be taken to be Clark's rather straightforward theory of the cognitive. In a nutshell, according to his view, a cognitive process is a process which is functionally equivalent to a process that common-sense knowledge classifies as cognitive. Thus, against Adams/Aizawa's claim, Clark does have a theory of cognition. And since his account combines the common-sense understanding of cognition with functionalism and Adams/Aizawa have neither given an argument against the common-sense notion of cognitive processes nor made a case against functionalism, their charge that Clark does not

have an adequate theory of cognition can be dismissed.[31] I will now turn to Adams/Aizawa's second argument.

IV. 2. Fallacious reasoning?

According to Adams/Aizawa's second charge, Clark commits a coupling-constitution fallacy. The fallacy concerns the question of how an environmental feature can become the locus of cognitive processing and a constitutive component of a cognitive system. Adams/Aizawa contend that Clark may give either of two answers to the question (2008, pp. 89/126). First, an environmental feature becomes part of a cognitive system once it is causally coupled to such a system. Second, some environmental features and a cognitive agent together constitute a cognitive system, and in a cognitive system every component does cognitive processing. Both answers to the question of how external features become loci of cognition are interrelated and, as Adams and Aizawa maintain, both involve fallacious reasoning. They involve either a simple coupling-constitution fallacy or a system coupling-constitution fallacy. However, as I will show in the following sections, as a matter of fact, Clark commits neither of these fallacies, because he proposes, as I have explained in the last section, a functionalist account of cognitive processes.

To see the point, let us consider the simple coupling-constitution fallacy first. The fallacy consisted of the following inference: because x is causally coupled to a cognitive process, x is thereby part of that cognitive process. This inference is evidently a fallacy because an agent may use and causally interact with a spatula to flip the burger on the

[31] Note that I do not intend to imply that there are no good reasons to reject the common-sense account of mental states which Clark favours. My point is merely that Adams/Aizawa have not yet made but need to make a case against common-sense functionalism in order to be justified in holding that Clark's account of cognition is inadequate.

grill, but that does not make the spatula a constitutive part of her body. Similarly, even though I may causally interact with pen and paper to accomplish the cognitive task of doing complicated arithmetic, my merely causal interaction with these items does not turn them suddenly into literal parts of my cognitive system.

However, Adams/Aizawa construe Clark as erroneously contending otherwise. They read Clark as saying that *because* the pen and paper are causally coupled to me when I do calculations both items become constitutive parts of my cognitive system. Owing to the causal coupling between me and the pen and paper, my cognitive process of calculating literally extends from my brain into these external objects.

But Adams/Aizawa's interpretation of Clark and subsequently their charge against him is confused. In Clark's argument for extended cognition, the causal coupling between internal and external processes does not play the role that Adams/Aizawa assume it does. It is not for the sole reason that I am causally coupled to the pen and the paper when doing arithmetic that I and these external objects together constitute a cognitive system.

Rather, in Clark's view, the reason I constitute such an extended cognitive system lies in the function or effect of the causal coupling. He is explicit on this issue: "It is not the mere presence of a coupling that matters but the effect of the coupling – the way it poises (or fails to poise) information for a certain kind of use within a specific kind of problem solving routine." (2008, p. 87) For Clark, since the causal interaction between me, the pen and the paper has the same effect or function as an internal cognitive process, the pen and the paper constitute, during my calculations, part of my cognitive system. This conclusion results from his functionalist account of cognition: the function of a resource or process determines the identity of a resource or process. If an external

resource fulfils the same function as some internal, cognitive resource, this external resource should, in the same ways as the internal resource, count as cognitive.

Clark would surely not deny that some sort of causal coupling between internal and external resources has to obtain for cognition to extend. But this is not all that it requires. In Clark's account, even though an agent may be tightly causally coupled to some external resource, if the coupling does not have the same effect of accomplishing a cognitive task as an internal process, the external resource would not constitute part of the agent's cognitive system. For instance, if I had not used the pen and the paper to solve a mathematical problem but to draw instead squiggles or funny faces while waiting to be called up at the dentist, my causal interaction with the pen and the paper would not have sufficed for considering these external objects as parts of my cognitive system. Mindlessly drawing squiggles does not represent a cognitive problem such as multiplying 386 x 417. Consequently, since drawing squiggles is not a problem-solving performance in which pen and paper could play the functional role of some internal resource, even though I causally interact with them, they do not thereby constitute parts of my cognitive system. So, unlike Adams/Aizawa's assumption, the mere causal coupling between me, the pen and the paper is in Clark's view not sufficient for these entities to qualify as proper parts of my cognitive system. Rather, the causal coupling is only one, albeit a very important prerequisite of extended cognitive systems. It may best be understood as a necessary condition for extended cognitive systems: the causal coupling has to obtain for extended cognitive systems to be possible, but it does not guarantee the presence of such a system.

However, only if Clark is committed to the stronger claim that the causal coupling does, in fact, guarantee an extended system will Adams/Aizawa's charge of a simple

coupling-constitution fallacy get off the ground. But, as mentioned, since Clark is clearly not committed to such a claim, Adams/Aizawa's argument misfires.

Nevertheless, the coupling-constitution fallacy comes in two versions. Adams/Aizawa claim that Clark cannot but commit either of the two versions (2008, p. 126). Does he commit the system fallacy? The system coupling-constitution fallacy consists in the inference that Otto's cognitive processes are extended into the notebook because he and his notebook constitute a cognitive system. The inference is fallacious since the "fact that something is an X system does not entail that every component of the system does X." (2008, p. 118)

But Clark does not commit this version of the fallacy either. Adams/Aizawa assume that Clark identifies a cognitive system first and then infers that a given process *x* is cognitive because it belongs to that very system. However, to the contrary, Clark first establishes that *x* is a cognitive process and then, subsequently, draws a conclusion about the boundaries of the cognitive system of which *x* may be part. So, for instance, Clark first establishes that Otto's process of information retrieval should count as cognitive by showing that the process is functionally equivalent to a cognitive process. After that, he asks what allows Otto to realise this process and finds that part of the physical realisation basis of Otto's recalling is the notebook. With this in mind, Clark draws the conclusion that Otto's cognitive system extends into the notebook. So contrary to Adams/Aizawa, Clark's reasoning does not proceed from cognitive systems to cognitive process but from cognitive process to cognitive systems. And what makes an object or process cognitive, in Clark's account, is not its mere belonging to a cognitive system but its functional role. So again, as with respect to the simple coupling-constitution fallacy, since Clark relies upon a functionalist account of cognition, Adams/Aizawa's point of criticism is off target.

IV. 3. Insufficient equivalence between external and cognitive processes?

Adams/Aizawa's third argument against Clark concerned his assumption of functional equivalence between internal and external processes. Adams/Aizawa maintain that Clark is mistaken in assuming that cognitive processes can occur in functionally equivalent ways in hybrid systems spanning the brain and parts of the environment. Their point was that there are a number of psychological differences between, for instance, Otto's notebook interactions and Inga's internal processes. Adams/Aizawa hold that since in free recall tasks and other psychological tests only Inga displays what is typical in human cognition, the processes between Otto and his notebook should not be considered as cognitive.

However, Clark (2008) offers a compelling reply to Adams/Aizawa's argument which is again based on his functionalist account of cognition. He argues that if one uses contingent facts about human cognition or "idiosyncratic features of terrestrial neural activity" (p. 93) to draw a line between what counts as cognitive, then alien beings whose recall is not subject to, for instance, recency effects would not count as cognitive agents. But this is, Clark contends, not very plausible. It contradicts the abovementioned functionalist assumption that cognition is multiply realisable.

To illustrate Clark's point, suppose you meet a Martian whose physiology is just like yours but owing to some odd circumstances on its home planet, the Martian evolved a memory that does not exhibit the weaknesses of human memory. For instance, it only needs to read a twelve-verse poem once to be able to recite it immediately and exactly. Is the price for his better memory exclusion from the group of cognitive agents? Similarly, suppose that, for some reason, a human happens to be extraordinarily good at recalling information and

does not make the typical mistakes his fellows make in psychological tests. Would his surprising memory ability not count as a cognitive ability?

> "Is the mutant human whose recall is fractionally slower, fractionally faster, or much less prone to loss and damage also to be banned from the ranks of true believers and rememberers? To demand identity of fine-grained causal role is surely to set the cognitive bar too high and way too close to home." (2008, p. 93)

It would be "to scale new heights of anthropocentrism and neurocentrism." (ibid)

In short, Adams/Aizawa's third argument against Clark according to which the psychological differences between Otto and Inga suffice to deny that Otto's notebook interactions count as cognitive processing is unconvincing.

IV. 4. Summary

To briefly sum up the present chapter and draw a conclusion, Clark can defend himself against Adams/Aizawa's first three major charges by emphasising his commitment to functionalism. First, he does have a theory of the cognitive: it is a functionalist extension of the common-sense account of mentality. Second, he does not commit a coupling-constitution fallacy: in his account, an external resource does not constitute a part of a cognitive system because it is causally coupled to the system, but because it is functionally equivalent to an internal resource. Third, the psychological difference between Otto and Inga which Adams/Aizawa put forward against considering Otto and his notebook as an extended cognitive system can be rejected as irrelevant by stressing the plausible functionalist assumption that cognition is multiply realisable. To conclude,

Adams/Aizawa's first three charges against Clark's HEC are unsatisfactory because Adams/Aizawa are not sufficiently aware of his commitment to functionalism.

Chapter V: Supersizing HEC – A critique of Clark's position

In the previous chapter, I dealt with three of Adams/Aizawa's arguments against Clark. In the following sections, I will turn to their fourth argument and show that Clark can again counter the argument. However, by thoroughly discussing the consequences of his possible response to Adams/Aizawa, I will gradually expound why there are good reasons not to embrace Clark's HEC. My main point for discouraging acceptance of his hypothesis is:

• Clark's 'glue and trust' conditions to identify PECS and his parity principle to identify TECS allow for 'cognitive bloat,' i.e., they allow for highly counterintuitive instances of cognition.

Before elaborating this claim, I will turn to Adams/Aizawa's fourth charge against Clark. I will refer to their argument first because it directly relates to my claim.

V. 1. On Clark's 'glue and trust' conditions

Adams/Aizawa's argument pertains to Clark's so called 'glue and trust' conditions. As mentioned, for an environmental feature to count as an extended part of a cognitive system, the coupling has to satisfy the following conditions.

• The environmental resource must be reliably available and typically invoked.
• Any information retrieved from the resource must be automatically endorsed.
• The information contained in the resource should be easily accessible when required.
• The information must have been consciously endorsed at some point in the past and is there as a consequence of this endorsement. (see Clark, 2008, p. 79)

Now, Adams/Aizawa's argument against Clark's conditions is that they are too stringent (2008, pp. 121) because they would rule that if, for instance, an agent does not trust her memory anymore she no longer has a memory. However, as Adams/Aizawa emphasise, this is unacceptable. When I cease to trust and to invoke my memory, my memory does not suddenly stop being at my disposal as a cognitive resource. Whenever I try to remember a friend's phone number, the number may still correctly 'pop up' before my mind even though I will be doubtful whether it is the right number. Clearly, my doubts do not change the fact that my memory still does provide correct pieces of information. Adams/Aizawa hold that since Clark's conditions would rule out well-functioning biological, internal parts of the cognitive system, they are too strong.

But Adams/Aizawa maintain that Clark has the problem that he cannot abandon the condition.

> "On the one hand, Clark needs his trust condition(s) to be necessary, so that such things as library books and internet search engines will not count as part of one's cognitive apparatus. One the other hand, Clark needs his trust condition(s) *not* to be necessary, so that such things as the memory […] one does not trust will count as part of one's cognitive apparatus." (2008, p. 124)

However, as I will argue now, Clark does not face this dilemma since he does not 'need his conditions to be necessary' at all. Adams/Aizawa's charge illustrates that they fail to see which party in the extended cognition debate needs necessary conditions and which merely requires sufficient conditions to make its case.

To clarify, the brain-bound party intends to show that cognition *solely* occurs in the brain. In order to do so, the brain-bound proponent requires conditions which allow judging that,

whereas some resources or processes belong to the cognitive system, external resources or processes do not belong to that system. The brain-bound proponent needs conditions that capture the features without which a process or resource cannot qualify as cognitive. Her conditions need to identify what is necessary to count as cognitive. To make her case, the brain-bound proponent must (i) mention the features that are plausibly deemed necessary for cognition and then (ii) show that only the resources and processes in the brain exhibit these features.

In contrast, the extended cognition party does not require such strict, necessary conditions. Since Clark's project is not to set up boundaries of cognition but rather to transgress them, he, unlike Adams/Aizawa, can do without necessary conditions. To make his case for extended cognition, Clark needs only conditions that specify what is sufficient for a process or resource to qualify as part of a cognitive system or cognitive process. His tasks are to (i) propose a set of tenable sufficient conditions and to (ii) show that they are in fact satisfied by some external resource or processes.

Since Clark does not require necessary conditions but only sufficient conditions, he avoids Adams/Aizawa's dilemma. As sufficient conditions, all that Clark's 'glue and trust' conditions say is that if an external or internal resource happens to satisfy the conditions, it is *guaranteed* that the resource constitutes a genuine part of a cognitive system. Whether a given resource which does not satisfy the 'glue and trust' conditions still constitutes a part of a cognitive system is a question that thereby remains simply left open. Thus Clark's conditions neither rule that library books, internet search engines or a mistrusted biological memory do constitute parts of an agent's mind nor that they do not. All that the conditions permit saying is that these resources do not qualify as guaranteed parts of an agent's mind. This is how Clark avoids the force of Adams/Aizawa's argument.

However, Adams/Aizawa may argue that Clark's sufficient conditions are still implausibly demanding. Since a mistrusted internal resource such as biological memory is indisputably cognitive, why not adopt more lenient sufficient conditions?

Clark could respond that his conditions should be accepted because they capture what is *typically* or paradigmatically and not merely *occasionally* true of our relationship with internal, cognitive resources. So, for instance, as Adams/Aizawa correctly observe, occasionally, as the result of, for instance, an accident, we may start mistrusting the information provided by an internal resource and not believe that we recall correctly, but typically we do. In general or most of the time, we automatically endorse the information retrieved from biological memory. Also, even though sometimes the information provided by an internal resource may not be easily available, typically it is thus available. Or, even though occasionally we may not have reliable access to our internal resources, typically we do. So, the conditions are mapped onto typical rather than occasional or exceptional features of internal cognitive processes.

Since the conditions are mapped onto typical rather than exceptional characteristics of internal cognitive processes, they become palatable. It is quite plausible to hold that an external resource must not only be accessed in the way an internal resource is sometimes accessed, but in the way it is typically accessed for the resource to qualify as a guaranteed part of a cognitive system. Since Clark's conditions merely make explicit some of the features of the relationship we typically maintain with internal resources, there is little reason not to accept them. The conditions appear to be adequate means to identify external candidate resources for the inclusion into a cognitive system.

V. 2. On problems with the conditions

That an external resource needs to be as reliably accessible, easily available etc. as an internal resource in order to be considered part of a cognitive system seems, as I have just explained, fair enough. However, Clark's sufficient conditions are not unproblematic. My point will be that, as Clark interprets them, they are unacceptably permissive.

To see why that is the case, let us take a closer look at the way Clark's uses the conditions. As mentioned, he argues that Otto's notebook satisfies the conditions. Otto's taking out the notebook and reading its entries meet the conditions of easy accessibility, reliable availability etc. By satisfying the 'glue and trust' conditions, the notebook becomes functionally equivalent to Inga's biological memory with respect to poise of information. Consequently, Clark claims, the notebook should count, in the same way as biological memory in Inga's case,[32] as a permanent part of Otto's cognitive system. Otto, together with his notebook, should be seen as a permanent extended cognitive system (PECS) (Clark/Wilson, 2009, p. 66). And his interaction with the notebook should just like Inga's memory retrieval be classified as a thought process (2008, p. 81).

But clearly, since Otto's interactions with his notebook involve motor action and visual perception, his process of 'recalling' information is in some respects quite distinct from Inga's turning to her biological memory. For instance, Otto does not have instantaneous speed-of-thought access to the information in the notebook. He needs to open the notebook,

[32] As an aside, Clark's very idea of comparing Otto's notebook with Inga's biological memory may be misguided since it relies upon a passive, mere storage view of memory. Contrary to this view, remembering may be a much more active and constructive process than what the comparison with Otto's notebook interactions suggests. It is noteworthy in this context that Clark (2001b) himself attacks Herbert Simon for subscribing to an "overly passive (mere storage) view of biological memory" (p. 140). See Weiskopf (2008, pp. 267) for a critique of HEC that takes up the issue of whether 'biological' standing beliefs are as inert as Otto's notebook entries.

visually perceive the entries in the notebook, etc. The information in the notebook is not available in the same way as the information that Inga stores in her biological memory and employs in her thought processes. Clark thus already 'stretches' the conditions that an external resource has to be as easily accessible and reliably available as an internal resource when holding that Otto's reaching for his notebook is sufficiently similar to Inga's paradigmatic memory retrieval.

Clark's 'stretching' of the conditions, i.e. his interpretation of them, becomes problematic. If the involvement of motor action and visual perception represents no longer an obstacle to meeting the criteria of easy availability and accessibility, then how much more motor action and visual perception can still be permissible to meet them? For instance, if Otto, instead of reaching for his notebook, needs to walk to some sort of easily accessible room where the notebook is located and readily available for him, would this extra motor action of walking to the room now result in the notebook not being part of his mind anymore? Or is the process of walking to the room and retrieving the information from his notebook in the room still sufficiently similar to his taking the notebook out of his pocket? If it is, then Otto's walk to the room is also sufficiently similar to Inga's memory retrieval. And, consequently, just as Otto's interactions with the notebook qualify as a thought process, so should his walk to the room.

In response, it can be argued that Otto's trip is not sufficiently similar to his reaching for the notebook to allow this inference. One may object that Otto's walking to the room involves more time and more action than his taking the notebook out of his pocket. Also, it certainly involves action of a different kind from the one involved in the pocket notebook case. Furthermore, it evidently requires visual perception which is distinct from that involved in Otto's reading notebook entries. For instance, it may involve perceiving traffic

signs, cars, buildings etc. So it could be argued that when Otto walks to the room his action and perception are not like his action and perception when interacting with the notebook. Thus, owing to this lack of sufficient functional similarity between the two processes, Otto's trip should not be classified as a cognitive process.

However, in reply to this reasoning it may be pointed out that Otto's reaching for the notebook differs equally obviously in action and perception from Inga's memory retrieval. Inga does not have to engage in any manual action of literally 'reaching for' her biological memory and does not visually perceive any 'entries.' If Otto's taking out the notebook meets the conditions of easy accessibility and availability even though it involves more time, more action and action of a different kind from that involved in Inga's typical memory retrieval, then why should Otto's walking to the room with the book not meet the conditions and thus, just like his interactions with the notebook, count as thought process too?

The argument against such an extension of the conditions cannot be that Otto's notebook in the room is not as immediately available and accessible as an internal resource. This argument does not work since it could also be put forward against accepting Otto's pocket notebook as meeting the easy accessibility and reliable availability criteria. Since it involves extra motor action and visual perception, Otto's pocket notebook is also not as instantaneously accessible and available as an internal resource. So, if the involvement of extra action and visual perception is not an argument against accepting Otto's interactions with his notebook as thought processes, then why should it serve as an argument against accepting Otto's walk to the room as a thought process?

But if Otto's walking to the room does not fail the easy accessibility and availability criteria, then the argument can be elaborated by further stretching the 'glue and trust' conditions. Instead of Otto's notebook in the room, we could imagine Inga's *Encyclopaedia Britannica* in her garage or books in a library. Suppose that for some reason Inga happens to have read the *Encyclopaedia Britannica* and a vast number of the books in the library and endorses all the information they contain.[33] Suppose further that whenever she requires some information that could be found in the encyclopaedia or the books in the library, she typically turns to these resources to obtain the information. Now if that is the case, then Inga together with the *Encyclopaedia Britannica* and the books in the library constitute an extended cognitive system.

The obvious fact that she cannot carry the entire *Encyclopaedia Britannica* or all the books of the library with her in the same way as Otto can carry his notebook does not work as a counterargument. For in reply one could point out that Otto does not carry his notebook with him in the same way as Inga 'carries' her biological memory either. But even though this clear difference in processing obtains between Otto and Inga, Clark still considers Otto and his notebook as a PECS, a permanent extended cognitive system.

One may also object that the accessibility and availability of the *Encyclopaedia Britannica* and the library depend upon some complicated motor action and extra visual perception that are entirely absent from Otto's reaching for his notebook. But then again, the accessibility and availability of Otto's notebook depend upon motor action and visual perception that are entirely absent from Inga's paradigmatic memory retrieval as well.

[33] This supposition serves to make the library book and *Encyclopaedia Britannica* meet Clark's fourth 'glue and trust' condition, namely the past-endorsement condition (see II. 3.). However, since Clark himself considers the status of the fourth condition as "uncertain" and "arguable" and in fact concedes that an individual can acquire beliefs without being conscious of this very acquisition (see Clark, 1998, p. 17), the supposition that Inga has read and at some point in the past consciously endorsed all what she read may even be dropped.

Now, Clark could argue that, with respect to accessibility and availability, the notebook in Otto's pocket is much more similar to an internal resource than the *Encyclopaedia Britannica* in Inga's garage or the books in the library. And since the interactions between Otto and his notebook are somehow more similar to internal cognitive processing than his trip to the mentioned room or Inga's drive to the library, only Otto's pocket notebook interactions should qualify as sufficiently similar to cognitive processing.

However, the problem with this point is that we do not have any means to specify 'sufficiently similar' or 'insufficiently similar' to decide whether the *Encyclopaedia Britannica* and the books in the library still meet the easy accessibility and availability criterion or not. In fact, any criterion that allows Otto's notebook to qualify as part of his mind can, by pointing at the inclusion of Otto's notebook, be 'stretched' to also include the *Encyclopaedia Britannica* and the library books. The criteria according to which an item needs to be as easily accessible and readily available as an internal cognitive resource are of little help once Otto's notebook satisfies them. The reason is that as soon as we follow Clark's reasoning and accept Otto's notebook as sufficiently similar to an internal resource, we have embarked on a slippery slope. Any 'stop' on the slide down this slope seems to be arbitrary. Once one agrees that Otto's notebook satisfies the conditions, the decision that Inga's *Encyclopaedia Britannica* and the books in the library do not satisfy them becomes arbitrary. The reason is that the argument for the claim that Otto's notebook satisfies the conditions is the same as the argument for the claim that Inga's *Encyclopaedia Britannica* and the books in the library do so too. If Clark wants to include Otto's notebook in the folder of cognitive extensions but not Inga's *Encyclopaedia Britannica* and the library books, it seems that he has replaced one arbitrary limitation to cognition, namely the skull, by another seemingly even more arbitrary one, namely, his particular interpretation of the 'glue and trust' conditions. But on the other hand, if Clark concedes that Inga's

Encyclopaedia Britannica and the books in the library do satisfy the conditions, then his conditions become unacceptably easy to satisfy.

Note that the present argument against Clark's conditions only continues his own line of reasoning. Just as there is functional equivalence between Inga's biological memory and Otto's notebook, so there is functional equivalence between Otto's notebook and Inga's *Encyclopaedia Britannica* and the library books. Together with the *Encyclopaedia Britannica* and the books in the library, Inga should count as a PECS in the same way as Otto and his notebook do.

Note also that Clark shares the view that this consequence is quite implausible. He writes that there are "absurd inflation[s] of an individual mind" and holds that, for instance, "counting the database of the *Encyclopaedia Britannica*, which I keep in my garage, as part of my general knowledge" is an "intuitively pernicious extension ('cognitive bloat')" (2001, pp. 155-156) of mind. He claims that his conditions would prevent such 'cognitive bloat.' But, as I have illustrated, they do not accomplish this task unless one stipulates a seemingly arbitrary stop to the 'stretching' of the conditions.

Before closing my discussion of Clark's conditions for PECS, it is worth remembering that he concedes that the 'glue and trust' conditions are

> "fairly stringent, and it is unlikely that any actual notebook currently carried by a human agent will meet the demands. In the context of future technology, however, it may be that reliable more permanent forms of personal cognitive augmentation will be relatively commonplace." (Clark/Wilson 2009, p. 67)

That is, Clark concedes that currently there may be no actual PECS. However, if PECS are presently unlikely to exist, Clark's rhetoric about the mind's 'leaking out' into the world loses some of its radical character (2008, p. xxviii). There may yet be no mind 'leaking out' into the world as external resources may not yet be sufficiently similar to internal resources to support mental states such as standing beliefs. If Clark's HEC involves a more interesting and more substantial modal claim than that cognition *could* in the future extend into the environment, it will be about the actual existence of extended cognition. It will not merely involve arguments for what could be the case but for what is in fact the case.

Also, as mentioned in my exposition of HEC in a previous chapter, Clark does make claims about actual cases of extended cognitive processes and systems. He argues that in contrast to PECS, TECS, transient extended cognitive systems, actually exist. In fact, he asserts that they are ubiquitous and constitute "quite literally, the soft-assembled circuitry of a great deal of practical human thought and reason." (Clark/Wilson, 2009, p. 73) For instance, Clark contends that whenever an agent solves mathematical problems with pen and paper or manually rotates geometric shapes to align them with depicted sockets then the agent, together with the external resources she employs, constitute a TECS. Clark's more substantial case for extended cognition will thus be about TECS rather than PECS.

V. 3. On problems with the parity principle

As with respect to PECS, whether or not Clark's case for TECS succeeds will depend upon the plausibility of the condition under which he thinks that cognition transiently extends. In the following sections, I will examine Clark's sufficient condition for TECS, that is, the parity principle. My claim is that Clark's principle leads even more easily and more radically to 'cognitive bloat' than his 'glue and trust' conditions for PECS.

First however, why not also use these conditions to identify TECS? The reason Clark cannot use the 'glue and trust' conditions to identify TECS is that they are too stringent to determine cases of merely temporary and soft-assembled cognitive extensions. Clark wants to argue that even if an agent may not always have a pen and paper reliably available or may not typically employ them to do calculations, on the occasions when the agent happens to use these external resources to do arithmetic, he or she together with pen and paper constitute an extended system (Clark/Wilson, 2009, p. 67). Since the 'glue and trust' conditions demand reliability and durability of the coupling between agent and external resource, they are inappropriate to identify soft-assembled and sometimes even only "one-off" extended cognitive systems (Clark/Wilson, 2009, p. 62). So Clark needs another means to identify TECS.

He suggests, the following parity principle as a "rule of thumb" for the "purpose of identifying material vehicles of cognitive states and processes" (2008, p. 77).

> "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 in recognizing as a part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process." (1998, p. 8)

Clark's parity principle helps to determine which external resources belong to the domain of the cognitive without making explicit and stringent demands on the reliability and durability of the relationship between agent and external resource. Since the principle does not involve such demands, it allows Clark to identify ephemeral extended cognitive systems.

However, this permissiveness has its costs. As it turns out, the principle is far too permissive and allows for an implausible inflation of cognition. To begin with, according to the parity principle, any object that is visually perceived would qualify as a proper part of an agent's cognitive system. In order to see the point, let us consider one of the psychological phenomena that Clark (2007) uses to bolster his case for extended cognition: the phenomenon of 'change blindness'.

Change blindness concerns visual awareness. When visually perceiving a scene, subjects are often convinced that they see all the details of the scene and can instantaneously notice any modification to it. However, Rensink et al. (1997) discovered that "when brief blank fields are placed between alternating displays of an original and a modified scene, a striking failure of perception is induced: identification of changes becomes extremely difficult, even when changes are large and made repeatedly" (p. 368).[34] Even though subjects will hold that the original and the modified scenes are identical, they are not.[35] This is the phenomenon of change blindness.

It suggests a surprising sparseness of what we are consciously aware. Clark writes that one

> "diagnosis of why we are not normally aware of any such sparseness is that our feeling of 'seeing all the detail' in the scenes (and hence the surprisingness of the demonstrations of unseen changes) is really a reflection of something implicit in the overall problem-solving organization in which vision participates. That organization 'assumes' the (ecologically normal) ability to retrieve more detailed info when needed, so we feel (correctly, in an important sense) that we are already in command of the detail." (2007, p. 275)

[34] During one of his presentations on TED, Dennett (2003) uses some interesting visual illusions to illustrate the phenomenon of change blindness, see http://www.ted.com/index.php/talks/dan_dennett_on_our_consciousness.html.
[35] See also Rensink et al. (1997) for a detailed discussion of the phenomenon.

To return to Clark's parity principle, it follows that whatever external object or scene I visually perceive already constitutes a physical component of my cognitive system. The reason is that any visually perceived external object or scene provides and stores information about its own details that could be provided by a mental representation of the object or scene. As Clark writes, the "external scene" functions as a "memory store" (2007, p. 275): it does what a biological, cognitive resource could do. According to the parity principle, since any visually perceived object or scene would be considered as a cognitive information resource if it were somehow stored in the head, any visually perceived object or scene should be considered as a cognitive resource and part of an agent's cognitive system.

However, the contention that any environmental feature which is visually perceived automatically belongs to a cognitive system because it stores information that could be provided by an internal resource or by mental representations, clearly results in an implausible over-extension of cognition. If, for instance, you strolled along the beach, walked through the park, city etc. – with open eyes – you would in fact be walking through your cognitive system.

As a further example of how the parity principle leads to an implausible inflation of cognition, consider the internet. Whenever an agent retrieves pieces of information from the internet, together with her computer and the World Wide Web, she would constitute one of Clark's transient extended cognitive systems. The reason is again grounded in the parity principle: since the process of retrieving information from an external storage device would be considered a cognitive process if it were performed inside the agent's head, retrieving information from an external resource such as the internet should be considered a cognitive process also. The computer and internet should be considered to be temporary

physical machinery of the agent's cognitive system because they do fundamentally the same job in providing pieces of information as an internal resource, as Otto's notebook or as a piece of paper when doing calculations. Just as an internal resource or Otto's notebook does, the internet serves as a storage device that poises information which guides an agent's action and thinking.

To give a final illustration of the problem with Clark's parity principle, according to the principle, one's cognitive system may even extend into the planetary system. Long before the invention of the magnetic compass, sailors determined their position, destination, and direction at sea mostly by the sighting of landmarks or the observation of the position of celestial bodies. The sun or other known celestial bodies were information resources which helped sailors navigate across the ocean.

In Clark's account, the celestial bodies that were and may still be used by sailors to determine their position when navigating a ship across the sea, should be considered as literal parts of the sailors' temporarily – for the time of the navigation – extended cognitive system. The reason is the parity principle: the stars fulfil fundamentally the same function as an internal mental map employed for the navigation to a particular destination. Since the stars would count as parts of a sailor's cognitive system if they were somehow internal, mental information resources, they should count as parts of that sailor's cognitive system even though they may be light years away.

To conclude, since soft-assembled TECS do not have to satisfy the four coupling conditions that Clark proposes to identify PECS, a highly implausible 'supersizing' of cognition and proliferation of cognitive systems ensue. The parity principle, which constitutes Clark's means of identifying TECS, is unacceptably permissive.

Clark may bite the bullet and hold that, even though it seems counterintuitive, the cognitive system does indeed extend into any object one visually perceives, spreads across the internet and comprises the planetary system.

But then the question of why one would want to make such assertion may be raised. To put it more bluntly, what is the point of holding that the mind is so extremely extended? What is to be gained by adopting a view that runs profoundly counter to common-sense intuitions? These questions will lead me in chapter IX to another reason for not accepting Clark's HEC. However, before delving more deeply into that, I will in the next chapters examine whether the brain-bound view proposed by Adams/Aizawa fares better than Clark's extended cognition view.

Chapter VI: Adams/Aizawa's case for brain-bound cognition

If there are reasons for rejecting Clark's conditions for cognitive systems and for not endorsing the view that cognition sometimes transgresses the boundaries of the skull, maybe the orthodox view that it occurs solely in the head represents the more palatable position in the current extended-cognition debate. In the debate, Adams/Aizawa are siding with orthodoxy and propose an internalist view. Since they want to restrict cognition to what occurs in the brain, they need to (i) propose tenable necessary conditions for cognition and then (ii) show that external resources do not satisfy these conditions. In the following sections, I will turn to an exposition of their necessary conditions for cognition and their argument for the brain-bound view.

According to Adams/Aizawa's view, the cognitive is distinguished from the non-cognitive by two features. The first "difference between cognitive processes and non-cognitive processes is that the former involve non-derived representations […]" (2008, p. 57). The second difference is that cognitive processes, unlike non-cognitive processes, are characterised by specific "kinds of mechanisms" (ibid). These are Adams/Aizawa's two necessary conditions for cognition. Both figure in their case for the brain-bound view. I will elucidate how they do so by considering them in turn.

VI. 1. Non-derived representations

The first condition for cognition is that genuinely cognitive processes always involve 'non-derived representations.' Adams/Aizawa differentiate between 'non-derived' and 'derived representations.' They hold that 'non-derived representations' are "mental representations that mean what they do independently of other representational or intentional capacities"

(2008, p. 31). Non-derived representations do not inherit their content from other content, convention, or interpretation by intentional agents but are originally contentful, that is, contentful in themselves. Adams/Aizawa hold that "[t]houghts, experiences, and perceptions are paradigm cases of items bearing non-derived content." (ibid)

In contrast, representations with derived content would be, for instance, traffic lights, gas gauges, words or flags. Unlike non-derived representations, traffic lights, words etc., do not mean anything in themselves or originally. Rather, their content depends upon intentional acts of interpretation and conventional agreement between agents. Items with derived content mean what they mean because agents, who already possess meaningful thoughts, agree upon a particular way of understanding and handling these items (2008, p. 32). This is Adams/Aizawa's distinction between derived and non-derived representations.

But why assume that there are representations with non-derived as opposed to derived content anyway? One of Adams/Aizawa's motivations for positing representations with non-derived content is that such content seems to be required to make sense of "lone thinkers" (2008, p. 33). One of their examples to illustrate the point is a solitary orang-utan in a rainforest. The orang-utan may reach for a banana because she visually perceives the fruit and has the desire to eat it. "A natural explanation of the animal's behavior is that she has cognitive states about the fruit and the state of her dietary needs. Yet, alone, in the forest, the current contents of her thoughts are not the product of any neighbouring cognitive agents" (2008, p. 34). The orang-utan is, just as the average human being, "capable of thinking about things all by him- or herself" independently of other intentional agents (ibid). To explain lone thinkers, Adams/Aizawa hold that one cannot avoid invoking representations with some sort of non-derived content.

Another reason for assuming that thinking involves representations with non-derived content is that if any thought possessed only derived content and always had to be interpreted, an infinite regress would ensue. Thought *x* with derived content *a* would have to be interpreted by thought *y* whose content *b* would itself have to be interpreted by thought *z* and so on. To prevent a regress of interpretations and to explain the possibility and determinacy of the first thought, it seems again inevitable to posit non-derived, interpretation-independent content.

However, the origin of non-derived content requires in turn an explanation. Adams/Aizawa themselves do not provide such an explanation. Instead they indicate that Fred Dretske's information theoretic account, Jerry Fodor's asymmetric causal dependency theory and Robert Cummins' picture theory of representation constitute theories of content which attempt to explain what Adams/Aizawa have in mind with non-derived content (2008, p. 36-37).

Nevertheless, with respect to these accounts, Adams/Aizawa admit that none of them is uncontroversial. They write that philosophers as well as psychologists have yet to formulate a widely accepted theory of content and concede that since an uncontroversial theory of content is currently not available, it remains unclear "what naturalistic conditions give rise to non-derived content […]" (2008, p. 55). Consequently, "it remains correspondingly unclear just exactly what objects bear non-derived content" (ibid).

Notwithstanding, Adams/Aizawa claim that "as a matter of contingent empirical fact," non-derived representations "happen to occur these days only in nervous systems" (ibid).[36] And since cognitive processes always operate on non-derived representations but non-

[36] Note that they do not offer any further argument for this claim.

derived representations only occur in the brain, Adams/Aizawa maintain that there is a "non-question-begging, defeasible reason to think that, contrary to what advocates of extended cognition propose" cognitive processing is solely located within the brain (ibid).

VI. 2. Cognitive mechanisms

I turn to Adams/Aizawa's second condition for cognition and argument for the brain-bound view. The condition in question is that cognitive processes differ from non-cognitive processes "in virtue of the kinds of mechanisms involved" (Adams/Aizawa, 2008, p. 57).

To determine what sort of mechanisms are characteristic of cognition, Adams/Aizawa point at those underlying human cognition. That is, they focus on the psychological regularities in human cognition that are studied by cognitive scientists and psychologists (2008, pp. 60). Examples of such regularities are the abovementioned recency and primacy effects of the human memory system (see III. 3) or George A. Miller's 'magical number seven' (Adams/Aizawa, 2008, p. 62). According to Miller, an average human being can only retain as many as seven plus or minus two items in short-term memory. Our capacity to recall information from short-term memory is confined to this number of digits, words, items etc.

Adams/Aizawa argue that in discovering such regularities as Miller's 'number seven' or primacy and recency effects, cognitive psychologists have uncovered some distinctive features of cognitive processing that can be utilised to demarcate the cognitive from the non-cognitive. The idea is that a process must exhibit the scientifically known mechanisms of human cognition to be classified as a cognitive process.

However, with this in mind, Adams/Aizawa contend that external processes are unlikely to be classified as cognitive. The reason is that external processes are unlikely to exhibit the

typical mechanisms of human cognition because they do not have the same intrinsic, microphysical profile as the internal, neuronal processes which normally implement cognition. Adams/Aizawa argue that since "differences in realizer properties and processes produce differences in realized properties and processes […]" external processes are as a matter of contingent fact highly improbable to implement cognition (2008, p. 70). Because brain processes typically realise cognitive processes and because brain processes differ in their intrinsic physical nature from brain-environment hybrid processes, there are again "defeasible" reasons to suppose that cognitive processes are "brain bound and do not extend from the nervous system into the body or environment." (ibid)

VI. 3. An assessment of Adams/Aizawa's case for brain-bound cognition

Having introduced the two necessary conditions for cognition upon which Adams/Aizawa base their case for the brain-bound view, I will in the following sections scrutinise the conditions. As will become clear, Adams/Aizawa's two conditions have the flaws of being either too permissive in that they allow for extended cognition, or too restrictive in that they do not permit classifying some processes as cognitive which are plausibly thus classified. Since Adams/Aizawa's conditions underlie their argument for the brain-bound view, the critical aspects which weigh against the former will equally affect the latter.

VI. 3. 1. On non-derived representations

According to Adams/Aizawa's first condition for cognition, all "cognitive processes have in common […] that they […] involve non-derived representations" (2008, p. 55).

Note that if Adams/Aizawa intend to base their case for the brain-bound view upon this mark of the cognitive, they must establish two points: (i) all cognitive processes operate only on non-derived content and (ii) no external item which contemporary human beings use can support non-derived content.

Certainly, Adams/Aizawa do not have to demonstrate these two points if they only wish to argue that in most but not all cases human cognition is brain-bound. However, if that were their goal, there would be little ground for disagreement between them and the proponent of HEC. After all, the proponent of HEC may very well concur that most of cognition takes place in the head. What is at issue in the extended cognition debate is whether cognition is currently extended at all. Since Adams/Aizawa want to deny this, they need to show (i) and (ii). In the following, I will focus my discussion on point (i).

With respect to (i), Adams/Aizawa already concede that they "have no good reasons to think that cognitive states must consist entirely" of non-derived representations (forthcoming, p. 5). And they write that "it must be admitted that it is unclear to what extent every cognitive state of each cognitive process must involve non-derived content" (ibid).

However, Adams/Aizawa immediately emphasise that to be considered a cognitive process any process must at least involve *some* non-derived content. If it does not involve any intrinsic content at all, "then the condition rules that the process is non-cognitive" (forthcoming, p. 5).

With this clarification on their condition, Adams/Aizawa invite the following response from the extended cognition proponent. Since Otto's interactions with his notebook do not

only involve linguistic symbols with derived content but also occurrent mental states about the notebook, Otto's notebook interactions would satisfy Adams/Aizawa's condition. The point is that at the particular moment in time when Otto retrieves the information about, for instance, the address of the MoMA for his notebook, the process of information retrieval does not only involve derived content. Rather it represents a hybrid process which operates on two different types of content and representations (Clark, 2007, p. 12). There are some internal representations involved which may be seen as bearing non-derived content (for instance, Otto's internal representations and thoughts about the notebook), and others which derive their content from interpretation and convention (namely the written symbols that the notebook contains). Consequently, at the time Otto consults his notebook, his notebook consultation satisfies Adams/Aizawa's condition according to which a process needs to operate at least on some non-derived content to qualify as cognitive.

Since this is evidently not in Adams/Aizawa's interest, what if they retracted their concession that a cognitive process need not involve only non-derived content? If Adams/Aizawa retreated from this concession and maintained that a cognitive process must not involve any interpretation-dependent element at all (i.e. if they adhered to the abovementioned point (i)), their condition for cognition would seem implausibly stringent.[37] Clark (2008, p. 91) offers a thought experiment to illustrate the point. He imagines Martians who are endowed with a biological mechanism that enables them to store bit-mapped images of visually encountered, conventional symbols. At will, the

[37] Furthermore, to rule out that cognition extends by utilising their condition, Adams/Aizawa would still have to show that Otto's notebook inscriptions cannot support non-derived content. To do so, a theory of content according to which external items cannot bear non-derived content is required. As mentioned, Adams/Aizawa do not propose such a theory. Rather, they admit that no widely accepted theory of content currently exists and that "it remains correspondingly unclear just exactly what objects bear non-derived content" (2008, p. 55). However, if that is the case then Adams/Aizawa's argument that external items cannot qualify as cognitive because these items cannot support non-derived content becomes questionable. After all, it may well be that the theory of content which will finally be accepted allows that external items bear such content. Since an uncontroversial theory of content still awaits discovery, Adams/Aizawa's claim about the location of non-derived content appears premature.

Martians can access these representations, interpret the stored symbols and employ them in their decision-making and action-planning. The question that Clark now raises is whether we would consider the bit-mapped images as part of the Martians' cognitive system. Since the images are internally stored in the Martians' heads, in the same way as biological memories, and since they fulfil the same functional role in guiding the Martians' behaviour as internal non-derived representations, it seems plausible to hold that they do constitute proper parts of these creatures' cognitive system.[38] However, if one subscribes to the intuition that such interpretation-dependent representations would qualify as genuine components of a cognitive system, then it becomes clear that Adams/Aizawa's non-derived content condition for cognition is too restrictive.

Furthermore, if one shares the intuition with respect to the Martians, there is little reason not to also apply its upshot to the familiar Otto-notebook scenario. As a result of such application, the mere fact that Otto's notebook inscriptions are interpretation dependent would no longer undermine the claim that they constitute parts of Otto's cognitive system. Otto's cognitive system could extend into his notebook even though the notebook contains only derivedly meaningful written symbols.

Since Adams/Aizawa surely want to avoid this conclusion, they need to address Clark's Martian thought experiment. In response to the thought experiment, Adams/Aizawa reject the intuition underlying the experiment and hold that it is "perfectly reasonable for us to stand by the view that these Martian representational states are not cognitive states." (2008,

[38] Instead of Martians, one could also imagine future humans with a neural implant that stores images of visually encountered written texts and projects these images, upon thought command, on a retinal display. Once they appear on the display, the images are interpreted and used to guide actions. For these future humans, the images would play the same functional, behaviour-guiding role as 'normal' internal representations. The only difference would be that the former, unlike the latter, require interpretation. Again the question could be asked of whether this difference should make a difference with respect to the cognitive or non-cognitive status of the images. Also, what would be the motivation for denying that they are cognitive? Adams/Aizawa would have to mention some disadvantages or problems that result from counting them as cognitive. I will return to this issue at the end of the next section.

p. 49) To support their stance, Adams/Aizawa give the following example. We are asked to imagine that future science discovers some strange animals that have the same external appearance and exhibit the same behaviour as ordinary ducks. However, even though externally they are entirely indistinguishable from ducks, unlike ducks, these animals do not have DNA in their cells. Rather, instead of DNA, the cell nuclei of these animals contain another chain molecule that happens to fulfil exactly the same biochemical functions as DNA in familiar, ordinary ducks.

Referring to this scenario, Adams/Aizawa write that it

> "seems to us to be perfectly defensible to maintain that these new things aren't ducks. What we have in the Martian case and our duck-like thing case are instances where we are invited to entertain surprising and even bizarre possible findings that appear to conflict with our theories of what constitutes cognition and what constitutes a duck." (2008, p. 49)

Now, what can be inferred from Adams/Aizawa's response to Clark is that they, unlike Clark, rest their demarcation of the cognitive upon intrinsic rather than the functional properties of cognitive processes.

Is Adams/Aizawa's way of defining the cognitive plausible? As a matter of fact, it is neither plausible nor very consistent with their avowed intention to base their case for brain-bound cognition upon current methodology and practice in cognitive science (see Adams/Aizawa, 2008, p. 9-12, p. 84). First, Adams/Aizawa's approach is not very plausible because it does not take into account the fundamental and widely accepted functionalist assumption that cognitive process can be realised in many different ways and by many different types of representations and physical substrates. Second, as Rupert

already notes, even though Adams/Aizawa repeatedly invoke cognitive science and psychology, their approach is "out of step with cognitive psychological practice" (forthcoming, p. 10).[39] In cognitive psychology, some neurons become cognitively relevant not because of their intrinsic properties or nature but because of their function and the way they manage to contribute to intelligent, goal-directed behaviour. Consequently, "we should expect cognitive psychology to deliver a functional criterion distinguishing the genuinely cognitive from the non-cognitive […]" (ibid). So to construe cognition functionally in terms of 'what it does' rather than in terms of its contingent intrinsic features is not only more likely to do justice to the view that cognition can be multiply realised, it is also more in line with current cognitive psychological custom.[40]

VI. 3. 2. On cognitive mechanisms

Having dealt with Adams/Aizawa's first necessary condition for cognition, I turn to their second condition. Adams/Aizawa's second condition was that cognitive processes, unlike non-cognitive processes, exhibit specific kinds of mechanisms. Examples of such mechanisms were recency and primacy effects and Miller's 'number seven.' To make a compelling case for brain-bound cognition by drawing on these mechanisms, Adams/Aizawa must again demonstrate two points. (i) All cognitive processes exhibit

[39] Rupert (forthcoming, p. 6) mentions research by Baron-Cohen, Nichols/Stich and Anderson as evidence for the view that in cognitive science, cognitive processes are predominantly specified in terms of functional role rather than intrinsic properties.

[40] Of course, if Adams/Aizawa proposed a theory of non-derived content which is functional, then they could avoid the present criticism. However, first, they have not done that yet. Second, if they did specify non-derived content functionally and thereby provided a functional criterion of the cognitive, the advocate of HEC may ask for the reasons one should accept Adams/Aizawa's functional criterion rather than Clark's. Two different functionalist notions of cognition would be on the table. Adams/Aizawa's criterion would probably be fine-grained in nature, as their reference to the microphysical structure, the DNA, of ducks in their abovementioned example indicates. In contrast, Clark's criterion is, as explained in a previous chapter, coarse-grain in character. The decision on adopting a fine-grained rather than coarse-grained functionalist specification of cognition would have to be motivated. Adams/Aizawa would have to provide further arguments to avoid begging the question against the proponent of HEC, see Rupert (forthcoming, p. 7) for a similar point.

these mechanisms and (ii) all brain-environment hybrid processes fail to display them. I will focus on (i).

First, contra Adams/Aizawa's point (i), it is quite plausible to assume that a process which does not exhibit the mechanisms of typical human cognition can still be a cognitive process. Put differently, against point (i), not all cognitive processes exhibit these mechanisms. As a brief illustration, consider the following. Suppose Jack, an ordinary human agent, hit his head on a kitchen cupboard. He is not sure whether the accident damaged his memory capacity and decides to undergo a psychological test. As part of the test, he is asked to recall a long list of words after someone has read the list out loud to him. After the test, it turns out that, unlike a normal human agent, Jack did not succeed in having better recall of the first and the last words on the list. For some reason, he happened to remember best, and with extraordinary accuracy, the words in the middle of the list instead. In other words, when recalling, Jack no longer exhibits the recency or primacy effects typical of normal human agents.

Now, if Adams/Aizawa demand that any process needs to exhibit the typical features of human cognition to qualify as cognitive, Jack's performance during the test will not qualify as cognitive processing. However, in the light of the fact that he does still remember the words in the middle of the list with high accuracy, this seems to be a counterintuitive judgement. Instead of denying Jack the cognitive process of retrieving information from memory, it appears much more tenable to hold that he does in fact realise this cognitive process but simply does not do so in the 'normal' or familiar way. Since the assumption that Jack realises a cognitive process is quite solid, Adams/Aizawa are wrong in holding that a process which fails to exhibit the known psychological regularities of human cognition cannot be a cognitive process.

That their way of demarcating the cognitive is unacceptably restrictive becomes also evident with the following consideration. The psychological regularities that Adams/Aizawa have in mind (Miller's 'number seven' etc.) are essentially deficiencies of human cognition. For instance, an average human has the weaknesses of being able to remember *only* up to seven plus or minus two items. Why should a creature have to be as imperfect as a human being to qualify as a genuinely cognitive agent? As a matter of fact, we are often stunned by imaginary characters such as aliens, psychics, deities or the like whose cognitive abilities exceed the limitations typical of the average human. The fact that the mental performances of such extraordinary beings are not characterised by the shortcomings a human normally exhibits clearly does not speak against considering them as genuinely cognitive. On the contrary, it rather suggests classifying them as more perfect cognitive abilities than those a typical human possesses. Adams/Aizawa's condition that any being must display the imperfections that characterise our species to qualify as cognitive is counterintuitive (Fisher, 2008, p. 350).

However, Adams/Aizawa are aware that circumscribing the cognitive by exclusively drawing on the contingent features of normal human cognition is too insular and stringent (2008, p. 70). They note that there may well be cases of atypical cognition. For instance, an otherwise normal human being who is only able to retain five items, plus minus two in short-term memory "would not, thereby, become a non-cognitive agent. Nor would such a person be said not to possess memory." (ibid)

To accommodate the possibility of atypical human and cases of non-human cognition in their account, Adams/Aizawa suggest that any processing only needs to possess a "family resemblance" to human cognition to qualify as cognitive (2008, p. 72). This move helps them avoid the charge of proposing an unacceptably strong mark of the cognitive.

Does their now modified criterion then serve as a tenable means to make a convincing case for brain-bound cognition? It may do so, if Adams/Aizawa can establish the point that no external process can bear a family resemblance to human cognitive processes. The demonstration of this point presupposes a specification of their notion of family resemblance. However, Adams/Aizawa have not provided any such specification. Consequently, their second criterion for cognition does not help to establish that cognition is entirely brain bound. In fact, since Adams/Aizawa have not defined the properties that are relevant to the resemblance they have in mind, the proponent of extended cognition receives the chance of arguing that Otto's notebook does bear a family resemblance to normal human biological memory. He or she may argue that the functions of Otto's notebook resemble the functions of Inga's biological memory. For instance, Otto and Inga both employ their sense organs to obtain information from the environment. They both convert this information into some sort of lasting and portable symbolism. Furthermore, they both have methods of retrieving that information when it is needed. And both Otto's as well as Inga's information storage resources are susceptible to malicious tampering by a third person. The proponent of extended cognition may contend that in all these respects, Otto's notebook bears a family resemblance to Inga's biological memory. Consequently, even according to Adams/Aizawa's second mark of the cognitive, the notebook would qualify as cognitive (Fisher, 2008, p. 350).

To avert such a conclusion, Adams/Aizawa must specify their criteria of family resemblance for this particular case. They must do so in such a way that these criteria are, on the one hand, permissive enough to allow for atypical and nonhuman cognition (so that they can avoid the above charge of proposing an implausibly demanding condition) and, on the other hand, restrictive enough to preclude extended cognition. Furthermore, Adams/Aizawa would additionally have to make sure they do not beg the question against

the extended cognition proponent in this balancing act (Fisher, 2008, p. 350). Nevertheless, despite all these difficulties, Adams/Aizawa may be able to define their notion of family resemblance in exactly the way they need it to be. My point here is that they have so far not done that but rather left their criteria for resemblance in the notebook-versus-biological-memory case completely undefined. Consequently, without further argumentation and conceptual clarifications, they cannot establish what they set out to establish, namely that cognition is entirely brain-bound.

Up until now, I have argued that, as it stands, Adams/Aizawa's second condition for cognition is either too strong in that it does not allow for atypical cognition, or too permissive in that it permits HEC to come out true.

A final reason for not adopting Adams/Aizawa's two conditions more generally is pragmatic in nature. The point is that their conditions may impede scientifically fruitful conceptualisations. For instance, if it turns out to be conducive to cognitive-scientific investigations to classify a process as cognitive even though it operates on derived content or does not exhibit the typical mechanisms of human cognition, then there is little reason not to do so. Since Adams/Aizawa's necessary conditions forestall such conceptual flexibility from the outset and they have not provided any argument for a cognitive scientist to commit herself to their intransigent and confining conditions, Adams/Aizawa's conditions seem unpalatable.

Adams/Aizawa may hold that one important argument for adopting their conditions is that the conditions help refute HEC. However, first, as mentioned, the conditions do not succeed in accomplishing this task. Second, even if they did preclude extended cognition, this by itself would only represent a scientific benefit if the adoption of HEC entailed some

sort of explanatory loss. Again, Adams/Aizawa have not yet shown, but need to show, that the latter would be the case in order to make their conditions appealing.

VI. 3. 3. Summary

To sum up the last sections, with respect to Adams/Aizawa's first condition – the condition that cognitive processes must always involve non-derived representations – I mentioned that Adams/Aizawa do not hold that cognitive processes need to operate solely on non-derived content. Owing to this qualification, Adams/Aizawa's first condition leaves the door open for HEC. If, to rule out HEC, Adams/Aizawa rigidified their condition so that a cognitive process must not involve any derived representation at all, then the conditions would become unacceptably strong. For it seems intuitively plausible to assume that a process which involves not only non-derived but also interpretation-dependent representations can still qualify as a cognitive process.

With respect to Adams/Aizawa's second condition for cognition – the point that a process needs to exhibit the typical features of human cognition or a family resemblance to human cognition to qualify as cognitive – I argued that it is again either too restrictive or does not suffice to make a case for brain-bound cognition. First, it is too restrictive to demand that any process must exhibit the typical features of human cognition to qualify as cognitive because there are arguably many processes that are still plausibly considered to be cognitive even though they are not characterised by these features. Second, Adams/Aizawa's attempt to account for this problem by holding that a process only needs to bear a family resemblance to human cognition renders their second condition ineffective in preventing extended cognition, for they leave their notion of family resemblance undefined.

Finally, I pointed out that there is a pragmatic reason for not accepting Adams/Aizawa's two conditions for cognition. There may be explanatory benefits from treating processes or resources as cognitive which do not satisfy their conditions. Since the two conditions would block such possible gains and Adams/Aizawa have not given arguments for embracing their restrictive conditions, their conditions remain scientifically unattractive.

Chapter VII: Rupert's critique of HEC

I have argued that Adams/Aizawa fail to provide compelling arguments for the brain-bound view. Robert Rupert (2004, forthcoming b) is another opponent of Clark's HEC. In the following sections and chapters, I will introduce and examine Rupert's challenges to Clark's view and his argument for non-extended cognition.

VII. 1. Rupert's challenges

In his most recent paper on Clark's HEC, Rupert (forthcoming b) mentions a number of criticisms. In what follows, I will focus only on the two arguments that I consider to be his main challenges to HEC. The two arguments concern the practical cognitive-scientific consequences of HEC and the possibility of reducing HEC to a more plausible and more ontologically parsimonious view.

In his first argument, Rupert questions the suitability of extended cognitive systems for cognitive-scientific research. He holds that "organismically bounded cognitive systems play an important role in cognitive science, a role to which the typical extended system is not well suited" (forthcoming b, p. 12).

His point is the following. Cognitive science is interested in explaining the regularities in an agent's cognitive performances across varying situations and across variations in available environmental resources. As an example of such regularities, consider your ability to solve mathematical problems. You can do arithmetic in a variety of situations. And you can do it mentally or with a number of different tools, for instance, with a pocket calculator, with a pen and a piece of paper, with a stick on the sand, with your fingers etc.

Furthermore, you could calculate when you were about 10 and you can still calculate today. In short, with respect to doing arithmetic, you display a consistent performance across different contexts and across variations in available resources.

When cognitive scientists attempt to explain an agent's consistent cognitive performance such as doing arithmetic in different situations and with different objects, they place organismically bounded systems in a privileged position. This is because such a system is integrated and persisting and thus allows for an explanation of why there are consistencies in an agent's behaviour. The common assumption is that an agent's cognitive behaviour is consistent and regular because it is grounded in the agent's persisting and integrated cognitive system (Rupert, forthcoming b, p. 17).

In contrast to the persisting organismically bounded system which traditionally constitutes the subject of cognitive performances and scientific investigations, Clark's transient extended systems[41] "vary greatly in their constitution and, by some obvious measures (e.g., constitution by functionally relevant subparts), do not persist beyond the time of the interaction between organism and external resources" (forthcoming b, p. 12). Since the identity and the capacities of an extended system change whenever the organism interacts with a new set of external resources, "with regard to the phenomena of focal interest in cognitive science, patterns of similarity and difference in behavior across contexts cannot be explained" by extended systems (ibid). They simply do not persist across contexts, and owing to their context-bound, fleeting character, extended systems are unlikely to beget scientifically interesting regularities. In other words, because they are transient in nature,

[41] Note that Rupert only targets actual cases of extended cognition in his critique. That is, he targets TECS, transient extended systems. He does not consider Clark's more permanent instances of cognitive extensions, such as the case of Otto and his notebook. This seems fair enough since, as I have previously mentioned, Clark himself admits that permanent cognitive extensions are currently unlikely to exist.

Clark's extended systems are of little interest in cognitive science. This is Rupert's first argument against adopting HEC.

In his second argument, Rupert holds that the phenomena that inspire HEC can be accounted for by a more plausible hypothesis than HEC. Rupert concedes that Clark's HEC represents one way of conceptualising the agent-environment interactions which result in the accomplishment of a particular cognitive task. HEC conceptualises these interactions in terms of an extended cognitive system. It proposes that in the course of such interactions the external resource constitutes a literal part of the agent's cognitive system. However, Rupert contends that there is another possible and, in fact, preferable alternative to HEC. The cognitive achievements of the coupling between an agent and some external resource that inspire HEC and Clark's postulation of an extended cognitive system can equally well be explained by viewing the external resource in the coupling as a mere tool of the organismically bounded cognitive system. This is the idea underlying Rupert's so-called 'hypothesis of embedded cognition' (HEMC).

According to the embedded cognition view,[42] cognitive activity routinely exploits structures of and is embedded in the environment. One simple illustration of embedded cognition would be one's writing a shopping list by jotting down the most important items to be purchased on the top of the list and the less important ones on the bottom of it. By composing the shopping list in this way, the cognitive work of remembering which of the items on the list is important is off-loaded onto the spatial ordering of words on the list. When the agent requires information about which items have to be purchased more urgently than others, he or she simply looks at the list. The cognitively expensive act of storing information about the items in the head is avoided by the use of external resources.

[42] Rupert is not the only proponent of this view. Other proponents include, for instance, Ballard et al. (1997) and McClamrock (1995).

According to HEMC, since we causally interact with the environment and exploit external objects or structures as scaffoldings to simplify and facilitate cognitive tasks, understanding the cognitive processes involved in managing certain cognitive tasks seems to require not only looking at what goes on in the head, but also at what happens in the interaction between individual and environment. This represents a point that HEMC has in common with HEC. However, even though HEMC proposes that cognitive processes may depend "*very* heavily, in hitherto unexpected ways, on organismically external props," it resists HEC's claim that the human cognitive system literally transgresses the physical boundary of the organism (Rupert 2004, p. 393). HEC and HEMC thus focus on the same phenomena but propose two different metaphysics of cognition.

Rupert holds that between these two different metaphysical pictures, HEMC should be preferred over HEC because it does the same explanatory work as HEC while being (i) more conservative and (ii) ontologically more parsimonious than Clark's hypothesis.[43]

To illustrate (i), according to HEC, there is the internal cognitive system, some external resources which are also considered as cognitive, and interaction between the two. According to HEMC, there is the cognitive system, some external resource and interactions between the two. Since HEMC already takes into account the contribution of the external resource, HEC can be rejected on the "grounds of conservativism: it is an uninteresting position that merely adds the label 'cognitive' to the external resources […]" (Rupert, forthcoming b, p. 14).

But Rupert also claims (ii) that HEC should be replaced on the grounds of simplicity. The reason is that HEMC already explains all that need be explained without postulating an

[43] See Rupert (2004) for a comprehensive explanation and illustration of how HEMC can be substituted for HEC.

additional extended cognitive system. HEC explains cognitively relevant agent-environment interactions by positing the same number of elements as HEMC but "then lumps these parts together under the label 'cognitive system'" (forthcoming b, pp. 14-15). This 'lumping together' of agent and external resources should, Rupert holds, be rejected for reasons of ontological parsimony. That is, if there is no plausible reason to introduce an additional entity, such as an extended cognitive system, doing so should be avoided. It is simply scientifically unhelpful to multiply entities without necessity. Also, merely re-labelling already explained processes as cognitive does not constitute any sort of significant scientific progress. Rupert concludes: "If HEMC accounts for the results that impress advocates of HEC, the more conservative, simpler HEMC wins the day" (forthcoming b, p. 12).

VII. 2. An assessment of Rupert's critique

I turn to Clark's responses to Rupert's arguments. Rupert's first point was that since an extended system is not integrated and persisting, it is unlikely to produce and help explain consistent regularities in an agent's behaviour across contexts. Since these regularities and the system from which they originate are the proper subjects of cognitive science, extended systems are unsuitable subjects of cognitive-scientific research.

In reply, Clark points out that the study of extended cognitive systems is just beginning and it may well be that future research will reveal that some temporary agent-resource interactions do in fact exhibit scientifically interesting regularities across contexts (2008, p. 115). To see whether or not that is the case, further investigations are required. So HEC should tentatively be adopted as a conceptual foundation for cognitive-scientific research.

However, why should the agent-resource interactions which Clark considers to be TECS be interpreted as extended cognitive systems rather than instances in which an agent merely employs some external resource as a tool? The search for scientifically interesting regularities in these interactions might very well be motivated by the assumption that these interactions are instances of tool use. In other words, the argument that there could be interesting regularities in temporary agent-resource interactions gives little reason to adopt HEC, as HEMC is available as an alternative to HEC. Here Rupert's second point against HEC comes into play. Since HEMC does the same explanatory work as HEC while proposing a more frugal metaphysics of cognition, HEMC remains preferable over Clark's hypothesis.

How could Clark respond to Rupert's second argument? A convincing response to Rupert would be if Clark could adduce empirical evidence that supports HEC or mention research that weighs in favours of HEC and against HEMC. However, so far, Clark has not successfully done that.[44] Consequently, Rupert's argument for reducing HEC to HEMC still stands.

[44] Rupert (forthcoming b, pp. 29) shows that the empirical research by Goldin-Meadow (2003), Paul (2006) and Gray et al. (2006) which Clark mentions for preferring HEC over HEMC might very well be seen as supporting HEMC rather than HEC.

Chapter VIII: Rupert's case for non-extended cognition

But Rupert does not only challenge the need to posit extended cognitive systems by proposing HEMC as an alternative way of accounting for the phenomena that inspire HEC. He also provides a positive argument for the claim that cognition is entirely organism-bound. In the following sections, I will introduce Rupert's argument for non-extended cognition and his necessary conditions for cognitive systems. Both are, just as Adams/Aizawa's argument and conditions, very closely related.

His argument for brain-bound cognition is the following. (1) The boundary of the cognitive system constitutes the boundary between causal contributors and genuine components of the cognitive system. (2) The most plausible candidate for the cognitive system is the persisting, integrated architecture. (3) With respect to human beings, this architecture is internally instantiated. Thus, with respect to humans, the cognitive system is internally instantiated (forthcoming b, p. 16).

The argument requires some explanations and justifications. For instance, why is the best candidate for the cognitive system the persisting and integrated architecture? To answer this question, Rupert points to the already mentioned regular patterns in an agent's cognitive behaviour when the agent is placed in a variety of contexts. The most plausible and simplest explanation of these regular patterns is the "reappearance of some organismically bounded system with persisting properties" (forthcoming b, p. 17). If there were no persisting and integrated system these regular patterns "would be a mystery" (forthcoming b, p. 21).

However, the questions of where this integrated architecture is located in the human case and whether or not it includes external resources require further argumentative work and will crucially depend upon how one specifies under which conditions a resource counts as integrated. Thus, to see how Rupert supports the assumption that, in the human case, the persisting, integrated architecture is internally instantiated, an explanation of his notion of systemic integration is required.

VIII. 1 On systemic integration

To determine precisely whether a given resource can qualify as systemically integrated, Rupert suggests the following procedure. He starts by noting that each completion of a cognitive task involves the causal contribution of a number of different mechanisms. So for a given agent at a given point in time, there is a particular set of mechanisms, abilities, external resources etc. which contributes to that agent's cognitive processing in a distinctive way. With respect to a particular cognitive process, each mechanism of the set has a conditional probability of, for instance, the frequency of its use relative to each of the other mechanisms in the set (ibid).

Rupert goes on to suggest identifying and then ranking these conditional probabilities on a list to separate sets with highly interdependent mechanisms from those with less interdependent ones. To illustrate this ranking, let us suppose there are three mechanisms X, Y and Z. All three contribute to an agent's cognitive processing, for instance, to his or her reading a journal article. Mechanism X is linguistic processing, mechanism Y represents short-term memory and mechanism Z is visual processing. Now, suppose that, based on past co-contributions of the mechanisms X, Y and Z, the conditional probability (P) of mechanism Z, given that X and Y are present, equals 0.6. Put formally, P (Z | X & Y)

= 0.6. The set {X, Y, Z} is written down on the ranking list and assigned the conditional probability value of 0.6. Subsequently, in comparison with the conditional probabilities of other sets of mechanisms which also distinctively contribute to cognitive processes, the set {X, Y, Z} will appear at a particular place on the list.

On the list of ranked conditional probabilities, Rupert expects a gap between the higher probabilities and lower ones. This gap separates highly interdependent mechanisms from less interdependent ones. Once the highly interdependent mechanisms with high probability values have been separated from the less interdependent mechanisms with low values, Rupert suggests counting how frequently a specific mechanism occurs on the list of sets with higher probabilities. And again, "a natural cut-off (another significant gap) separates those mechanisms that appear frequently on the list – i.e., are highly interdependent and heavily co-employed – from those that appear rarely" (forthcoming b, p. 23).

Rupert holds that this two-step method of hierarchically ranking the mechanisms that causally contribute to cognitive processing "indicates which mechanisms are parts of the integrated set to be identified with the cognitive system and which are, in contrast, resources used by the cognitive system" (forthcoming b, p. 23).

To summarise Rupert's diagnostic procedure to separate genuine parts of a cognitive system from mere tools of the system, a resource or mechanism constitutes an integrated part of a cognitive system if it satisfies two quite interconnected conditions. First, it has to be highly interdependent upon other mechanisms contributing to a cognitive

phenomenon.[45] Second, it needs to appear frequently in sets of high conditional probability, that is, it has to be frequently employed in different combinations with other mechanisms contributing to cognitive phenomena.

Rupert notes that his method of determining when a resource or mechanism is systemically integrated does not preclude the possibility of extended cognitive systems – if an external resource manages to satisfy the two conditions then it belongs to the cognitive system. But he assumes that for a normal human being, external items will not "appear in sets determined by higher conditional probabilities or, if they do appear, they fail to appear in very many of these, their being dedicated to the solution of specific problems" (forthcoming b, p. 26). Thus, external resources are likely to fail Rupert's two necessary conditions for systemic integration. This is Rupert's argument for non-extended cognition.

VIII. 2. An assessment of Rupert's case for of non-extended cognition

In the following sections, I will take a critical look at Rupert's account of cognitive systems which underlies his argument for non-extended cognition. The problems with his account concern the details of how he defines systemic integration. First, his way of deciding which resource counts as systemically integrated seems arbitrary. Second, delineating cognitive systems in numerical terms appears scientifically unwise.

With respect to the first point, as mentioned, Rupert assumes that on his ranking list which hierarchically orders the interdependency of the mechanisms contributing to cognitive

[45] In order to prevent misunderstandings about what Rupert aims at with his conditions, note that his conditions for cognitive systems presuppose an already given understanding of what constitutes cognitive explananda. Rupert's conditions are, unlike Adams/Aizawa's, not intended to capture the unique feature that distinguishes cognitive from other phenomena. Rather, he assumes that we are already able to pinpoint some behaviour as cognitive. His conditions are only intended to help identify the mechanisms that are part of a cognitive system.

processes, there will be a ‘significant gap’ that separates highly interdependent mechanisms from those mechanisms that are insufficiently interdependent to still qualify as systemically integrated.

However, suppose there is one discernible gap on the list between mechanisms that cluster around the conditional probabilities of 0.1 and 0.3 and another significant gap between mechanisms clustering around 0.3 and 0.5. Now, if we go into the details of Rupert’s account, the question arises of whether only the mechanisms around 0.5 should count as systemically integrated or those around 0.3 as well. Are the mechanisms that cluster around both 0.1 and 0.3 insufficiently interdependent or only those around the value of 0.1? Which of the two gaps in the list is Rupert’s ‘significant gap’ and why is the one gap more significant than the other? These questions are more substantial than they may appear for two reasons: first, the answers to them determine whether or not a given resource constitutes part of a cognitive system and, second, in the light of these questions, it becomes evident that Rupert’s way of defining systemic integration is arbitrary.

The problem of arbitrariness with Rupert’s cognitive-systems delineation reappears if one supposes that there is a continuum between highly integrated and less integrated mechanisms rather than a ‘significant gap’ between the two. Rupert suggests that in such a case, “0.5 would seem to mark the relevant cut-off point” (forthcoming b, p. 23). However, he remains silent on his motivation for choosing this particular value of conditional probability. Why would 0.4 be too low and 0.6 too high a probability value? My straightforward point here is that specifying systemic integration, as Rupert suggests, in numerical terms invites many different, arguably eventually arbitrary demarcations of the cognitive system.

Admittedly, if there is a clear, bi-modal distribution, with a large gap between highly integrated and less integrated mechanisms, Rupert's criterion may after all not be arbitrary. The gap would mark the line between systemically integrated and not systemically integrated mechanisms. Furthermore, it may be argued in support of Rupert that in order to do justice to his approach, one should not focus so much on its fine-grain details. What is more important and should primarily be considered is whether the general direction of the way he accounts for cognitive systems is on the right track.

So let us then suppose that there is a clear bi-modal distribution on Rupert's ranking list and ask whether his account is on the right track. In fact, there is, as with Adams/Aizawa's approach, a pragmatic reason for denying the question. The point is that defining systemic integration and delineating the cognitive system in numerical terms does not only invite arbitrary decisions but is also cognitive-scientifically wrongheaded. Whether or not a mechanism should be considered as part of the cognitive system should not depend upon some specific mathematical value but upon the explanatory benefits of considering it as part of the cognitive system. If it proves scientifically advantageous to consider a mechanism which does not satisfy Rupert's conditions as a component of the cognitive system, then doing so seems to be well justified. After all, conceptual distinctions in empirical sciences are strategic distinctions. What motivates them is their scientific pay-off. Consequently, conceptual flexibility appears scientifically more fruitful than conceptual rigidity which potentially limits scientific gains.

But one may respond that the fact that there *could* be some scientific pay-off in considering a process or resource as part of the cognitive system even if it does not satisfy Rupert's conditions, does not serve as a strong argument against adopting Rupert's conditions. After

all, the argument only relies upon the possible or potential benefits of using the concept of a cognitive system more flexibly.

However, here one needs to remember on whose shoulders the burden of proof rests. To give a drastic analogy to illustrate the point, does an individual have to provide reasons against being incarcerated in order to not be incarcerated? Or is it rather the task of those who wish to incarcerate the individual to deliver persuasive reasons for incarcerating him or her? Quite obviously, the latter is the case.

Similarly, Rupert has the burden of proof to show why a cognitive psychologist who currently uses the concept of a cognitive system flexibly should confine herself to his conditions. Rupert needs to mention some significant benefits from committing oneself to his conditions. Until now, he has not done so.

To avoid gratuitous theoretical or conceptual obstructions to cognitive-scientific research, it is more appropriate and plausible to abandon the idea of a necessary condition of the cognitive, such as a specific numerical value of systemic integration altogether, and leave the question of when a mechanism or process qualifies as cognitive and as a genuine part of a cognitive system open. In the following sections, I will investigate what such a move would imply for the case for extended cognition.

Chapter IX: (Un)necessary conditions - Why adopt HEC?

Would abandoning necessary conditions for cognition play into the hands of the advocate of extended cognition? Necessary conditions for cognition rule that unless a process or resource has property *x*, it cannot qualify as cognitive. In the absence of necessary conditions, there is no longer a theoretical justification for denying that a particular resource or process can qualify as part of the cognitive system. That is, without necessary conditions, there is, correspondingly, also no way to preclude extended cognition.

But, as I have just mentioned with respect to Rupert, abandoning the attempt to strictly delineate cognition by a set of necessary conditions does not automatically mean that cognition extends. More argumentative work needs to be carried out to justify such a conclusion. For a convincing case for extended cognition, Clark needs to (i) propose tenable sufficient conditions for counting something as part of the cognitive system and (ii) show that these conditions are met by some contemporarily employed external resources.

However, as shown above, Clark already fails in accomplishing point (i). The 'glue and trust' conditions and the parity principle, which represent Clark's sufficient conditions, are not plausible. As a result, since his conditions can be repudiated, Clark's case for extended cognition does not succeed either.

But let us for the sake of argument put aside the issue that Clark has not yet provided plausible sufficient conditions according to which HEC comes out true and let us ask whether there are more pragmatic, cognitive-scientific benefits from adopting HEC. Clark claims that cognitive science would be impoverished if HEC is not adopted. The "costs" of not accepting the extended cognition view are, he contends, "great indeed" (Clark/Wilson,

2009, p. 73). As an apparent reason for this assertion, he writes that extended cognitive systems are "quite literally, the soft-assembled circuitry of a great deal of practical human thought and reason. We ignore or downplay the importance of these ensembles, treating them as merely ersatz cognitive circuitry, at our theoretical peril […]" (ibid).

However, Clark fails to elucidate just what exactly this 'theoretical peril' might be. He arguably assumes that if HEC is not accepted, cognitive science and psychology will disregard the role that environmental resources play for an agent's ability to accomplish certain cognitive tasks. An important feature of human cognition would be ignored and left unexplored.

But, contra Clark, impugning the metaphysics of HEC does not mean rejecting, 'ignoring' or 'downplaying' the cognitive significance of the agent-environment interactions which Clark interprets as extended cognitive systems. The above mentioned alternative to HEC, HEMC, can retain the view that such interactions are cognitive-scientifically important and should be investigated. In fact, as Rupert writes, it is HEMC's "expressed agenda […] that cognitive science focus on ways in which the human cognitive system interacts with and exploits external resources" (forthcoming b, p. 28).

However, not even a commitment to HEMC may be needed to motivate psychological or cognitive-scientific research on agent-environment interactions. Rupert notes that "the empirical research taken to support HEC was motivated not by a specific commitment to HEC *or* to HEMC, but rather by a general sense that interaction with the environment plays an important role in cognitive processing" (ibid). For instance, in a paper which Clark repeatedly quotes in support of HEC (see, for instance, 2008, pp. 11, p. 69, p. 201), Dana Ballard et al. (1997) write the following about their research motivation: "Our central

thesis is that intelligence has to relate to interactions with the physical world, meaning that the particular form of the human body is a vital constraint in delimiting many aspects of intelligent behavior" (p. 723). As Rupert remarks, "This thesis entails neither HEC nor HEMC, yet it captures the approach of some of the most influential empirical work supposed to support HEC" (forthcoming b, p. 28). In other words, since there already exists a cognitive-scientific interest in agent-environment interactions, Clark's metaphysical picture and flamboyant rhetoric of extended cognitive systems are not required to highlight the important fact that these interactions deserve scientific attention.

If HEC is not needed to initiate some sort of reorientation in cognitive science, could there be other scientific benefits from conceiving of external resources as literal parts (HEC) rather than mere tools of an agent's cognitive system (HEMC)? It seems that this question can be denied since, scientifically, the differences between HEC and HEMC are quite negligible. Whether an external resource is interpreted as a literal part of a cognitive system or merely as a tool that the system employs to solve a cognitive problem makes little difference with respect to what is empirically observable in the problem-solving performance. Whether viewed through the 'lenses' of HEC or HEMC, what is observable remains the same: an agent, an external resource and the interactions between the two. Consequently, a cognitive scientist could flip back and forth between HEC and HEMC without modifying her empirical work (Sprevak, 2008, p. 21). Since HEC and HEMC are merely two different ways of conceptualising the same empirical phenomena, for a scientist, the question whether HEC or HEMC should be adopted seems an extraneous matter.

However, there could be the following advantage of adopting HEC rather than HEMC. HEC may offer a more attractive taxonomy of cognitive processes than HEMC and should

thus be given preference. Chalmers (2008) claims that "explanatory unification" constitutes "the real underlying point" and benefit of HEC (p. xiv).

The idea seems to be that HEC allows for the explanatory unification of extended processes with internal cognitive processes and can thereby offer simpler and shorter explanations of behaviour than HEMC's or traditional cognitive-scientific explanations. For instance, HEC would explain Otto's walking to 53rd Street by proposing that he had the desire to go to the MoMA and believed that it was in this location. In contrast, the traditional, internalist view of cognition would "explain Otto's action in terms of his occurrent desire to go to the museum, his standing belief that the Museum is on the location written in the notebook, and the accessible fact that the notebook says the Museum is on 53rd […]" (Clark, 1998, p. 13). Clark holds that this explanation is "pointlessly complex, in the same way that it would be pointlessly complex to explain Inga's actions in terms of beliefs about her memory […]" (ibid). Since in explanations of Inga's actions no belief about her memory is involved, analogously, in HEC's explanations of Otto's actions no belief about his notebook is involved either. By explanatorily unifying Otto's notebook interactions with Inga's memory retrieval, HEC yields a shorter and simpler explanation. And since in "an explanation, simplicity is power," Clark claims that this is an advantage of HEC (ibid).

However, the first aspect to note about Clark's argument is that it effectively only concerns the number of words in an explanation. But even if we follow Clark and consider this terminological issue as non-trivial, there is still a reason for taking exception to his argument for adopting HEC. The reason is that a longer explanation that includes an extra step in accounting for Otto's behaviour remains well justified and, in fact, more accurate than HEC's explanation of Otto's actions. Otto would not look in his notebook for the

address of the museum if he did not believe that it stores this information. The longer and more complex account which involves the extra step of Otto's beliefs about his notebook does not only explain what HEC explains, namely why Otto decides to walk to 53rd Street. It also explains, unlike HEC, why Otto decides to look up his notebook. The longer account explains all the actions that actually occur, whereas HEC only accounts for Otto's going to 53rd Street. Thus, even though the traditional cognitive-scientific account is longer and more complex than HEC's explanation, it is more accurate in explaining the actions that lead Otto to his trip to the MoMA (Shapiro, 2008, p. 26). Clark's argument that HEC should be adopted since 'simplicity is power,' is thus not persuasive. If simplicity is only possible at the expense of accuracy, it is surely not preferable over longer explanations.

But, if Clark's explanatory-simplicity argument does not yield an actual benefit of HEC, then, again, what are the benefits of embracing HEC? Since Clark does not answer this question satisfactorily, there seems little reason to adopt the view that cognition sometimes extends. Thus, even if no plausible necessary condition for cognition may be available for *rejecting* the view, there are simply no cogent arguments for *endorsing* it either.

IX. 1. Summary and conclusion

In the following sections, I will summarise my discussion and draw a conclusion. My discussion concerned two different answers to the question of where cognition is located. Clark purports that cognition sometimes extends into parts of the environment. As mentioned, to make a persuasive case for his view, Clark needs to provide plausible sufficient conditions for classifying a resource or process as part of a cognitive system and then establish that currently available external resources satisfy his conditions. In contrast, Adams/Aizawa and Rupert argue that cognition is entirely brain-bound. To make their case,

Adams/Aizawa and Rupert need to offer plausible necessary conditions for cognition and then demonstrate that no currently available external resources satisfy them.

In the course of my discussion of Clark's extended and Adams/Aizawa's and Rupert's brain-bound views, I offered reasons for rejecting both Clark's sufficient and Adams/Aizawa's necessary conditions for cognition. I discouraged acceptance of either view on the location of cognition.

As illustrated, one reason for demurring at Clark's hypothesis of extended cognition is that the sufficient conditions that he proposes are unacceptable. Both conditions – the 'glue and trust' conditions as well as the parity principle – lead to a counterintuitive inflation of cognition (see V. 3 and V. 4). This represents the first reason for resisting Clark's extended cognition view. A second reason constitutes the fact that Clark's HEC can, without explanatory loss, be supplanted by the more plausible and ontologically more parsimonious hypothesis of embedded cognition (HEMC) (see VII. 1-2). Thus, in my discussion, two points operate together to repudiate Clark's extended cognition view.

After explicating reasons for opposing the extended cognition view, I turned to an investigation of whether Adams/Aizawa's and Rupert's internalist proposals would be more plausible and acceptable. With respect to Adams/Aizawa, I argued that their necessary conditions for cognition are unpalatably restrictive and in fact insufficient to establish that cognition is brain-bound (see VI. 3. 1.). With regard to Rupert's conditions for cognitive systems, I argued that his way of drawing the boundaries of the cognitive system appears arbitrary and, just as Adams/Aizawa's conditions, scientifically unwise (see VIII. 2.).

In a final step of my discussion, I raised the question of whether it would help the extended cognition proponent if Adams/Aizawa's and Rupert's necessary conditions for cognition were abandoned. I argued that not being able to say that cognition occurs only in the head does not mean to say that cognition extends. In the absence of necessary conditions which preclude extended cognition, the advocate of HEC still needs to present reasons for assuming that cognition extends. However, as it turns out, Clark does not succeed in providing such reasons. He fails to elucidate why HEC should not be supplanted by the explanatorily equivalent but more ontological frugal HEMC. Since that is the case, HEMC remains, as the common-sense and default position on the location of cognition, more plausible than HEC. HEMC wins over HEC. That is – to come back to the question introducing this thesis –, even though cognition may depend very heavily upon external features, there is no persuasive reason to assume it literally "leaks out" into the world. (2008, p. xxviii)

What about the court case of Michael T. Arnold that I mentioned at the beginning of the thesis? Was the officer at the Los Angeles Airport Customs not allowed to inspect Arnold's computer hard drive because it constituted a literal part of Arnold's mind? Do electronic storage devices such as computer hard drives constitute a physical "extension of our own memory,"[46] as Judge Pregerson holds? Maybe, maybe not. The ball is in Judge Pregerson's and Clark's court.

[46] Liptak (2008)

References

Adams, Frederick and Aizawa, Kenneth (2008). "*The Bounds of Cognition.*" Malden, MA: Blackwell Publishing.

Adams, Frederick and Aizawa, Kenneth (forthcoming). "Defending the Bounds of Cognition." In: Menary, R., (ed). *The Extended Mind*, Aldershot, Hants: Ashgate. Retrieved on the 20.04.2009 from http://philpapers.org/asearch.pl?strict=1&searchStr=Adams, Frederick R.&filterMode=authors

Baars, B. J. (1986). *The Cognitive Revolution in Psychology*. New York: Guilford.

Ballard, Dana et al. (1997). "Deictic Codes for the Embodiment of Cognition." In: *Behavioural and Brain Sciences*, 20.

Block, Ned (2005). "Review of Alva Noë's *Action in Perception*." In: *Journal of Philosophy,* 102.

Braddon-Mittchell, David and Jackson, Frank (2007). *The Philosophy of Mind and Cognition.* Malden, MA: Blackwell Publishing.

Chalmers, David (2008). "Foreword to Andy Clark's *Supersizing the Mind.*" In: Clark, Andy, *Supersizing the Mind. Embodiment, Action, and Cognitive Extension.* New York: Oxford University Press.

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

Clark, Andy (2001). *Mindware*. New York: Oxford University Press.

Clark, Andy (2001b). "Reasons, Robots, and the Extended Mind." In: *Mind and Language*, 16.

Clark, Andy (2007). “Re-inventing Ourselves: The Plasticity of Embodiment, Sensing, and Mind.” In: *Journal of Medicine and Philosophy*, 32, 3.

Clark, Andy (2008). *Supersizing the Mind. Embodiment, Action, and Cognitive Extension.* New York: Oxford University Press.

Clark, Andy and Chalmers, David (1998). “The Extended Mind.” In: *Analysis*, 58, 1.

Clark, Andy and Wilson, Robert (2009). “How to Situate Cognition: Letting Nature Take its Course.” In: Aydede, M. and Robbins, P. (eds.), *The Cambridge Handbook of Situated Cognition.* Cambridge: Cambridge University Press.

Dennett, Daniel C. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.

Dennett, Daniel C. (2003). Presentation: On our Consciousness. Retrieved on the 20.06. 2009 from http://www.ted.com/index.php/talks/dan_dennett_on_our_consciousness.html

Fisher, Justin C. (2008). “Critical Notice: *The Bounds of Cognition*. Frederick Adams and Kenneth Aizawa.” In: *The Journal of Mind and Behavior*, 29, 4.

Gazzaniga, Michael, S. (2004). *The Cognitive Neurosciences III*. Cambridge, MA: MIT Press.

Gertler, Brie (2007). “Overextending the Mind?” In: Gertler, Brie and Shapiro, Lawrence (ed.), *Arguing about the Mind*. New York: Routledge. Retrieved on the 24.04.2009 from http://people.virginia.edu/~bg8y/

Goldin-Meadow, Susan (2003). *Hearing Gesture: How Our Hands Help Us Think.* Cambridge, MA: Harvard University Press.

Gray, W. D. et al. (2006). “The Soft Constraints Hypothesis: A Rational Analysis Approach to Resource Allocation for Interactive Behavior.” In: *Psychological Review*, 113, 3.

Green, David W. et al. (1996). *Cognition. An Introduction*. Oxford: Blackwell Publishers.

Grush, Rick (2001). "The Philosophy of Cognitive Science." Retrieved on the 21.05.2009 from http://mind.ucsd.edu/papers/phil_cogsci/phil_cogsci.html

Hampson, Peter and Morris, Peters E. (1996). *Understanding Cognition*. Cambridge, MA: Blackwell Publishers.

Haugeland, John (1998). *Having Thought: Essays in the Metaphysics of Mind*. Cambridge, MA: Harvard University Press.

Horgan, Terry and Kriegel, Uriah (2008). "Phenomenal Intentionality Meets the Extended Mind." In: *The Monist*, 91.

Howieson, D. et al. (2004). "Neurobehavioral Variables and Diagnostic Issues." In: Lezak, M. D. et al. (eds.), *Neuropsychological Assessment*. New York: Oxford University Press.

Kandel, E.R. et al. (2000). *Principles of Neural Science*. New York: McGraw-Hill

Kandel, E.R. et al. (1995). *Essentials of Neural Science and Behavior*. Norwalk, CT: Appleton & Lange.

Kolak, D. et al. (2006). *Cognitive Science. An Introduction to Mind and Brain*. New York: Routledge.

Kosslyn, S. M. (1986). "Toward a Computational Neuropsychology of High-Level Vision." In: Knapp, T.J. and Robertson, L. C. (eds.), *Approaches to Cognition: Contrasts and Controversies.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Levin, J. (2004). "Functionalism." In: *The Stanford Encyclopaedia of Philosophy*. Retrieved on the 12.05.2009 from http://plato.stanford.edu/entries/functionalism/

Liptak, Adam (2008). "If Your Hard Drive Could Testify ..." In: *New York Times*, January 7, 2008. Retrieved on the 16.02.2009 from http://www.nytimes.com/2008/01/07/us/07bar.html?_r=1&pagewanted=print

McClamrock, Ron (1995). *Existential Cognition: Computational Minds in the World*. Chicago: University of Chicago Press.

Millikan, Ruth (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Miller, George A. (2003). "The Cognitive Revolution: A Historical Perspective." In: *Trends in Cognitive Sciences*, 7, 3.

Noë, Alva (2007). "Magic Realism and the Limits of Intelligibility: What makes us Conscious." In: *Philosophical Perspectives*, 21.

Normand, Emmanuelle et al. (2009). "Forest Chimpanzees (*Pan troglodytes verus*) Remember the Location of Numerous Fruit Trees." In: *Animal Cognition*, (online: May 31, 2009).

Paul, C. (2006). "Morphological computation: A Basis for the Analysis of Morphology and Control Requirements." In: *Robotics and Autonomous Systems*, 54.

Plato (1920). *Phaedo*. *The Dialogues of Plato*. (trans.) Jowett, B., New York: Oxford University Press.

Rensink, Ronald et al. (1997). "To Se Or Not To See: The Need for Attention to Perceive Changes in Scenes." In: *Psychological Science*, 368, vol. 8, 5.

Rupert, Robert (2004). "Challenges to the Hypothesis of Extended Cognition." In: *Journal of Philosophy*, 98.

Rupert, Robert (forthcoming). "Systems, Functions, and Intrinsic Natures: On Adams and Aizawa's *The Bounds of Cognition*." Retrieved on the 18.04.2009 from http://spot.colorado.edu/~rupertr/Adams_Aiz_Review_Rupert.pdf

Rupert, Robert (forthcoming b). "Keeping HEC in CHEC: On the Priority of Cognitive Systems." Retrieved on the 20.04.2009 from

http://spot.colorado.edu/~rupertr/HECinCHEC.pdf

Segal, Gabriel (1997). “Review of Robert A. Wilson, Cartesian Psychology and Physical Minds: Individualism and the Sciences of Mind.” In: *British Journal for the Philosophy of Science*, 48, 1.

Shapiro, Lawrence A. (2008). “Functionalism and Mental Boundaries.” In: Marsh, L. and Onof, C. (eds.), *Perspectives on Social Cognition*, a special issue of *Cognitive Systems Research* 9, 2008. Retrieved on the 12.02.2009 from http://philosophy.wisc.edu/shapiro/HomePage/FMB.pdf

Sprevak, Mark (forthcoming). “Extended Cognition and Functionalism.” In: *The Journal of Philosophy* (forthcoming). Retrieved on the 16.06.2009 from http://people.pwf.cam.ac.uk/mds26/files/Sprevak---Extended%20Cognition.pdf

Sterelny, Kim (2004). “Externalism, Epistemic Artifacts, and the Extended Mind.“ In: Schantz, R. (ed.). *The Externalist Challenge*. Berlin: Walter de Gruyter.

Sterelny, Kim (2003). *Thought in a Hostile World. The Evolution of Human Cognition*. Oxford: Blackwell Publishing.

Thatcher, Robert W. et al. (ed.) (1994). *Functional Neuroimaging: Technical Foundations.* San Diego, California: Academic Press.

Uttal, William R. (2001). *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain.* Cambridge, MA: MIT Press.

VandenBos, Gary R. (2007). *APA Dictionary of Psychology*. Washington DC: American Psychological Association.

Van Gulick, Robert (2009). “Functionalism.” In: McLaughlin, Brian P. et al. (eds.), *The Oxford Handbook of Philosophy of Mind.* Oxford: Oxford University Press.

Weiskopf, Daniel (2008). “Patrolling the Mind’s Boundaries.” In: *Erkenntnis*, 68.

Wilson, Margaret (2002). “Six Views of Embodied Cognition.” In: *Psychonomic Bulletin & Review*, 9, 4.

Wilson, Robert (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences*. Cambridge: Cambridge University Press.

http://homepages.nyu.edu/~jdc372/Coursenotes/Functionalism.htm