Location, location, location:
On the metaphysical commitments of the extended mind thesis

K. Vergobbi Vold

Department of Philosophy
McGill University
June 2017

ABSTRACT

While the standard view in the brain sciences is that the human mind is (in some sense) a product of operations in the brain, the extended mind thesis, a popular emerging view in the philosophy of cognitive science, challenges this, instead claiming that objects located in an agent's environment, beyond the brain and body, can serve as partially constitutive of the agent's mental states. Thus, the extended mind thesis advocates a non-traditional understanding of the human mind and cognition. There is a common presumption that the extended mind thesis relies on certain metaphysical commitments, including a functionalist view about the nature of the mind. This is unsurprising given that the most influential argument for the view, due to Andy Clark and David Chalmers, does appeal to functionalism along with the claim that there can be 'functional parity' between internal and external resources.

The main purpose of this thesis is to explore how Clark and Chalmers's extended mind thesis interacts with functionalism, a pillar of 20th century metaphysics of mind. Functionalist-related debates, including objections to both Clark and Chalmers's appeal to functionalism and their claim that there can be functional parity between internal and external objects, have dominated the discussion of the merits of the extended mind thesis. My goal is to show that the core principle of their argument for the thesis does not depend on the metaphysical commitments that it is widely thought to do.

The central argument I develop in this thesis severs the allegedly tight connection between the thesis and functionalism, thereby insulating it from nearly all the standard objections. The crux of the argument I make turns on the distinction between the multiple realizability thesis, another pillar of 20th century philosophy of mind, and a thesis I refer to as multiple localizability. I argue that the extended mind thesis is about the *location* of mental states rather than their *composition* – and thus depends only on a commitment to multiple localizability rather than multiple realizability or functionalism. Once it is stripped of its relationship to functionalism, the extended mind thesis can be evaluated on its own merits. I conclude that the seemingly radical extended mind thesis is fact quite conservative and thus very plausible.

RÉSUMÉ

Il est communément admis dans les sciences du cerveau que l'esprit humain est (en quelque sorte) le produit de l'activité du cerveau. La thèse de l'esprit étendu s'oppose à ce point de vue en affirmant que des objets situés dans l'environnement d'un sujet, au-delà du cerveau et du corps humain, peuvent être des parties constitutives de ses états mentaux. Cette thèse préconise donc une compréhension non traditionnelle de l'esprit humain et de la cognition. On présuppose souvent que la thèse de l'esprit étendu repose sur certains engagements métaphysiques, notamment une vision fonctionnaliste de la nature de l'esprit. Cela n'est pas surprenant, car l'argument le plus influent appuyant cette perspective, dû à Andy Clark et à David Chalmers, fait appel au fonctionnalisme et affirme qu'il peut exister une « parité fonctionnelle » entre les ressources internes et externes.

Le but principal de cette thèse de doctorat est d'explorer la façon dont la thèse de l'esprit étendu de Clark et Chalmers interagit avec le fonctionnalisme, un pilier de la métaphysique de l'esprit du 20ᵉ siècle. Les débats à propos du fonctionnalisme, et notamment les objections au recours de Clark et Chalmers au fonctionnalisme et à leur affirmation selon laquelle il peut y avoir une parité fonctionnelle entre les objets internes et externes, ont dominé les discussions sur les mérites de la thèse de l'esprit étendu. Mon objectif est de démontrer que le principe fondamental de leur argument pour la thèse ne repose pas sur des engagements métaphysiques, contrairement à ce que l'on pourrait croire.

L'argument central que je développe dans cette thèse rompt le lien prétendument nécessaire entre la thèse de l'esprit étendu et le fonctionnalisme, le mettant ainsi à l'abri de presque toutes les objections habituelles. Au cœur de cet argument est la distinction entre la *réalisation multiple*, un autre pilier de la philosophie de l'esprit du 20ᵉ siècle, et la *localisation multiple*. Je soutiens que la thèse de l'esprit étendu concerne la *localisation* des états mentaux plutôt que leur *composition*; et donc, ne repose que sur un engagement à l'égard d'une localisation multiple, mais ne demande pas de s'engager dans la réalisation multiple ou le fonctionnalisme. Une fois le lien avec le fonctionnalisme rompu, la thèse de l'esprit étendu peut être évaluée en fonction de ses qualités intrinsèques. Je montre ainsi à quel point la thèse de l'esprit étendu, au premier abord extrême, s'avère finalement plausible et, par conséquent, difficilement contestable.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

INTRODUCTION

0.1 Preface

The extended mind thesis is a popular emerging view in contemporary philosophy of mind and philosophy of cognitive science. It is one part of a larger push to see the mind as involving more than just the operations and states of the brain. This movement is known as 'situated cognition' (Robbins and Aydede 2009). While traditional cognitive science and neuroscience consider the brain to be the sole machinery of the mind, situated views of cognition reject this 'intracranialist' commitment and instead maintain that mental states and processes can be partially constituted by mechanisms beyond one's brain. The goal of this thesis is to elucidate the central metaphysical commitments of what is often considered to be the most radical of these situated views—the extended mind thesis. I begin the first chapter with a discussion of the standard argument for the extended mind thesis, which I then go on to deconstruct in subsequent chapters. To set the stage for this, in this introduction I give an overview of the commitments of traditional cognitive science and consider why and how situated views of cognition reject this picture. I survey some of the scientific and philosophical antecedents of situated views as well as its main varieties, including embodied, embedded, extended, enactive, and distributed views of the mind, with the purpose of explaining how the extended mind thesis relates to and differs from these other situated views. I then discuss the aims and contributions of this thesis, and end with a summary of the chapters.

0.2 Traditional cognitive science

Cognitive science is still a relatively young field, sitting at the intersection of neuroscience, computer science, linguistics, anthropology, psychology, and philosophy.

1

The field was born out of the so-called 'cognitive revolution', which rejected the dominant view in psychology at the time, behaviorism (Bechtel and Graham 1999; Thagard 1992, 1996). Behaviorism, as an explanatory project, tries to explain all of our mental life in terms of our outwardly observable behaviors. As a metaphysical view, behaviorism reduces all of our mental activity to outwardly observable behaviors (see Skinner 1977). But, while rejecting Descartes's dualist views about an immaterial soul, behaviorism still inherited the Cartesian view of the body as a machine whose operations were dependent on external stimuli and whose behavior could be predicted on the basis of its history of reinforcement, without any need to consider internal psychological mechanisms. Because the operations of the brain are not outwardly observable, behaviorists were charged with ignoring their contributions to our mental life. Being concerned only with outwardly observable behaviors forced behaviorists to treat the mind as a black-box, the contents of which could be ignored in explanations of mental functioning. Chomsky (1959) most famously rejected behaviorism on just these grounds in his critical review of B.F. Skinner's (1957) behaviorist account of language acquisition. His review is widely seen as having discredited behaviorism as a theory of mind, and cognitive science has since filled the vacuum left by the rejection of behaviorism.

In addition to rejecting the behaviorist paradigm, cognitive science established its own paradigm. While improving upon the framework that preceded it, the science that emerged from the cognitive revolution still preserves much of the Cartesian doctrine of the mind. In this introduction I will discuss several commitments of traditional cognitive science that preserve Cartesianism and that have been the target of attacks by a new group of revolutionaries—situated cognitive theorists (Rowlands 2010). These commitments are: (a) scientific reductionism; (b) representationalism and computationalism; and (c) intracranialism. The final view—intracranalism—is the main target of this thesis.

(a) Scientific reductionism: Traditional cognitive science inherited the

reductionist scientific methodology, advocated by Descartes, which studies the world by breaking it down into smaller parts and then trying to understand how those parts work. This approach assumes that parts can be understood independently from each other and from the whole within which they exist (Clancey 2009). By adopting this methodology in its study of the mind, cognitive science treats the mind as a distinct part, or entity, that can be completely understood on its own, independently from the body or the world that it is a part of, or at least intimately connected to. This methodology, coupled with other theoretical commitments, (including intracranialism, to be discussed below) contributes to cognitive scientists' treatment of the body and the wider world as peripheral to an understanding of mental functioning.

(b) Representationalism and Computationalism: Two more central commitments of traditional cognitive science are to the representational theory of mind and to the computational theory of mind. Representationalism, in this context, maintains that mental states are representational, that is, they are information-bearing structures (or symbols) that represent external reality (Fodor 1980). The representational theory of mind is also a Cartesian view: Descartes maintained that much or all of our thought is representational.[1] The computational theory of mind maintains that mental processes are the manipulation of these mental representations in accordance with a set of formal (syntactic) rules (Fodor 1980). Both of these views contribute to yet another commitment that characterizes traditional cognitive science: a commitment to some version of machine functionalism, the view that mental states are states of a Turing machine (Putnam 1960, 1967). (I will explain the functionalist theory of mind, and the machine functionalist version in particular, in some detail in the second chapter.)

(c) Intracranialism: Traditional cognitive science subscribes to physicalism, the view that minds are physical and that the only existing substance is physical. While

---

[1] In fact, the representational theory of mind predates Descartes, going back to at least Aristotle (Pitt 2017). There is debate over how much of our thinking Descartes held to involve representations, e.g. whether this is true of our sensations or passions (1641, Third Meditation).

rejecting dualism as a metaphysical position, cognitive science still held on to the Cartesian idea that the mind is, in some sense, distinct from the body and the world it inhabits in that it can be disassociated or separated from these external factors, at least for the purpose of explanation. In particular, cognitive science subscribes to the view that, whatever else is true of cognitive processes, they take place in the head, in particular in the brain, which is the organ of the mind. This view maintains that all of the *vehicles*, i.e. the information bearing structures, of our mental representations are within the skull. In other words, mental representations are encoded in the brain and the relevant computations over representations, or cognitive processes, all take place in the brain (Rowlands 2010). This position, known as vehicle internalism, should be distinguished from content internalism, the view that the *content* of our mental representations are fixed, or determined, by properties of our brains.[2] The vehicle internalist/externalist distinction varies independently of the content internalist/externalist positions (Chalmers 2008; Theiner 2011). Thus, to keep these two versions of internalism clearly distinguished I will hereafter refer to vehicle internalism as 'intracranialism'.[3]

Intracranialism makes room for the possibility that a humanlike mind could exist in a non-humanlike body. For example, a humanlike mind could exist as a brain floating in a vat, since all our cognitive processes take place in the brain (Rowlands 2010; Shapiro 2011). This view, known as the 'separability thesis', is compatible with the other commitments of traditional cognitive science that maintain that the body only plays a supporting role in bringing about cognition (Shapiro 2011). Thus, while cognitive processes may lean heavily (i.e. be causally dependent) on the non-neural body and the extra-bodily environment, cognitive processes are not, on the traditional view, constituted by the extra-neural body or by anything beyond the agent's brain.

---

[2] See Putnam (1975) and Burge (1979) for arguments against content internalism.
[3] Intracranialism is the view that Clark (2008b: xxviii) calls 'BRAINBOUND' (his capitalization).

0.3 The emergence of situated views of cognition

In recent years within cognitive science new revolutionary winds have been blowing. Many feel, and indeed some have felt all along, that the new science did not go far enough in its rejection of Cartesianism. The new revolutionaries defend what they call a 'situated' view of cognition that rejects key commitments of the traditional doctrine. The seeds of this new revolution have been germinating for some time. Situated views have both scientific and philosophical antecedents that contributed to their development.[4] Many situated theorists take inspiration from discoveries and ideas in a range of scientific fields, including passive dynamics research in robotics, the concept of feedback in cybernetics, work on cognitive scaffolding and ecological views of perception in psychology, the notion of autopoiesis in biology, and others (Gallagher 2009). Situated theorists also take inspiration from the work of twentieth-century philosophers, such as Heidegger (1927), Wittgenstein (1953), and Merleau-Ponty (1962), who responded critically to the early modern philosophies of Descartes and Kant (Gallagher 2009; Wilson and Clark 2009) I will briefly survey some of the views that have, for some time, questioned the key commitments of the traditional doctrine, thereby paving the way for situated views of cognition to emerge. In doing so, I will also introduce the four main situated views of cognition—(i) embedded cognition, (ii) embodied cognition, (iii) enactivism, and (iv) extended cognition[5]—and explain how they challenge these core commitments.[6] As we will see, each situated view of cognition questions some aspect of traditional cognitive science, but not every situated view rejects the traditional doctrine wholesale.

---

[4] See Clancey (2009) and Gallagher (2009).

[5] I will use 'extended cognition' and the 'extended mind thesis' interchangeably throughout the thesis with the exception of a few sections where I will explicitly address the distinction.

[6] More detailed surveys are available in Robbins and Aydede (2009); Rowlands (2010); and Shapiro (2011).

0.3.1 Rejection of scientific reductionism

A key concept that cuts across various sciences and that eventually manifested itself in situated views of cognition, is 'systems thinking' (Clancey 2009). This idea can be seen as a rejection of scientific reductionism. Systems thinking instead supports scientific holism, which advocates a holistic methodology. Scientific holists maintain that in order to fully understand the world we need to understand how the parts relate to each other, or work together, and how they work in relation to the whole (Esfeld 2001). Systems thinking looks to understand the causal relations among the parts of a system and the emergent processes that arise from the interactions of these parts, in order to give a full explanation of a system of whatever kind, whether it be naturally occurring or artificial (Clancey 2009). Systems thinking, for example, views human cognition as developing and occurring within a wider system involving biological, social, and cultural practices. The holistic approach emerged in a number of different scientific disciplines, including cybernetics and robotics.

Pre-dating cognitive science was cybernetics, which studies feedback and control of systems, both artificial and biological. Its founding father, Norbert Weiner, was a mathematician working in wartime who became especially interested in goal-directed systems, especially gun-control systems (Eliasmith 2009). Weiner worked on what is now called 'classical control theory', which aims to design a control system that can alter its own inputs in order to achieve its desired output (Clancey 2009). Weiner worked on two kinds of control systems: open-loop control and closed-loop control. In the former, the system controller provides inputs that, under normal circumstances, should achieve the system's desired outputs. The weakness in open control systems is that circumstances are often changing and so-called normal circumstances are often hard to define (Mindell 2002; Clancey 2009). Closed-loop systems offer a more sophisticated kind of control based on 'feedback', where the inputs provided to the system depend on their internal state and current outputs. Drawing from these ideas in cybernetics, situated cognitive

theorists often emphasize the role of feedback in the control of cognitive systems (Eliasmith 2009; Clancey 2009).

Another example of systems thinking in the sciences that inspired situated views of cognition comes from work in robotics (Clancey 2009). Because of its commitment to machine functionalism, cognitive science has always been open to the possibility of artificial intelligence (AI; more on this in the next chapter). But AI programmers and robotics engineers ran into several key problems in their machine designs. For example, on its website, Honda calls its robot known as ASIMO (Advanced Step in Innovative Mobility) the world's most advanced humanoid robot[7], yet it consumes inhuman levels of energy in order to perform basic human tasks, such climbing stairs. And it performs these tasks with very un-humanlike movements (Clark 2008b). ASIMO had been built with a scientific reductionist methodology: it was made to be a freestanding entity that relies entirely on internal representations of the external world and on an internal energy supply.

By adopting a holistic systems thinking approach, roboticists observed that humans rely on a fundamentally different kind of processing, which explains the vast difference in energy efficiency. Humans rely on 'passive dynamics': our dynamical behaviors make maximal use of external resources, including information and energy, rather than relying entirely on internal resources (Clark 2008b; Shapiro 2011). The slinky toy provides a primary example of passive dynamics in action. The toy is simply a large spring: it has no complex inner workings, no internal energy supply, and, arguably, no internal representations. Yet, the slinky is able to make its way down a set of stairs in a fairly reliable and smooth way by depending on external forces such as gravity and friction (Collins, Wisse, and Ruina 2001; Shapiro 2011). Humans too rely on these sorts of external forces, only in more complicated ways. These insights from passive dynamics informed engineers who built robots that can now perform motor tasks with much more

---

[7] As on Honda's website for ASIMO: <http://asimo.honda.com>. (American Honda Motor Co., 2017.)

human-like movements and at comparable energy consumption rates (Collins and Ruina 2005; Clark 2008b).

Holistic approaches in various scientific fields provided insights that laid the groundwork for the situated view of cognition to emerge. Situated views advocate a holistic approach to understand human cognition. One of the earliest situated views that emerged from this approach was the embedded theory of mind, which maintains that the functioning of a cognitive system depends crucially on states of its local environment. One advocate of this view, Herbert Simon (1969), for example, argues that much of the apparent complexity of cognitive systems is actually external to the agent, residing in the environment. Cognitive systems lean heavily on this complexity without needing to internalize it. Simon offers a parable of an ant making its way across the sand:

> We watch an ant make his laborious way across a wind- and wave-molded
> beach. He moves ahead, angles to the right to ease his climb up a steep dunelet,
> detours around a pebble, stops for a moment to exchange information with a
> compatriot. …Viewed as a geometric figure, the ant's path is irregular, complex,
> hard to describe. But its complexity is really a complexity in the surface of the
> beach, not a complexity in the ant. An ant viewed as a behaving system, is quite
> simple. The apparent complexity of its behavior over time is largely a reflection
> of the complexity of the environment, in which it finds itself. (1969: 51-2).

The embedded view of cognition tells us that, like the behavior of the ant, human cognition occurs within a wider complex system. And, in order to understand and explain human cognition, cognitive science cannot just study the internal processes of computation instantiated in the brain. Instead, cognitive science needs to study the way that structures in the local environment of an agent facilitate the success of the agent's cognitive processes. One example of this is the phenomena known as 'non-trivial causal spread', which refers to the utilization of external factors and forces to complete cognitive tasks (Clark 1998; Wheeler and Clark 1999; Shapiro 2011). The *mise en place*

method of lining up one's ingredients in the correct order for cooking, for instance, is widely used by cooks to save them from having to remember the ordering of their ingredients while cooking (Kirsh 1995; Clark 2008b)[8]. For the embedded mind theorist, this is an example of how structures in the local environment can facilitate the agent's cognitive successes and why we need to consider those wider structures in order to fully understand how human cognition works. The embedded view thus takes a holistic approach to understanding human cognition. Its insights offer us explanatory, or epistemic, reasons to look beyond the brain (Rowlands 2010).

0.3.2 Rejection of representationalism and computationalism

The embedded view also calls for an attenuated role for internalized representations of the world, thereby challenging cognitive science's reliance on internal mental representations (Rowlands 2010). The work on passive dynamics in robotics also contributed to this challenge. As noted, the simple slinky toy is able to make its way down a set of stairs without any internal representation of the world. A number of situated views, including embedded, embodied, and enactive views, have likewise argued that humans do not need to store a complete representation of the world they find themselves in (Clark 2008b). To do so is too demanding in terms of storage, energy, and computation. Thus, situated theorists emphasize how agents can rely on the world around them rather than producing an internal representation of the world and relying on that. Using the *mise en place* method of cooking, cooks are able to structure their environment to store the information they need and then to rely on these external structures instead of on internal representations of the external world. Thus, the agent saves herself from having to make use of costly internal representations (Kirsh 1995; Clark 2008b; Shapiro 2011).

---

[8] Hutchins's (1995) 'distributed' view of cognition also emphasizes how intelligent workspace design facilitates our cognition.

A key figure who challenges the need for internal representations in human perception was the psychologist J.J. Gibson. Gibson took a systems thinking approach to the study of cognition that explained behavior as a dynamic relation between the agent and her environment (Clancey 2009). Applying his theory to visual behavior, Gibson (1977, 1979) developed an ecological view of perception, which maintains that perceivers do not rely on an internal representation of the world but instead rely on the world itself. As such, perceivers are constantly monitoring the world for changes. On this view, there is an open channel between the perceptual system of the agent and the world such that the perceiver is tuned to exploit invariances in the 'optic array'. In order to uncover these invariants, the agent must move around its environment in a process known as 'information self-structuring' (Shapiro 2011; Chemero 2003). An example of this process is how we can use head movements to structure information for the purposes of depth perception. We see objects that are closer to us as moving faster than objects that are further away, and this depth cue, known as motion parallax, signals to us (often implicitly) the relative distance of objects. For example, by fixating on an object at a far distance and moving one's head, any objects that lie at a mid-distance (i.e. between oneself and the point of fixation) will appear to move in the opposite direction from that which one's head is moving in. Their movement in the opposite direction provides the visual cue that these objects are at a mid-distance. Meanwhile, any objects that lie even further away, beyond the point of fixation, will move in the same direction providing a visual cue that they are further away. (Shapiro 2011). While the dominant account of perception in traditional cognitive science has been David Marr's (1982) computational account of vision, which applied a reductionist methodology and posited the need for internal representations of the world, situated theorists are generally quite sympathetic to Gibson's rival account (see Robbins and Aydede 2009).

Both embodied mind theorists and enactivists have rejected the representational and computational commitments of Cartesian cognitive science. There are many different conceptions of what embodiment amounts to. Wilson (2002), for example, distinguishes between six different views. While Rowlands (2010) and Shapiro (2011)

10

each distinguish between three different interpretations.[9] Some embodied mind theorists aim to replace our need for mental representations by appealing to how the organism's body interacts with its environment, just as Gibson had (Shapiro 2011). This *replacement* project, which supported by the developmental psychologist Esther Thelen, for example, is clearly anti-representational (Shapiro 2011). Thelen states her view of embodiment as follows:

> To say that cognition is embodied means that it arises from bodily interactions with the world. From this point of view, cognition depends on the kinds of experiences that come from having a body with particular perceptual and motor capacities that are inseparably linked and that together form the matrix within which memory, emotion, language, and all other aspects of life are meshed. The contemporary notion of embodied cognition stands in contrast to the prevailing cognitivist stance which sees the mind as a device to manipulate symbols and is thus concerned with the formal rules and processes by which the symbols appropriately represent the world. (Thelen et al. 2001: xx)[10]

---

[9] For example, the three versions of embodiment that Rowlands (2010) distinguishes between are: an *epistemic interpretation*, which maintains that cognitive processes cannot be fully understood or explained without an understanding of the bodily structures within which these processes occur; a *dependence interpretation*, which maintains that cognitive processes (causally) depend on wider bodily structures; and a *constitution interpretation*, which says that cognitive processes are partially composed, or constituted, by bodily states and processes. Rowlands describes the constitution interpretation as the strongest of the three, as it alone challenges intracranialism, while the first two are actually compatible with what traditional cognitive science maintains. For this reason, I only discuss the constitution interpretation. Meanwhile Shapiro (2011: 3-7) distinguishes among three different research projects undertaken by embodiment theorists: the *conceptualization project*, which claims that the concepts an agent requires to interact with its environment depends on the particular form of the agent's body; the *replacement hypothesis*, which I will discuss in this section; and the *constitution hypothesis*, which maps on to Rowlands' 'constitution interpretation'.

[10] Quoted in Clark (2008b: xxvi).

According to Thelen, and like-minded embodied theorists, traditional cognitive science is wrong in thinking that in addition to our perceptual and motor capacities, there must also be representational structures and concepts stored in our heads (Shapiro 2011). On this view, it is akin to a category mistake to say that in addition to all of our bodily capacities, we have mental representations and states (Shapiro 2011). Thus, cognitive science would be better off replacing the study of internal representations and states with direct empirical studies of our bodily capacities.[11]

Enactivism is a third situated view of cognition. It is perhaps the most radical situated view in so far as it rejects all of the Cartesian commitments that I have outlined and, as a result, enactivists advocate a major departure from the distinguishing features of cognitive science (Varela, Thompson, and Rosch 1991; Thompson 2007; Hutto and Myin 2013). While traditional cognitive science says that a system needs to be both representational and computational in order to be a cognitive system, enactivism says that a system must be *autonomous* in order to be a cognitive system. A system is *autonomous* in the relevant sense if it is self-producing and self-regulating: it regulates its own interactions with its environment so as to bring about, or create, information and meaning (Thompson 2007; Di Paolo and Thompson 2014). Enactivists also reject intracranialism, instead maintaining that our cognition is enacted by our bodily interactions with the external world. Thus, on their view, it is more than just the brain that constitutes the autonomous system responsible for human cognition (Thompson 2007; Gallagher 2013).[12]

---

[11] Some situated views advocate a reduced need for internal representations, while other reject the need for any internal representations at all, thus advocating a more radical rejection of the traditional doctrine.

[12] The enactive theory of mind could, thus, be seen as a version of extended mind theory, since (on most versions) it maintains that objects in the environment can be partially constitutive of an agent's mental state. But it is important to distinguish enactivism from the version of the extended mind thesis that I will be concerned with. Clark and Chalmers's extended mind thesis, as I will explain in the next section, accepts both representationalism and computationalism: it only challenges the intracranialist commitment of traditional cognitive science (see Clark 2008a,b). Thus enactivism advocates a more radical departure from traditional cognitive science than the extended mind thesis does.

0.3.3 Rejection of intracranialism

Recall that intracranialism is the view that the brain is the whole constitutive base of the mind. In other words, all the states and processes of the mind are instantiated by brain states and brain processes. Thus, the intracranialist maintains that the skull marks a significant boundary between the mind and the world. Historically, a number of philosophers have emphasized that intentionality, language, or cognition more generally are constituted by a larger web of linguistic, social, and cultural practices. On their view more than just more than just the brain constitutes our mental life. I will limit this section to a review of a few key twentieth-century thinkers who challenged intracranialism, thereby setting the stage for situated views of cognition to emerge. I will also discuss several situated views that reject intracranialism, including embodied, enactive, and extended theories.

Developing Brentano's and Husserl's idea of intentionality in *Being and Time* (1927), Heidegger argues that standing in an intentional relation with the world requires already being in the world "in a more basic, ontological way" (Gallagher 2009: 38). Thus, as beings with intentional states, we are necessarily situated, or embedded, in the world: we are beings-in-the-world. Heidegger explores the question: "what kind of existence does the human being have such that it is necessarily situated or embedded in the world?" (Gallagher 2009: 39) His response is that there is no possible alternative for an organism but to be embedded in the world. It is part of the very nature of *Dasein* (human existence) to be deeply situated in the environment and to stand in relations to others (Gallagher 2009). Much has been written about Heidegger's views on embeddedness and his notion of tools being *ready-to-hand* in a way that allows them to be constitutive parts of our cognition, along with our brains.[13] It is sufficient here to note that many situated cognition theorists are inspired by Heidegger's work, which serves as one of the philosophical

---

[13] See Gallagher (2009) and Rowlands (2010) for lengthier discussions.

foundations for situated cognition, especially the embedded and extended theories of mind. Rowlands (2010), for example, draws on Heidegger's account of intentionality, as well as his notion of *readiness-at-hand*, in his own argument for extended cognition (see also Clark 2008b).

Another key figure that challenged intracranialism is Merleau-Ponty (1962) was working in the same phenomenological tradition as Heidegger, but was focused more on aspects of our embodiment than Heidegger. In *Phenomenology of Perception* Merleau-Ponty argues that our embodiment is central to the way we experience the world. His anti-Cartesian view was influenced by Bergson, who, in *Matter and Memory* (1911), also defended an embodied view of cognition and by Husserl's concept of embodiment, which was developed in several of his unpublished manuscripts (Gallagher 2009). But he also drew on work from psychology and neurology, discussing concepts such as *body schema*, the body's internal awareness of its own position, which have become central to contemporary arguments for embodied theories of mind (Gallagher 2005). Merleau-Ponty emphasizes that both spatiality and temporality are dimensions of our situated existence and, as such, they permeate almost every aspect of our lived experiences. His view laid the groundwork for later embodied theories of mind.

In the previous section I discussed Thelen's version of embodiment, which challenged representationalism and computationalism. Some defenders of embodied theories of mind, including Noë (2004) and Gallagher (2005), also challenge intracranialism. This version of the embodied mind thesis says that (at least some) cognitive processes are partially constituted by bodily states and processes.[14] Embodied mind theorists argue that there is no reason to privilege neural processes over the contributions of extra-neural bodily processes which, on their view, play just as essential of a role in bringing about cognition. By maintaining that human cognitive processes are partially *constituted* by human bodily processes, these embodied mind theorists deny the

---

[14] As noted earlier, Rowlands (2010) and Shapiro (2011) both distinguish the 'constitution hypothesis' from several other theories or projects that often fall under the same umbrella of 'embodied' theories. Furthermore, Rowlands and Shapiro both agree that the constitution version of embodiment alone challenges intracranialism.

separability thesis, the view that a humanlike mind could exist in a nonhuman-like body (Shapiro 2011).

The extended mind thesis takes the rejection of intracranialism a step further, maintaining that the mind can be partially constituted by extra-bodily processes working together with neural and bodily processes (Clark and Chalmers 1998; Clark 2008b). In other words the thesis maintains that while minds might require brains, they can sometimes *extend* to be instantiated by objects located beyond their core biological shells.[15] Thus the view makes a claim that the mind could be constituted by: things outside of our brains and bodies. But it is not an ontological thesis: it is compatible both with dualism or physicalism, for example (Chalmers 2008). Extended mind theorists are also likely to resist a reductive scientific methodology and to prefer a more holistic systems thinking approach to understanding the mind—one that considers the contributions of extra-bodily resources.

I will hold off on giving a full characterization of the extended mind thesis here because this is precisely what is at issue in this thesis. But it is worth distinguishing it from the other situated views I have discussed. As we've seen, the embedded theory of mind also emphasizes the contributions of the extra-bodily environment. But the embedded view falls short of making a *constitutive* claim. The extended mind thesis, on the other hand, maintains not only that cognition causally depends on wider processes in the environment, but also these wider processes can constitute that cognition. This constitutive claim is what distinguishes these two theses and explains why the extended mind thesis is incompatible with intracranialism, while the embedded mind view is compatible with it (Wilson and Clark 2009; Rowlands 2010).

What distinguishes the constitution claim made by extended mind theorists from the embodiment constitution hypothesis is that the former view, but not the latter, maintains that cognitive processes can be constituted by *extra-bodily* processes and not

---

[15] I use the term 'object' loosely here (and throughout). First of all, it is the *information bearing structure(s)* in the object that partially constitute the mental type, not necessarily the entire object. Second, I remain open to the possibility that 'objects' could include physical properties, including, for example, a particular ordering of objects.

just by extra-neural bodily processes. Thus, the extended mind thesis is unique in making a constitutive claim about extra-bodily parts or processes. It is in this sense that the extended mind thesis goes further than the embodied view in its rejection of intracranialism: it maintains that mental states and processes can *extend* not only beyond one's brain, but also beyond one's body. For this reason some see the extended mind thesis as more radical than other situated views (e.g. Theiner 2011). But part of the aim of this thesis will be to argue that the extended mind view is not as radical as one might initially think. We have already seen that the extended mind thesis is less radical than at least one other situated view, enactivism, which rejects not only the intracranialist commitment of traditional cognitive science but also the commitments to representationalism and computationalism (Rowlands 2009a).

0.3.4 Situated views of cognition

I have characterized traditional cognitive science as involving several Cartesian commitments: (a) scientific reductionism, (b) representationalism and computationalism, and (c) intracranialism. We've seen that not every situated view of cognition rejects *all* aspects of the traditional doctrine. In fact, some accept much of the traditional doctrine. Some have even pointed out that certain situated views of cognition are actually at odds with some others (e.g. Clark 2008a; Rowlands 2010). This raises the question of what situated theories of cognition all have in common and to what extent these views really belong under the same umbrella (see Robbins and Aydede 2009; Rowlands 2010; Shapiro 2011). As a brief answer to this question, all situated theories push back against traditional Cartesian cognitive science in some way or other, but are otherwise quite diverse. They all emphasize either (i) a systems thinking approach, (ii) an "attenuation" of the role of representations and computations (Rowlands 2010: 33); or (iii) a rejection of intracranialism. In the case of some situated views of cognition, such as enactivism, all three are emphasized. On the view to be defended over the next few chapters, the core

claim of the Extended Mind thesis is compatible with most of the traditional doctrine, as Clark (2008a,b) himself says; it only challenges the intracranialist commitment. Thus my main preoccupation in this thesis will be to show how and why intracranialism is false. For an overview of the landscape, Table One outlines which commitments are accepted and which are rejected by each of the four main situated views: (i) embedded cognition, (ii) embodied cognition, (iii) enactivism and (iv) extended cognition. From here on, the focus of this dissertation is Clark and Chalmers's version of the extended mind thesis, which of the family of views that have emerged within the approach of situated cognition offers perhaps the most radical rejection of intracranialism.

Table One: On situated views and Cartesian commitments.

	THREE COMMITMENTS OF TRADITIONAL COGNITIVE SCIENCE		
SITUATED VIEWS OF COGNITION	(a) SCIENTIFIC REDUCTIONISM	(b) REPRESENTATIONALISM & COMPUTATIONALISM	(c) INTRACRANIALISM
(i) EMBEDDED	Rejects	Some accept, some reject	Accepts (or is consistent with)
(ii) EMBODIED	Some accept, some reject	Some accept, some reject	Rejects
(iii) ENACTIVISM	Rejects	Rejects	Rejects
(iv) EXTENDED[16]	Rejects	Accepts	Rejects

---

[16] Here I refer to Clark and Chalmers's version of the Extended Mind thesis.

0.4 Aims and contributions

Clark and Chalmers (1998) defend the extended mind thesis on the grounds of what they call *parity*. They argue that when the representations in an external resource are functionally isomorphic to internal neural-based representations we should credit those external representations as partially instantiating the mental state, on par with how we treat internal representations. Many have understood Clark and Chalmers's parity argument as depending on a functionalist theory of mind. Functionalism maintains that the mind can be characterized on an abstract level in terms of its functions or relations, while the actual physical realization is not essential to an understanding of mental functions. The fact that a functional characterization makes no reference to the realization of mental states makes room for the possibility that not only the brain but also other physical realizers could support mental functions (see Putnam 1960, 1967). This possibility is known as *multiple realizability*. Functionalism and the multiple realizability thesis are two pillars of 20th century metaphysics of mind. Because these views continue to be mainstream in philosophy of mind, it makes sense that Clark and Chalmers would have appealed to them in formulating an argument for the extended mind. And it further makes sense why some have thought that the parity version of the extended mind thesis follows as an implication of the commitment to functionalism.[17] The main purpose of this thesis is to explore how Clark and Chalmers's view interacts with functionalism.

While functionalism is a familiar view in philosophy of mind it is not uncontroversial; thus, insofar as the extended mind thesis is thought to require functionalism critics have had reason to object to the thesis.[18] Furthermore, there are other concerns over whether functionalism really can support the extended mind thesis (e.g.

---

[17] Wheeler (2010a: 245), for example, says that "[t]he claim [the extended mind thesis] is in some way a form of, dependent on, entailed by, or at least commonly played out in terms of, functionalism is now pretty much part of the received view of things." Drayson (2010) makes similar remarks.

[18] Searle, a prominent critic of functionalism, for example, has called the extended mind thesis "nonsense" (Boag 2014).

Rupert 2004; Sprevak 2009; Rowlands 2010; Wadham 2016). Indeed, debates about functionalism have dominated discussions of the merits of the extended mind thesis since its conception. Some have responded to these concerns and objections by giving up on the original parity argument entirely and instead providing arguments for the view that do not depend on functionalism.[19] The resulting views, however, never look quite the same. Some provide vastly different pictures of what the mind is and where it can be located – so different that some conflict with the picture that Clark and Chalmers (1998) originally put forth. Some of the new arguments lead to a more liberal view, that some would describe as *over*-extending the mind,[20] while others are relatively conservative.[21] Furthermore, each of these new arguments face objections of their own.

In what follows I argue that the emphasis on the relevance of functionalism in the extended mind debate is misplaced and leads to a misunderstanding of the import of the extended mind thesis. My strategy is to deconstruct Clark and Chalmers's parity argument, by showing that it can be stripped of any commitment to functionalism or multiple realizability. The central argument of the thesis severs the allegedly necessary connection between the extended mind thesis and functionalism. In doing so, my aim is to insulate the extended mind thesis from the objections that confront functionalism and from the major objections that have surrounded the original functionalist version of the parity argument.

Having stripped the parity argument of functionalism and multiple realizability, I then argue that what is left at the core of the argument is a claim about the location of mental states, rather than a claim about their nature or their composition. In particular, I argue that the parity-driven extended mind thesis is committed to a view that I call the thesis of the *multiple localizability* of mental states. The multiple localizability thesis maintains that within a single subject the same mental type can be located in (sometimes drastically) different kinds of places. Thus, I make a distinction between multiple realization and multiple localization. I argue that the multiple localizability of mental

---

[19] I review some of these arguments in Chapter One.
[20] This is Gertler's (2007) term.
[21] I discuss these differences is in Chapters One and Three.

state types is quite plausible. But what is most important for my purposes is that it is supported by neuroscientific evidence and thus less controversial than the multiple realizability thesis is. On my view then, the parity-driven version of the extended mind is really a thesis about the location of mental states, while functionalism is about the nature of mental states and the multiple realizability thesis is about their composition. By pulling these views apart, I show that the extended mind thesis need not rely on functionalism or multiple realizability.

My hope is that this new multiple-localization based parity argument for the extended mind thesis will shift the focus of the current debate. Stripped of its relationship to functionalism, we are better able to judge the extended mind thesis on its own merits. What we find is that the metaphysical requirements for the extended mind thesis are in fact quite minimal. And so the once seemingly radical thesis becomes difficult to deny. This is true even though, as I argue, the view also has a wider scope than Clark and Chalmers maintain. I argue, for example, that they must accept that consciousness can, at least in principle, extend, along with non-conscious states –something they have argued against.

The focus in the philosophy of mind has long been on the nature and composition of psychological function. But by clarifying that the extended mind is really a distinct thesis about the location of mental kinds, we can see how it, and perhaps other situated views of cognition as well, have shifted the focus from the issue of the composition to the issue of location. The extended mind thesis is not, therefore, simply an implication or extension of the functionalist view that has been around since the 1960s. Instead, it offers new insight into mentality.

0.5 Summary of Chapters

Chapter One: The Extended Mind thesis

I begin this thesis by introducing the central argument under discussion: Clark and Chalmers's parity argument for the extended mind thesis. In this chapter I consider two key examples that they give: Tetris and Otto and Inga. Central to their argument are the glue-and-trust conditions for parity. I will introduce each of these and then consider a series of objections that have been raised against them. I end with a discussion of how the parity-driven version of the extended mind thesis differs from other accounts of cognitive extension.

Chapter Two: An Overview of Functionalism

The aim of the second chapter is to introduce functionalism in some detail. Here I will distinguish between some of the different kinds of functionalism that have emerged over the last half-century, since the view was first defended by Hilary Putnam (1960). This will also include a discussion of some of the key features of functionalism, such as the Lewis-Ramsey method and multiple realizability. Ultimately, the goal of this chapter is to set the stage for the arguments that will come in subsequent chapters.

Chapter Three: Functionalism and the Extended Mind thesis

A crucial part of Clark and Chalmers's parity argument for the extended mind thesis is the parity principle, which they say relies on a functionalist account of the nature of mental states and processes. Most commentators agree that there is a tight relationship between the extended mind thesis and functionalism: some claim, along with Clark and

Chalmers, that a parity-driven argument requires functionalism, while others say that certain versions of functionalism entail the truth of the extended mind thesis. The main objective of this chapter is to explore the relationship between functionalism and the extended mind thesis. In doing so, I will review a host of functionalist-related objections that have been raised against the parity argument. I will argue that at every turn the functionalist-driven parity argument runs into a new problem. I end by considering what has been a popular reaction to these objections: many second-wave extended mind theorists have moved away from considerations of parity entirely and instead advance non-parity arguments.

Chapter Four: Pulling functionalism and parity apart

In the fourth chapter I argue that, contrary to the popular wisdom, the parity argument does not require the truth of functionalism or even the multiple realizability thesis. This is because the parity principle that motivates the argument rests on a theoretically distinct issue from these other views. Crucial to my position is an understanding that mental types being multiply realizable is compatible with a non-functionalist conception of mind. I begin the third chapter by making this point. I then spend some time diagnosing why functionalism and the parity argument are thought to be closely associated by Clark and Chalmers, and by others. But I maintain that, while the original version of the parity argument appeals to some kind of functionalism, a non-functionalist parity-style argument that relies only on multiple realizability is possible. Thus, Clark and Chalmers emphasize the connection between functionalism and parity, but in fact their argument could do without it: functionalism is not among the core, or most central, theoretical commitments of a parity-supported extended mind thesis. This leads to a consideration of how a non-functionalist parity-driven argument might go: showing that the mind could extend, in just the way that Clark and Chalmers describe, without even multiple realizability. Indeed in the next chapter I argue that the real content

of the extended mind thesis consists in a claim about the location of mental states.

Chapter Five: The what and where of mental states: on what is distinctive about the extended mind thesis

Clark and Chalmers's version of the extended mind thesis provides an answer to the question *where is the mind*? I argue that the real import of the extended mind thesis consists in a claim about the whereabouts of mental types. To make this clear, I present an argument for the extended mind thesis that does not rely on either functionalism or multiple realization. The argument instead requires what I call the *multiple localizability thesis*, which says that particular kinds of mental states need not be strictly or uniquely located in any particular place, e.g. the brain or one of its regions. I begin the chapter by arguing that evidence of neural plasticity shows that mental states are multiply localizable. The aim of advancing a new parity-driven argument for the extended mind thesis is to clarify what I believe is distinctive about the view: that it offers us new insight about the location of mental types. Thus distinguishing between multiple realizers and multiple locations helps illuminate the idea that the extended mind is essentially a thesis about the location of mental states, not a claim about their nature or their composition, which is why it can remain uncommitted to functionalism and multiple realizability.

There are two appendices to this chapter. The first, Appendix A, offers additional details on the whereabouts of mental states beginning with a brief overview of some of the key figures who established the tradition of cerebral functional localizationism. I then argue that the evidence that established this tradition, a tradition that continues today, does not support strict localizationism. I end by making clear how multiple localizability is compatible with mechanistic explanations in modern cognitive neuroscience. In Appendix B I take a critical look at the evidence of neuroplasticity that philosophers have cited in support of multiple realizability. I then highlight how the distinction I make in Chapter Six between multiple realization and multiple localization fits in to some contemporary discussions about brain mapping.

Chapter Six: The parity argument for extended consciousness

Some take consciousness to be the mark of the mental, so an extended mind would not be possible without extended consciousness. But, Clark and Chalmers deny that an agent's conscious mental states could extend. According to them only non-conscious mental states can extend, while consciousness supervenes on internal brain states alone. I argue that their position on this matter is untenable: they have no good reason to deny that their parity argument can support extended consciousness. Instead, it supports extended consciousness just as well as it supports extended non-conscious states. Furthermore, this is true whether or not one appeals to a functionalist-driven version of the parity argument or to a non-functionalist version of the sort I present in the previous chapter.

Chapter Seven: Conclusion

In the final chapter I offer some concluding remarks. I also consider and respond to several objections to the central claims that I have defended in this thesis. I begin with some objections directed at the arguments in Chapter Five: both to the non-functionalist version of the parity argument and to the claim that the extended mind thesis is primarily a view about the location of mental states. I then consider two objections to the argument in Chapter Six for the possibility of extended consciousness.

CHAPTER ONE

THE EXTENDED MIND THESIS

1.1 Introduction

By now many arguments have been advanced in support of the extended mind thesis, understood as the claim that extra-bodily states and processes can partially constitute mental states and processes. My primary concern in this thesis is with the most influential of these arguments—the parity argument advanced by Clark and Chalmers (1998).[22] I focus on Clark and Chalmers's argument not only because it is has received the most attention in philosophical discussions (and perhaps more widely as well) but also because, as I will argue, it has been misunderstood in both an interesting and revealing way. I begin this chapter by introducing the parity argument and discussing some of Clark and Chalmers's supporting examples. Following this, I provide an overview of some of the main objections that have been raised against the parity argument. This will include objections to their proposed set of conditions, but not objections specific to the argument's reliance on functionalism, which I will discuss in chapter three.

1.2 Clark and Chalmers's extended mind thesis

In the introduction, I described the extended mind thesis as rejecting intracranialism, a core commitment of traditional cognitive science, but as accepting other commitments, such as representationalism and computationalism. In rejecting

---

[22] Wadham (2016: 137) similarly describes the parity argument as the "most well known and influential" argument for the extended mind thesis.

intracranialism, the extended mind thesis makes a constitutive claim about the objects and resources that could instantiate the mind. The extended mind theorist maintains that extra-bodily states and processes can be partially constitutive of the mind. The view is not that extra-bodily states and processes *on their own* could constitute a mind, but rather, working together with neural and bodily processes, extra-bodily resources can serve as *part of* the supervenience base of the mind.[23]

A further point is that, while the view makes a claim about what the mind could be constituted by—things outside of our brains and bodies—it is not an ontological thesis. That is to say, the extended mind thesis is compatible with either dualism or physicalism, for example.[24] Although most extended mind theorists are physicalists (of some kind), nothing prevents one from being a dualist (of some kind) and also endorsing the extended mind. Another point, as we have seen, is that the thesis is compatible with representationalism and computationalism, two core commitments of traditional cognitive science. There is, however, an ongoing debate about what the extended mind thesis is really about—whether it is about the nature of mental states, about their composition, or about their location. This is the issue at the heart of this thesis and so I will not take a position at the outset. Instead this will be explored over several chapters. In this chapter the main objective is to introduce Clark and Chalmers's influential parity argument. But first, it is worth situating the parity argument within the landscape of other arguments that have been given the for the extended mind thesis.

Clark and Chalmers's argument was one of the earliest in a now burgeoning literature on cognitive extension within contemporary philosophy of mind. There are now several other arguments for the extended mind thesis, most of which move away from considerations of parity entirely. Rowlands (2010), for example, appeals instead to Heidegger's notion of intentionality as revealing activity in order to support cognitive extension. Sutton's (2006, 2010a,b) argument is based on the complementarity of inner

---

[23] I will not distinguish between different metaphysical relations such as constitution, realization, or supervenience. Instead, I use these as interchangeable as I believe that nothing I say in this thesis will hinge on which relation one opts for.

[24] Chalmers (2008: xv).

and outer processes rather than their functional equivalence. DiPaolo (2009) appeals to the enactivist school to give a non-functionalist development of extended cognition. And more recently Palermos (2014) looks to dynamical systems theory to provide two distinct arguments in support of a qualified version of extended cognitive systems, which treats continuous mutual interactions (or "loops") as necessary for extension.[25]

As a consequence of this move away from parity, however, some of these arguments support versions of cognitive extension that look quite different from Clark and Chalmers's parity-driven view. On Palermos' version, for example, someone using a notebook would not count as a case of extension because one cannot mutually interact with a notebook—a requirement on Palermos' view. On Rowlands' account many more objects in the environment will count as partially constitutive of an agent's cognition. Thus, for those who raise concerns about Clark and Chalmers's view leading to cognitive bloating (Rupert 2004, 2009) or over-extension (Gertler 2007), Rowlands' more liberal versions of cognitive extension would certainly be too permissive.

An in-depth discussion of all of these views falls outside the scope of this thesis, but it is worth mentioning these other non-functionalist versions if only to distinguish them from the parity-driven version that I will introduce in this chapter. I will be concerned only with the version of the extended mind thesis that results from an argument driven by considerations of parity (understood as the symmetric treatment of internal and external states and processes). Thus, my aim over the next few chapters is to develop a parity-driven argument for the extended mind thesis that does not need to rely on functionalism. My main objective is to clarify what the claim at the heart of Clark and Chalmers's influential version of the extended mind thesis really is.

---

[25] Other arguments for this claim can be found in Chemero (2009), Clark (1996), Hurley (1998a,b, 2010), Hutchins (1995), Menary (2007), Noe (2004, 2009), Wheeler (2005), and Wilson (1994, 2004), among others. And many of these thinkers were inspired by similar views from outside of philosophy, e.g. Hutchins (1995); Dawkins (1982); Vygotsky (1930), Gibson (1979), and Donald (1991) (as Rupert notes 2004: 389).

1.3 Extended cognition: The Tetris example

Clark and Chalmers (1998: 7) begin their article with an example involving "three cases of human problem-solving." The example describes three ways of solving a problem that one might encounter while playing Tetris, a familiar computer game:

(1) A person sits in front of a computer screen which displays images of various two-dimensional geometric shapes and is asked to answer questions concerning the potential fit of such shapes into depicted "sockets." To assess fit, the person must mentally rotate the shapes to align them with the sockets.

(2) A person sits in front of a similar computer screen, but this time can choose either to physically rotate the image on the screen, by pressing a rotate button, or to mentally rotate the image as before. We can also suppose, not unrealistically, that some speed advantage accrues to the physical rotation operation.

(3) Sometime in the cyberpunk future, a person sits in front of a similar computer screen. This agent, however, has the benefit of a neural implant which can perform the rotation operation as fast as the computer in the previous example. The agent must still choose which internal resource to use (the implant or the good old fashioned mental rotation), as each resource makes different demands on attention and other concurrent brain activity. (7)

In case (1) there is some kind of cognitive process going on in the person's head. Clark and Chalmers contend that there is likewise a cognitive process going on in case (3) only it involves the use of a neural implant. But they contend that the cognitive process in case (3) is clearly on a par with case (1). By the time their article was written, the possibility of a silicon chip one day being implanted in the brain to serve as a functionally equivalent replacement of neurons was already a familiar idea in cognitive science (see Pylyshyn

1980).[26] Thus, the idea that the neural implant described in case (3) could perform some of the cognitive functions that the biological brain performs in case (1) is something that philosophers of cognition would have been comfortable with. Presumably, this is why Clark and Chalmers do not *argue* that cases (1) and (3) are on a par—they simply assert this and assume that the comparison does not need any further motivation. But, Clark and Chalmers (1998: 7) then go on to argue that case (2) is also on a par with (3):[27]

> [C]ase (2) with the rotation button displays the same sort of computational structure as case (3), although it is distributed across agent and computer instead of internalized within the agent. If the rotation in case (3) is cognitive, by what right do we count case (2) as fundamentally different? We cannot simply point to the skin/skull boundary as justification, since the legitimacy of that boundary is precisely what is at issue. But nothing else seems different.

Recall that traditional cognitive science is committed to computationalism. Thus, when Clark and Chalmers argue that case (2) and case (3) have the same computational structures, they are appealing to a standard commitment of cognitive science. The neural implant described in case (3) would count as on a par with the neurons being used in case (1) precisely because their operations have the same computational structure. Thus, they use the same line of argument to show that case (2) is similar to case (3) (and case 1) that they use to argue that cases (1) and (3) are on par. All three cases, they maintain, have similar computational structures. And, because traditional cognitive science subscribes to (some variety of) functionalism, the idea that physical things as diverse as the human brain and a silicon-chip could realize a mental state would not have been all that new or surprising. Case (2) also involves a diverse set of physical things—a computer and a rotation button—but in a distinct location. Clark and Chalmers's insist that as long as its

---

[26] I will discuss functionalism in some length in the next chapter.

[27] See Kirsch and Maglio's (1994) work on 'epistemic actions'. They found that using the rotation button is more time efficient than relying on mental rotation (as in the first case).

computational structure is on par with the other cases, it should equally be seen as a cognitive process.

So given that many of these views (e.g. computationalism and functionalism) have been around for sometime, what explains the fact that the extended mind thesis has been received as a new and radical view? The new case that Clark and Chalmers present, case (2), alone challenges intracranialism—a commitment of traditional cognitive science. Clark and Chalmers argue that case (2) shows how cognitive processes can sometimes *extend* to be located in objects beyond an agent's brain and core biological shell. Thus, their thesis appeals to some of the core commitments of traditional cognitive science—representationalism and computationalism—to argue that another key commitment—intracranialism—is false. The more familiar case involving a neural implant—case (3)—does not challenge intracranialism, since the cognitive process takes place entirely in the head. To motivate their view, Clark and Chalmers (1998: 8) offer the following principle:

> 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 part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process.

Many now refer to this as their 'parity principle,'[28] while others call it the 'fair treatment principle' (e.g. Sprevak 2009; Drayson 2010), as it maintains that we should treat equivalent processes with "the parity they deserve" irrespective of whether they are internal or external to the skull (Clark and Chalmers 1998: 8). Clark and Chalmers use this Tetris example to support the possibility of cognitive extension—case (2) is an

---

[28] I have a longer discussion of the parity principle in a later chapter.

extended cognitive process. They then give a second example that is meant to show that more standard mental states, such as beliefs, can also extend.[29]

1.4 The parity argument: Otto and Inga

Armed with their parity principle, Clark and Chalmers describe a case in which, they argue, an object in the environment plays the same role for one agent that neurons in the brain—something we would surely count as a part of the supervenience base of the mind—do for another. The case involves two people, Inga and Otto. Inga decides to go to an exhibition at the museum and to do so, "[s]he thinks for a moment and recalls that the museum is on 53rd Street, so she walks to 53rd Street and goes into the museum." (Clark and Chalmers 1998: 12) Meanwhile, we imagine that Otto suffers from Alzheimer's disease and has to rely on information he stores in a notebook to help structure his life. When he decides to go to the same exhibition he consults his notebook, where he has written the address and directions for how to get there. He then walks to the museum and heads inside to enjoy the exhibition.

Clark and Chalmers maintain that in the relevant respects information in Otto's notebook "functions just like" the information in Inga's brain that constitutes an ordinary belief and thus both should count equally as part of the constitutive machinery of his mind (12). Although Otto's actions may differ from Inga's in trivial ways, e.g. how he accesses the information in his notebook, the functional role that the information plays in the two cases is analogous.[30] And, for them, it is this functional role that matters. Thus,

---

[29] Some maintain that cognitive states and processes are a subset of mental states and processes, while others maintain that the two sets overlap. For most of this thesis I will use the terms 'cognitive' and 'mental' as interchangeable. Thus, I make no distinction between cognitive processes or cognitive states on the one hand, and mental processes or mental states on the other. Some do draw a distinction, e.g. Drayson (2010), and I will have a brief discussion of her view in a later chapter. I believe my arguments throughout this thesis are not affected by this distinction.
[30] Clark and Chalmers (1998: 12) claim, "the essential causal dynamics of the two cases mirror each other precisely".

because the information plays the same role in both cases and the information in Inga's case surely counts as her belief about where the museum is located, it follows that the information in Otto's notebook ought to equally count as a part of his belief. And, since beliefs are widely considered a part of one's mind, it follows that a part of Otto's mind is constituted by a resource located beyond his brain and body. It is in this sense that Clark and Chalmers maintain that our minds can extend beyond our biological bodies. The parity argument can be summarized as follows:

> (P1) A physical state (or content-bearing structure) *p* is constitutive of a mental state of type *m* when *p* plays the causal role characteristic of *m* in the system.

> (P2) A physical state (or content-bearing structure) located beyond (or partially beyond) an agent's biological body can play the same causal role as physical states of the biological body that surely constitute an ordinary mental state of type *m*.

> (C) Therefore, physical states located beyond (or partially beyond) the biological body can be constitutive of an agent's mental state.

The first premise expresses a commitment to a variety of functionalism, a view that I will explain in the next chapter and explore over the course of the next few chapters. The example involving Otto and Inga is meant to support the second premise. The type of mental state in question is a non-occurrent (i.e. not currently being entertained), non-conscious belief state. Thus the second premise would maintain that content-bearing structures in the notebook, a physical state, can play the same causal role for Otto as the content-bearing structures in Inga's brain—which we surely count as constitutive of her ordinary non-occurrent belief—play for her. And from this it follows that the content-bearing structures in the notebook can be partially constitutive of Otto's non-occurrent belief.

The parity argument, as stated, might also support the possibility that a wide range of mental states types can extend, from non-occurrent, non-conscious beliefs and desires to occurrent and conscious beliefs and desires. And Clark and Chalmers remain open that the argument could also work for other non-conscious states such as desires (18). But they limit the scope of their thesis to non-conscious states, saying it is "far from plausible" that conscious mental states (or processes) extend (10). Thus, on their view non-conscious mental states can extend, while consciousness supervenes only on internal brain states. Eventually I will argue that both extended non-conscious states and extended conscious states are, at least in principle, possible, and that my main arguments about (and for) the extended mind thesis apply to both. But, for the next few chapters I will focus on examples of non-conscious states, without discussing the possibility of extended consciousness until Chapter Six.

1.5 The glue-and-trust conditions for parity

While the parity principle does not appear as a premise, it plays an important role in motivating Clark and Chalmers's argument. Clark and Chalmers defend this principle in (at least) two ways. First, they argue that to deny the principle outright would beg the question. The parity principle simply demands neutrality on the issue at hand—namely, the whereabouts of the mind. Thus, to assert *a priori* that only things with certain privileged locations, e.g. being located within the skull, can realize our minds would be to take a stance on the very matter under debate.[31] The second way that they defend their principle is by offering a more precise characterization of the functional role that it refers to, e.g. the role that they argue is played by both Otto's notebook and Inga's biological memory alike. This is the causal role referred to in their second premise:

---

[31] Clark and Chalmers (1998: 15-6).

(P2) A physical state (or content-bearing structure) located beyond (or partially beyond) an agent's biological body can play the same causal role as physical states of the biological body that surely constitute an ordinary mental state of type *m*.

For Clark and Chalmers this characterization takes the form of a set of conditions. These conditions have come to play a crucial role in their argument. Notice at the outset that because the parity principle is used to motivate an argument for the extended mind thesis, it is essential that Clark and Chalmers characterize this functional role in a sufficiently abstract way so to allow that resources as diverse as Inga's brain or Otto's notebook could play that role. In other words, they need a way of explaining why these resources should count as constitutive parts of an agent's mind that does not make reference to any particular physical properties. But, at the same time, their description cannot be so abstract that just anything could be recognized as a constitutive part of one's mind. Even an extended mind theorist, after all, will want to maintain the distinction between the mere use of a tool and a device that actually constitutes one's mind. If any object a person uses becomes a part of their mind this would threaten to "overextend" the mind in a way that no longer seems plausible.[32] This is what Weiskopf (2010) refers to as the 'Goldilocks' problem' for extended mind theorists.[33] Their conditions need to distinguish between objects that are genuine extensions of the mind and those that are mere tools, in a way that is "neither too restrictive nor too permissive" (Weiskopf 2010: 313).[34] It is the

---

[32] This is Gertler's (2007) term.

[33] The Goldilocks in reference is from the children's story *Goldilocks and the Three Bears,* in which Goldilocks sits in three different chairs and finds one too big and one too small, but the third to be just the right size. The story emphasizes the idea of something's having just the right amount of some quality, e.g. size or temperature. The 'Goldilocks problem' for the extended mind theorist is to find just the right grain of description for the relevant functional role that the parity argument appeals to. This problem also rears its head when elucidating a set of conditions for parity that are neither too conservative nor too liberal.

[34] This problem is not unique to Clark and Chalmers's goals. Ned Block (1978) points out the difficulty of actually giving an appropriate functional characterization, one that specifies the

role of the conditions for parity to make such a distinction viable. For this reason, these conditions are essential for Clark and Chalmers's argument and, as such, have become the object of much scrutiny. In total Clark and Chalmers outline four conditions to explain in what ways the information in Otto's notebook has a similar causal role in his mental life to the one the information that Inga makes use of (stored in her brain) has in her mental life.[35] Clark (2008, 2010) calls these the *glue-and-trust conditions*, which is how I will refer to them.

Constancy: The constancy condition maintains that the agent must use the external resource regularly. In Otto's case, "the notebook is a constant in (his) life—in cases where the information in the notebook would be relevant, he will rarely take action without consulting it" (Clark and Chalmers 1998: 17). Thus, Otto carries the notebook with him everywhere he goes, storing and accessing information in it constantly. This parallels how information in Inga's internal memory is a constant in her life. The inclusion of this condition is motivated by the intuition that if the contrary were true and Otto were to use the notebook infrequently or as a "one-off" we would be less likely to view his case as functionally analogous to Inga's (15).

Accessibility: The accessibility condition maintains that the resource must be directly and easily available for use. The accessibility of the information is essential for it to play the right kind of role in guiding the agent's actions.[36] In Inga's case, the information stored in her memory can be easily accessed when she needs it, and the same needs to be the case for any external resource. And Clark and Chalmers describe the case of Otto's notebook in the same way—"the information in the notebook is directly available without difficulty." (17) This condition captures the intuition that to the extent that the relevant information is not easily or directly available, and the agent has to struggle to access it,

---

inputs and outputs of a mental kind without being too liberal or too chauvinistic. This has continued to be a challenge for functionalists ever since.

[35] Clark and Chalmers (1998: 17); Clark (2008b: 79-80).

[36] Clark and Chalmers (1998: 15).

we would be increasingly reluctant to call this a belief. This is true for the cases of both Otto and Inga—if Inga, for example, had to take significant time to reflect and search her internal memory for some piece of information, we would be less willing say that it was one of her beliefs, precisely because it was not easily accessible to her.

Reliability: The reliability condition maintains that the agent must *rely upon* the information in the resource, trusting and endorsing its veracity without hesitation. In Inga's case, she takes just a moment to recall the location of the museum from her internal memory and she immediately trusts the veracity of the information. Likewise in Otto's case, once he accesses the information from his notebook (his memory), he takes no time to doubt or question its accuracy. He "automatically endorses" the information and allows it guide his actions in a seamless way (17). Thus, in both cases the agent *relies upon* the information. In both these cases the information also turns out to be *reliable,* in that it successfully plays the role of guiding the agent's actions to the correct destination. But the accuracy of the information is not what is important for the reliability condition: it is the relation between the agent and the information that matters (not a relation between the information and the world).

Prior Endorsement: Finally, Clark and Chalmers tentatively propose the condition of prior endorsement, which maintains that the information in the resource must be there as a consequence of having been consciously endorsed at some prior time.[37] Inga has already endorsed where the museum was located; this is how the information got stored in her memory to be recalled at a later time. Likewise, Otto is responsible for having written down the information about where the library is; he has already endorsed the information as true.

---

[37] I say 'tentatively' because Clark and Chalmers (1998: 18) say this fourth condition is "arguable," but the first three conditions "certainly play a crucial role."

With these conditions characterizing the relevant functional role that Clark and Chalmers refer to in their parity principle, we are now in a better position to understand their parity argument. The argument contends that because the information in both Otto's and Inga's case satisfies these four conditions, despite being realized in starkly different ways and in different places, and because we would surely recognize the information stored in Inga's biological memory as constitutive of her belief, then, given the principle, we ought to recognize the information in Otto's notebook as constitutive of his belief as well. And, since Otto's notebook is external to his brain and body, and beliefs are a part of his mind, it follows that Otto's mind is partly constituted by an external resource.

There is debate over how exactly to understand these conditions. The first issue is whether or not they are intended as independently necessary and jointly sufficient conditions, and the second issue is *what* they are intended to be conditions for. A number of commentators have understood them as independently necessary and jointly sufficient conditions for functional equivalency.[38] This is how I will understand them as well. It is possible, however, that Clark and Chalmers never intended them this way. Clark, for example, later describes the conditions as "a rough-and-ready set of additional criteria to be met by non-biological candidates for inclusion into an individual's cognitive system" (2008b: 79). His use of 'additional' here suggests that they are not intended to characterize the functional role, but are somehow add-ons to the characterization of functional role equivalency.

A further issue is whether the conditions are meant to characterize functional equivalency only for mental states of a particular kind or for all mental states and processing in general. Drayson (2011) has made the case that the conditions are only meant to characterize functional equivalency for non-occurrent beliefs. This is the kind of mental state that Clark and Chalmers are describing when they present these conditions—

_____

[38] Aizawa (2008: 2), for example, characterizes the conditions as a "set of necessary and sufficient conditions under which a resource is a part of an agent's cognitive processing." Sprevak (2009) seems to understand them in the same way. Furthermore, although Clark responds to these commentators on a number of occasions he never takes issue with this characterization of the conditions.

the state that is realized by the information in Otto's notebook and in Inga's brain. Clark and Chalmers write: "Insofar as increasingly exotic puzzle cases lack these features, the applicability of the notion of 'belief' gradually falls off." (17). Thus, they seem to intend them specifically for the relevant notion of belief that they are discussing. This is how I will understand the conditions: as independently necessary and jointly sufficient for the functional equivalency of non-occurrent beliefs, rather than for all kinds of mental states and processes.[39] This is important because, as we will see, there is disagreement over how to interpret this set of conditions and whether they are compatible with the form of functionalism that Clark appeals to. I will return to these issues in a later chapter.[40]

1.6 Objections to the independent necessity of each condition

Usually objections to the conditions for parity take one of three forms: (1) they maintain that it is unlikely or impossible for one (or more) of the conditions to be met by an external resource, such as Otto's notebook; (2) they maintain that one (or more) of the conditions is not independently necessary; or (3) they maintain that the conditions are not jointly sufficient and propose that another condition be added to the set. In this section I will consider objections of the first two kinds and in response I will defend the inclusion

---

[39] This is also the more charitable interpretation of these conditions. While the fourth condition would be satisfied in Otto's case, for example, Clark and Chalmers give a number of other examples of cognitive extension that would not meet this condition, e.g. the extended cognitive process in their Tetris example. They also describe a blind man who uses his walking stick as a cognitive extension of his visual system. Neither of these cases would meet the prior endorsement conditions. The difference seems to be that Otto's extended belief is non-occurrent, while in the other cases the processes they describe are occurrent. Thus, the most charitable interpretation is to take the conditions as specifically for non-occurrent beliefs, as Drayson (2011) argues, otherwise the fourth condition in particular will need to be reconsidered.

[40] Furthermore, the issue of giving a defining 'mark of the cognitive' (or 'mark of the mental') has come up in debates between extended mind theorists and their critics (see Adams and Aizawa 2005, 2008; Fodor 2009; Clark 2008b, 2010). Thus, these conditions should not be understood as defining the 'cognitive'. I will discuss this line of objection in the next section.

of each of these conditions. In the next section I will consider objections of the third kind, i.e. to the joint sufficiency of the conditions.

1.6.1 Objections to the constancy condition

The constancy condition maintains that use of the relevant external resource must be a constant in the agent's life. Clark and Chalmers (1998: 17) introduce the condition by describing Otto as "rarely" taking action without consulting his notebook. Gallagher and Crisafi (2009: 46) have questioned the necessity of this condition; they ask: "[W]hy should we not count as a case of extended cognition the use of a machine that did exactly the same thing, but worked only once, or is no longer available?" Gallagher and Crisafi maintain that as long as the resource is available when needed, even if it was just a one-off use, this should be enough. I, however, share what I take to be Clark and Chalmers's intuition—that something used only once is less obviously a part of one's mind than a resource that is frequently used. We frequently rely on information stored in both hemispheres of our brains, but imagine that one day we discover that (for whatever reason) we only used our left hemisphere once, early on in life, at which point it becomes entirely inactive for the rest of our lives. We would be much more inclined to treat the left hemisphere as peripheral to our cognitive system than the right hemisphere, which supports our cognitive functions at all other times in our lives. The constancy condition tracks the fact that the less disposable a resource is to the functioning of a system, the more likely we are to treat that resource as a constitutive part of the system.

1.6.2 Objections to the accessibility condition

The accessibility condition maintains that the resource must be directly and easily available for use. On Clark and Chalmers's view the accessibility of the information is

essential for it to play the right kind of role in guiding the agent's actions.[41] This condition faces less pressing objections than some of the others but Gallagher and Crisafi (2009: 46) do raise two minor concerns; they ask: "Does easy accessibility mean something different from 'reliably available,' and how should we measure it?" The first concern is whether the accessibility condition is too similar to the constancy condition. I take it that Clark and Chalmers do mean something like availability, but not just that. A resource is 'accessible' in the relevant sense if it is able to reached, understood, used, or obtained. Thus it is *easily* accessible if it is *easily* available, *easily* reached, *easily* understood, and so on. This is different from the constancy condition, which has to do with the regularity or frequency of use.

Finally, regarding Gallagher and Crisafi's second concern, my interpretation is that the accessibility of a resource cannot be measured in any straightforward way, as with the other conditions. The inclusion of this condition is nonetheless defensible on the grounds that, as with our internal resources, the more that we struggle to recall or to put to use some information the less likely we are to describe that information as a belief. It, thus, captures our intuitions on how a belief ought to function.

1.6.3 Objections to the reliability condition

The third condition maintains that the information in the resource must be reliable, such that the agent trusts and endorses it without hesitation. There have been a number of objections to this condition. Gallagher and Crisafi (2008) argue against its inclusion because it is too demanding, while Sterelny (2004) argues that the condition is necessary but that external resources are unlikely to satisfy it. I will consider both of these objections.

Gallagher and Crisafi argue that there are instances in which we subject our cognitive states to critical scrutiny and that this scrutiny should not exclude the states

---

[41] Clark and Chalmers (1998: 15).

from belief status: "[W]hy should some process that would otherwise count as a cognitive process not count as a cognitive process because it requires critical scrutiny, which is itself a cognitive process?" (46)[42] When someone questions the reliability of their memory we nonetheless call their internal states 'beliefs' and recognize them as a part of their minds. Thus, they argue that this condition is too strong and should not be accepted as necessary for beliefs. After all, some people's internal memories are unreliable and would not meet this condition. As a further point they argue that some people's memories are more trustworthy than others.

There are really two issues here. The first is how the agent relates to the information: whether the agent would trust and endorse the information (from that source) without hesitation. The second is how the information relates to the world: whether it truly or falsely represents reality. Clark and Chalmers's must intend their condition in the first sense. The reliability condition isn't about how 'good' the information is (i.e. whether it has a high probability of being true). One's internal beliefs are sometimes false and so too can one's extended beliefs be. Thus, the accuracy or trustworthiness of memory is not what matters. What matters is the agent's sense of trust in the information. The relevant question then is if we distrust our memory—be it internal or external—does it cease to be a part of our mind? Gallagher and Crisafi argue that even when someone questions and critically scrutinizes their (internal) memories we nonetheless call their states 'beliefs'. They maintain that the same should be true of extended beliefs. Thus, requiring that one endorse external informational states without hesitation (as the reliability condition maintains) is, on their view, too strong: if it is not a requirement for our internal beliefs, then it should not be a requirement for external beliefs.

I disagree with Gallagher and Crisafi on this point. I maintain that to the extent that one does critically scrutinize some piece of information before putting it to use we are less likely to call it their standing belief. And, in this way, the reliability condition does capture our standard treatment of (internal) memories. Descartes, for example, had a

---

[42] See also Gallagher (2011: 7).

large number of standing beliefs about the external world. But after he begins his *Meditations* (1641) he 'threw out' his beliefs by subjecting them to critical scrutiny. This scrutiny was such that he would not endorse what he had once believed, at least for the purposes of his attempt to arrive at secure foundations for his knowledge. Thus, the reliability condition tracks our intuitions on how beliefs should function, whether they are internal or external.

Sterelny (2004) raises another concern about the reliability condition; he argues that it is unlikely to be met by an external resource. Sterelny maintains that internally stored information is reliable in a way that externally stored information would not be. The difference, for him, lies in how we access that information. We access internally stored information through introspection, a process that is inherently more reliable than perception, which is how we must access externally stored information. Perception is less stable than introspection and puts the agent at increased risk of being deceived. For example,

> [Otto's external memory is] less reliable after dark; when he forgets his glasses; when his pen leaks or his pencil breaks; when it rains and his book gets wet. And we have not yet considered the issue of epistemic sabotage by other agents. For [Otto] is at risk of thought insertion and deletion to the extent that others have access to his notebook. (246)

Otto also risks having his beliefs stolen. So externally stored information, Sterelny argues, is unreliable and unstable. It is subject to threats in ways that do not parallel internal memory. For Sterelny this is a result of the fact that artifacts and tools are used in shared and sometimes contested spaces—"perception operates in an environment of active sabotage by other agents." (246) For this reason, externally stored information is unlikely to ever meet the reliability condition.

I maintain that even if we agree with Sterelny that the external sphere is less stable and more at risk of interference, it is still possible for agents to rely on information in

external resources, at least some of time. External resources might require extra cognitive demands to protect against these threats, but they can be protected. Parsell (2006: 3), for example, argues that retrieving information from the external sphere will require "the deployment of social guards for the information to remain reliable." Deploying social guards would increase the cognitive load on the agent, but we sometimes also "off-load" these social guards (3). Sterelny considers the possibility of off-loading social guards; for example, being able to recognize his own handwriting might protect Otto from believing something that has been written in his notebook by someone else. But Parsell argues that more sophisticated resources allow for more sophisticated guarding. For example, we create passwords to protect our digital information. By doing so we only need to remember the passwords instead of all the information that they unlock. We can also back up our data to protect it from loss and theft, just as our biological brain is protected. Parsell argues that backing up data to a web-based server removes "even the remote possibility of loss" and "being non-physical (the data) is immune to theft." (8) Both of these claims go too far in my view. Backed-up data can still be lost and non-physical data can still be stolen, in the sense that an unwelcome party can access it, even if a copy of it remains available to the rightful owner. But Parsell's argument shows how an external resource can meet the reliability condition and can be functionally equivalent to internal resources, which can also be lost or stolen (by being accessed by unwelcome parties). Thus, there are ways that we alter our environments to protect externally stored information that can make this information just as reliable as what we store internally.[43]

---

[43] Sterelny's objections are stronger in the case of socially extended cognition—where one person relies on information stored by another agent, instead of information stored in an external artifact. Other social actors are potentially deceptive and we cannot employ the same kinds of safeguards, e.g. passwords or padlocks. But Parsell (2006: 9) argues that as a safeguard in social extensions we establish a trust relation: "To treat others as trustworthy involves refraining from the type of cross-checking that endangers endorsability." I find cases of social extended cognition to be plausible, but I will not try to defend them in this thesis.

1.6.4 Objections to the prior endorsement condition

Recall that Clark and Chalmers's prior endorsement condition requires that the information in the resource be there as a consequence of having been consciously endorsed at some point in the past. And they describe Otto as having entertained and endorsed the information about where the museum is located when he initially enters it into the notebook, thereby meeting this fourth condition. Rowlands (2010: 93), however, argues that some of our beliefs are formed "subliminally," such that we never consciously endorse them. This may be true of implicit beliefs, for example, which sometimes carry content that contradicts our consciously endorsed beliefs. Rowlands (2010: 93; 2009b: 638) argues that we should not regard a state's "mode of formation" as automatically excluding it as a belief. In light of this concern, I suggest that we revise the condition to say that prior endorsement is necessary but prior *conscious* endorsement is not. Standing beliefs that were formed subliminally may have been *implicitly* endorsed at some prior point, even if they were never consciously endorsed.[44] This prior endorsement of the propositional content of the belief, whether consciously or implicitly, helps explain why the agent relies on that information – namely, because she previously endorsed it.

One might contest that what I am suggesting amounts to a rather weak sense of 'endorsement.' Much of our experientially grounded knowledge, for example, is acquired subliminally but is only 'endorsed' in a passive way. The endorsement of this information might be described as merely a failure to challenge incoming information rather than any kind of *active* endorsement on the part of the agent, whether consciously or unconsciously.[45] My proposed modification does weaken the endorsement condition, but not so much so that it cannot play an important role in the set of parity conditions. The condition is needed to distinguish between what should count as a person's standing belief and what should not. There is a significant amount of information that could meet the other three conditions (i.e. be a constant in the agent's life, available to the agent, and

---

[44] All I intend by '*implicit* endorsement' here is that the information was endorsed, but not consciously so.

[45] I thank David Davies for raising this concern.

which the agent might even rely on) but that should not count as a standing belief precisely because the agent has never previously endorsed it, even in this weak sense. All of the information stored in the Google Maps application on Otto's iPhone, for example, (or on the Google search engine in its entirety) might be a constant in his life, easily available to him, and reliable (such that Otto *would* trust and endorse the information from the application without hesitation). But, intuitively, we should not count *all* of this information, the vast majority of which he has never entertained or endorsed, as a part of his mind. If Otto never uses the Google Maps application to visit his friend at the NYU Philosophy Department, for example, then (even if this information meets the first three conditions for Otto) we should not view him as having a standing belief that the Department is on Washington Place. It is only if Otto actually entertains and endorses the information, whether consciously or unconsciously, that we would consider him as having formed the standing belief about where the Department is located. What is key here is that Otto is somehow involved in accepting the information for it to be a standing belief, even if this involvement was nothing more than a failure to challenge the information as he uses it. This is sufficient to exclude all of the other information on the Google Maps application that he has never used and may never use in the future. Thus, the modification that I suggest to the endorsement condition does weaken it, but it is not so weak that it fails to serve an important purpose. Some kind of endorsement, or acceptance, of the propositional content is essential in the formation of a belief. If the information, whether it be somewhere on the Internet or elsewhere, is never endorsed by the agent then it cannot be one of their standing beliefs.[46]

Another concern with my proposed modification is that some might insist that endorsement requires consciousness because it involves some kind of evaluation of the information. But there is no reason why this evaluation cannot be done unconsciously. On the standard computational account of vision, much of our perceptual information is processed through unconscious inferences that involve some kind of assessment of the

---

[46] This condition also rules out foreign entries into the notebook: if a stranger enters information into Otto's notebook while he is sleeping we should not say that he has acquired new standing beliefs while he was asleep, because the new entered information was never endorsed by Otto.

incoming visual information (Marr 1982). This process might only enter into our conscious awareness when things go wrong, while the rest of the time it will go unchallenged. Ultimately, whether we choose to call this process one of 'endorsement' or something else, what matters is that our external beliefs are formed *in the normal way*, that is, through the same kind of mechanisms or processes that form our internally stored beliefs. Since we acquire many of our normal beliefs subliminally, especially experientially grounded beliefs, the same can be true of external beliefs.[47]

1.7 Objections to the joint sufficiency of the conditions

In this section I consider two objections to the joint sufficiency of the conditions for parity. Both objections propose the addition of some new independently necessary condition: (1) original, or underived, intentionality and (2) information integration. I will argue against their inclusion.

1.7.1 Original, or underived, intentionality

Frederick Adams and Kenneth Aizawa argue that even if Otto's notebook satisfies the conditions for parity it is not constitutive of a mental state because it does not have original content.[48] Things that have original content are intrinsically representational: their content is not derived from anything outside of them. On the standard view, our biologically instantiated mental representations have intrinsic content, but representations in the world do not. Things like words on a page, road signs, or even the rings of a tree may be interpreted as representations, but none of these external representations are intrinsically representational. They represent because an observer interprets them as

---

[47] I thank Ian Gold for raising this concern and line of response.

[48] Searle (1980, 1984) sets the distinction between original and derived content out, but it is not uncontroversial (see Dennett 1986).

having some content. Thus, the notebook's content is derived from something that is mental, namely from Otto's biologically constituted mental processes that were occurrent when he wrote the information down and when he subsequently accesses it. For Adams and Aizawa, original content is the mark of the cognitive—it is both necessary and sufficient for cognitive status.[49] So the conditions that Clark and Chalmers provide are, on their view, neither independently necessary nor jointly sufficient.[50]

First notice that this objection might work against the case of Otto and his notebook, but it would not block all cases of extension. In cases of what I call 'socially extended minds' or 'socially extended mental states' the mental states of one individual are partially constituted by the mind of another agent. So, for example, instead of having the information in the notebook as a part of his memory, Otto might rely on the information in Inga's head. This would require that Inga play a big part in Otto's daily life, meeting all the conditions that the notebook would have to. But in this case the information in Inga's memory would be original and thus the objection would not apply.

Furthermore even if original content were the mark of the cognitive, it would not follow that Otto's notebook is not a part of his mind. After all, it is not the case that each and every neuron in Inga's brain that constitutes her mental states display original content, and yet we recognize the information stored in her brain as unproblematically constituting her belief. Those who defend original intentionality view it as 'emergent' from the lower-level physical activity of the brain. But notice that in the case of Otto, all of his mental states, whether extended or non-extended, crucially involve a brain.[51] So while not all the constitutive parts of Otto's non-extended beliefs will have original

---

[49] This is how Adams and Aizawa characterize it, but for our purposes we can also understand it as a proposed 'mark of the mental'.

[50] Rowlands (2009b, 2010) also includes non-derived content as a condition on functional equivalency, although for him it is just one independently necessary condition among others. Likewise, Fodor (2009) argues that there is a principled distinction between the information in Otto's notebook and the information in Inga's brain: only Otto's notebook has derived content while Inga's brain has original content.

[51] In this respect, the extended mind thesis is an organism-centered view, that is, there is a single organism at the center of the cognitive system whose cognition is extended into some resource beyond what is typically seen as the accepted boundary of the organism, e.g. the skin or skull.

content, instead this content is (somehow) emergent from lower-level activity of the brain, likewise not all the constitutive parts of Otto's extended beliefs will have original content. Instead the original content of his extended beliefs will be emergent from the lower-level activity of his brain in conjunction with his notebook. In other words, it is possible that there is original content that is emergent from the wider system that includes both his brain and the external resource, which, on Clark and Chalmers's view, is only partially constitutive of his belief. In fact, this is a lesson I take from Adams' and Aizawa's own 'system thesis', which maintains that there can be "an X system, where only a portion of the thing identified as the system does X" (Aizawa 2005: 2). The idea is that if a system displays some essential property, p, it is not the case that every single part of that system must display p. Aizawa (2005: 2) gives the example of an air conditioner:

> In an air conditioning system, only a fraction of the apparatus is dedicated to actually cooling the air. The majority of a typical system merely ducts the air about the building, compresses the refrigerant, directs the flow of refrigerant, monitors the room temperature, and so forth.

All these parts of the air-conditioning system play a constitutive role in bringing about the higher system level property of cooling air, but they do not themselves exhibit this property. Adams and Aizawa use this to argue that even if the information in Otto and his notebook together constitute a cognitive system, it doesn't follow that Otto's cognitive processing extends.[52] I agree with Adams and Aizawa's system thesis, but I think it shows why even if original content is the mark of the cognitive it doesn't follow that every single part of that system needs to have (or be capable of having) original content. Thus, even if we accept original content as a mark of the cognitive, it would not exclude the

---

[52] This is the so-called 'coupling-constitution' issue, which has spurred a lengthy exchange between Clark and Adams and Aizawa, which I will not discuss. See Adams and Aizawa (2001, 2008); Aizawa (2008); Wilson and Clark (2009); Clark (2008b, 2010). See Shapiro (2011) for a review of this debate.

notebook from being a part of Otto's extended belief. And, if we include original content as a condition for parity, external objects could meet it.[53]

1.7.2 Information integration

Weiskopf (2008) argues that Clark and Chalmers's conditions are not jointly sufficient because they fail to capture a necessary feature of belief: "To count as a belief a state has to be part of system of states in which processes of integration and updating function to keep the subject's mental contents in epistemic equilibrium to some degree or other." (268) In other words, beliefs must be 'informationally integrated' with each other: they must be sensitive to changes in the person's overall set of beliefs such that they are updated to be in concert (265). Weiskopf maintains that informational integration is indispensable from the point of view of belief's explanatory function. But he argues that externally stored information will always fail to satisfy this condition of belief (273).

Imagine that Otto writes in his notebook that the museum is on 53rd Street but then later learns it has been torn down and moved to 63rd Street. Weiskopf argues that upon learning this new information, a normal subject will revise the first belief and instead come to believe that the museum *was* on 53rd street—"Beliefs tend to be updated to reflect novel information the believer acquires." (269) But information stored externally,

---

[53] It is also worth noting that not everyone accepts the distinction between derived and original content in the first place (Dennett 1986). Denying the viability of the distinction might be another way to respond to Adams and Aizawa's objection. Notice that the first reply that I offer, which appeals to Adams and Aizawa's system thesis does not question the viability of the distinction between derived and original content. One might try to argue, for example, that all bearers of derived content in fact have original content because they stand in a systemic relation to us, in just the way that the notebook stands to Otto. But I think that it would be wrong to say that all texts, pictures, or other bearers of content, in fact stand in the appropriate kind of relation to some organism to count as a part of their cognitive system. On the line that I am suggesting my 'system thesis' reply, if Otto's notebook were to be sent into space and never recovered it would no longer be a constitutive part of the relevant activity that brings about his beliefs with underived content, but it would continue to bear derived content.

e.g. in the notebook, will not be updated automatically and unconsciously as normal internal beliefs are. In this case Otto would believe both that the museum is on 53rd Street and that it is not on 53rd Street, an irrational position. Thus, if mental extension is widespread, then this sort of irrationality is also widespread (270). And since we would not want to attribute this kind of widespread irrationality, Weiskopf argues that we should include information integration as a necessary condition. But its inclusion means that, while cognitive extension is not impossible, it is likely nonexistent (or at the very least, it is much less common than Clark suggests).

I contend that information integration is too strong a condition. It would end up excluding too many of our ordinary beliefs. Weiskopf considers this counterpoint: he acknowledges that "there are stubborn beliefs that no evidence seems likely to dislodge." (273)[54] But he argues these are not counterexamples to his condition because they are not typical. The conditions are meant to describe the standard or typical case of a belief, so exceptional cases are not counterexamples. But in fact the kind of irrationality Weiskopf describes *is* widespread. People frequently have contradictory sets of beliefs.[55] Furthermore, while there are different accounts of self-deception, at least on one popular view the self-deceiver maintains contradictory pairs of beliefs about the topic of her deception.[56] As another example, many argue that contradictory beliefs explain our implicit biases, which sometimes carry content that contradict our explicit attitudes.[57]

Another possible response to Weiskopf's objection would be to add a further condition to our set of conditions that maintains that there must be procedures for monitoring the content of the external resource in light of new incoming information on the part of the individual. Perhaps Otto would have to review the information in his notebook. This is not an unreasonable condition: many external resources, including Otto's notebook, could meet it. Some devices might be structured in a way that helps

---

[54] Studies suggest that we can be very resistant to revising our beliefs even when faced with clear counterevidence; e.g. Garrett (2011); Garrett et al. (2013); Garrett et al. (2016).

[55] See Cherniak (1984, 1986).

[56] See Rorty (1988); Pears and Pugmire (1982); Davidson (1985).

[57] See Schwitzgebel (2010); Egan (2008, 2011).

prevent this type of contradictory beliefs. Imagine that the notebook is an address book, for example, and every time that Otto forms a belief about the location of an important building he scans the relevant entry in his book to check for consistency. Returning to Weiskopf's example, when Otto learns that the museum has been torn down and moved to 63rd Street he automatically jumps to the 'M' section in his address book to find his entry for the MoMA to update his belief. In this case Otto would be acting just like the 'normal subject' that Weiskopf describes, who would revise her beliefs upon learning the new information. Thus, a second response to Weiskopf's concern is to simply bring it on board as a new condition, as there an external object, such as the notebook, could still meet it.

There is a third response to Weiskopf's objection that is also available. This response hinges on how exactly we understand the parity principle and the conditions. Are the conditions meant to apply to both internal and external resources? And where should these conditions come from? Rowlands (2009b) argues that the parity principle does not rely on a similarity-based criterion, as many have understood it to (including, for example, Rupert 2004 and Weiskopf 2008). The principle does not say: "if an external process is sufficiently similar to an internal cognitive process, then it too is a cognitive process." (2009b: 635) Instead, the principle says if some process that takes place externally is such that, were it to take place in the head we would surely recognize it as cognitive, then that external process should also be recognized as a cognitive process. In other words, we should not let the fine-grained properties that are unique to biological memory (or to non-occurrent mental states more generally) dictate the conditions for parity. Wheeler (2010a) shares Rowlands' interpretation and argues that there are certain accidental features of human biologically-based beliefs that are not essential to what memory is, and thus the difference between exhibiting or failing to exhibit these features should not mark the boundary between having a belief and not having one. Thus, from the fact that some particular fine-grained feature, such as informational integration, has received attention from contemporary cognitive scientists one cannot infer that this is an essential feature of beliefs or that this should be a condition for functional equivalence.

To do so would be to privilege our internal beliefs in a way that risks begging the question against the extended mind thesis. I will return to this issue in Chapters Three and Four, when I discuss the relationship between functionalism and the parity argument.

Weiskopf's objection also raises the question of how widespread extended cognition is. Weiskopf does not deny that the extended mind thesis is in principle possible, but he denies that it actually happening—that is, that any minds are currently extended into external objects, such as notebooks. This position, now known as 'contingent intracranialism' has been taken by a number of prominent critics to the extended mind thesis and thus it merits a brief discussion.

1.8 Contingent intracranialism

Are minds currently extended? Or is it merely possible that one day they could extend? The parity argument presented earlier in this chapter is for the logical or metaphysical possibility of extended mental states. I believe that this is the main claim that Clark and Chalmers's have tried to defend with their parity argument. In describing what proponents of the extended mind thesis are committed to, Clark says: "[T]hey paint mind itself (or better, the physical machinery that realizes some of our cognitive processes and mental states) as, under humanly attainable conditions, extending beyond the bounds of skin and skull" (2008b: 76). And in a later paper, Clark says that the extended mind thesis maintains that extension is possible "with no giant leaps of technology or technique." (2010: 82). Both of these quotes express the modal claim that it is possible (in this world) for mental states to extend.

A number of critics of the extended mind thesis agree with this interpretation. Weiskopf (2008), for example, argues that the "extended mind thesis makes a modal claim: it is possible for mental states to be externally constituted" (266). It is a separate question whether the extended mind thesis is true of us, that is, whether or not some of our mental states are currently constituted (in part) by states beyond the body (267).

Similarly, Gertler (2007) says, "the mere possibility of extended minds is all that's needed for what is perhaps [Clark and Chalmers's] central contention" (9). But, as Weiskopf notes, defenders of the extended mind thesis "often think that extended mentation is a virtually commonplace phenomenon."[58] It is clear that Clark's own view is that mental states *can* and *do* extend.[59] It may be for this reason that others describe the extended mind thesis as requiring more than just the possibility of extension—that is, as requiring actual cases of extension.

It is important to be clear about where the disagreement is between defenders and opponents of the extended mind thesis, as some interpret the thesis as requiring that mental states be currently extended, not just that extension is principle possible. Long-time opponents of the extended mind thesis, Adams and Aizawa (2001, 2008), for example, accept the possibility of extended mental states in principle but object that there are not currently any actual cases of cognitive extension. They call their opposing position *contingent intracranialism* (2001).[60] Meanwhile, Sprevak (2009: 510) too describes the parity argument as concerning "the actual existence of extended cognitive processes, not their mere possibility."[61] More recently, commentators Miyazono (2015) and Wadham (2016) both agree with Sprevak that the extended mind thesis is committed to more than just the metaphysical possibility of extended belief; it is committed to there being at least some real cases of extended cognition in the actual world.[62]

We can at least distinguish the claim that (1) mental states *can* extend from the claim that (2) at this point in time some mental states actually are extended. The second claim is stronger in the sense that it requires the truth of the first, while the first claim

---

[58] Weiskopf (2008: 266).

[59] Clark (2008b); Sprevak (200).

[60] Weiskopf is also a contingent intracranialist, but on his view this is compatible with what Clark and Chalmers argue for, whereas Adams and Aizawa interpret Clark and Chalmers as defending the view that minds are currently extended. Thus, they see contingent intracranialism as an opposing position.

[61] Sprevak maintains that it would be uncontroversial to say that it is "logically or nomologically possible for cognition to extend" (2009: 503).

[62] Rupert 2004 also seems to understand the view in this way.

could be true whether or not there are any actual cases of mental state extension. I will be concerned primarily with the first claim that mental states could (in this world) extend; my aim is to show what particular commitments about the nature or composition of mental states, if any, are required for this (weaker) claim, as defended by the parity argument.

1.9 Conclusion

The aim of this chapter has been to introduce Clark and Chalmers's parity argument for the extended mind thesis. The argument is supported by the parity principle, by several examples, including their Tetris and Otto and Inga examples, and by their four conditions for parity. I argued that these conditions should be understood as independently necessary and jointly sufficient for the functional equivalency of a non-occurrent belief. Finally, I defended the conditions against a number of objections that have been brought against them.

CHAPTER TWO

AN OVERVIEW OF FUNCTIONALISM

2.1 Introduction

We've seen that the extended mind thesis maintains that mental states and processes can *extend* in the sense that they can be partially constituted by mechanisms beyond one's brain-and-body, while traditional cognitive science takes the brain to be the sole constitutive base. And Clark and Chalmers support the extended mind thesis with the parity argument, a crucial part of which is their parity principle, in which they appeal to a functionalist account of the nature of mental states and processes. Most commentators agree that there is a tight relationship between the parity-driven extended mind thesis and functionalism: some claim that any parity-driven argument requires functionalism, while others say that certain versions of functionalism entail the truth of the extended mind thesis. In this chapter I will offer an overview of functionalism and the different varieties that it comes in. This will allow for a deeper discussion in subsequent chapters about the relation between the parity argument and functionalism. It will help the reader understand why Clark appeals to the version of functionalism that he does, why his appeal to this version of functionalism is problematic, and why a supporter of the parity argument cannot easily appeal to another version.

2.2 An introduction to functionalism

For the past half-century the dominant philosophical theory about the relation between the mental and the physical has been functionalism of one variety or another. The approach was first advanced by Hilary Putnam (1960, 1967) and soon adopted and

further developed by Jerry Fodor (1974, 1981). Functionalism now comes in several different varieties and functionalists continue to disagree on certain important issues, which have led to different functionalist views or strands, some of which I will discuss in this section. But the overarching distinguishing feature of the functionalist family of views is that the mind can be characterized on an abstract level in terms of its functional states while the actual mechanistic realization of mental functions does not bear directly on questions about the nature of the mind.

According to functionalism mental properties just are functional properties (Kim 2011: 133-4). These functional properties are characterized by their causal role in the system—that is, in terms of the input and output relations that they causally mediate (169). These inputs and outputs can include physical or behavioural states, but can also include other internal, mental states including conscious ones. Thus, like the behaviourist account of the nature of the mind that preceded it, functionalism gives a relational account of mental properties. But, for the functionalist, mental states are not just behavioural dispositions, and, functionalism defines mental states in terms of more than just their relations to stimuli (inputs) and responses (outputs); it also includes their relations to other internal states of the system. By thus characterizing mental states, functionalism was able to overcome a major problem that confronted behaviourism, as many charged that behaviourists ignored internal states entirely.[63] This definition of mental states allows the functionalist to include non-manifest mental states of the system that have an effect on the behavioural outputs of the system. This is an advantage over behaviourism, which identifies mental states with behavioural states and thus struggles to preserve the explanatory role of mental states in bringing about behaviours.[64]

---

[63] Some versions of behaviorism did a better job of handling this concern. Psychological behaviorist, such as Watson (1930) and Skinner (1953) thought that a science of the mind had to be restricted to observable entities and thus excluded the non-observable internal states that the functionalist includes in their definitions. Arguably, certain logical or analytic behaviorists, such as Ryle (1949), were better able to accommodate internal states in their characterization of mental kinds. Rylean behaviorism, also known as 'adverbialism', appealed to the notion of multi-tracked behavioral dispositions, which include references to other mental states (1949: 43-5).
[64] See Putnam's (1963) objection to Rylean behaviourism (1949).

So, on the functionalist account, mental kinds are interdefined and each mental kind is associated with a distinctive input-output relationship. This is what distinguishes mental properties of one kind from those of another kind. In other words, what makes some state count as a mental state of a particular type is the causal role that it plays in the system that it is a part of, rather than, any of its physical properties (or the properties of its constitutive base).[65] This feature distinguishes functionalism from the mind-brain type identity theory about the nature of mental states (advanced by Place 1956, Feigl 1958, and Smart 1959), which identifies types of mental states with types of physical states of the brain and central nervous system.

2.3 Machine functionalism

The earliest form of functionalism, advanced by Putnam (1960, 1967), is known as machine functionalism. On Putnam's model, a mental state is characterized by its role in a 'Turing machine' table, named after Alan Turing, who first described these machines (1936). The Turing machine provides a model of a system whose internal states can be defined in purely functional terms. Turing machines are a class of computers that include any mechanism with a finite number of program states, where the inputs and outputs of the machine are written on a long tape that is divided into squares each of which contain a symbol from a finite alphabet. The machine is able to scan just one square at a time; it can erase the symbol on the scanned square and replace it with a new one, erase the symbol and leave the square blank, leave the square unchanged, or inscribe a symbol on a blank square. A Turing machine's entire set of mechanical operations comprise scanning,

---

[65] Beliefs, desires, hopes, and fears, are all examples of different types of mental states. But one can make a further more fine-grained distinction between different types of belief states, for example, the belief that *p* and the belief that *q*, are different types of beliefs. When I speak of mental kinds or mental types, I have in mind the more finely-grained distinction. But, for the functionalist all of these types are distinguished by distinct input-output relations, where inputs and outputs can also includes other internals states of the system.

erasing, and printing on a square, moving the tape, and changing its own internal state (Fodor 1981).

In order to clarify the notion of a Turing machine, Fodor (1981) gives an example of a simple machine whose states are individuated in functional terms. His example is of a machine that dispenses a Coke for 10 cents.[66] The states of this machine are defined by reference to their causal roles, which includes reference to other internal states of the system. The simple machine has just two interdefined states, *S1* and *S2*. *S1* is defined as the state that the machine is in if, and only if,

(1) when it is fed a nickel it dispenses nothing and proceeds to *S2*, and

(2) when it is fed a dime it dispenses a coke and remains in *S1*.

Meanwhile, *S2* is defined as the state that the machine is in if and only if,

(1) when it is fed a nickel it dispenses a coke and proceeds to *S1*, and

(2) when it is fed a dime it dispenses a coke and a nickel and proceeds to *S1*.

Notice that *S1* and *S2* are defined in terms of the machine's inputs and outputs, and that they are also interdefined, as both of their definitions include transitions to the other state (Fodor 1981). While this example helps to convey the idea of a functional state, any functional definition of a mental state would, of course, be significantly more complex.

2.4 Multiple realizability

The machine analogy illustrates two concepts that are central to functionalism: the first is that mental states are interdefined; the second is that functional properties can be

---

[66] To be sure, the Coke machine is not itself a Turing machine, it is only intended to illustrate how the states of a Turing machine are defined in terms of relations to inputs, outputs, and transitions to other states.

realized by many different kinds of systems (Fodor 1981). Indeed, nothing in the description of the states of the Coke machine puts any constraints on what the machine could be made of (Fodor 1981). This follows from the fact that functional states are defined by their distinctive causal role and have no deeper essence beyond that. This distinct causal role is shared by all instantiations of properties of a particular kind. But the functional definition leaves room for the possibility that mental properties of the same kind could be instantiated by vastly different mechanisms (Kim 2011; Polger and Shapiro 2016). On this matter, Putnam (1960) says:

> The functional organization (problem solving, thinking) of the human being or machine can be described in terms of the sequences of mental or logical states respectively (and the accompanying verbalizations), without reference to the nature of the 'physical realization' of these states. (149)

Thus, it is the fact that a functional characterization makes no reference to the physical realization of mental states that makes it possible for systems other than the brain to realize mental functions (Putnam 1960, 1967). And the possibility of multiple realizers of mental properties was one of the main reasons philosophers came to favour functionalism over preceding theories about the nature of mental states. Multiple realizability allows for the possibility that other species of animals can share mental state types with humans, despite distinct behavioural dispositions (Putnam 1963) or distinct neurobiological constitutions (Putnam 1967). Neither behaviourism nor the mind-brain type identity theory makes conceptual space for multiple realization.[67]

The 'multiple realizability thesis', following Putnam (1967) and Fodor (1974), requires that higher-level properties, in this case mental properties, can possess multiple lower-level realizers or instantiations. Scientifically respectable natural kinds should share similar causal powers. But multiply realizable higher-level kinds have different

---

[67] Putnam (1988) later criticized functionalism precisely because it could not accommodate multiple realizability to the extent that he thought necessary. I will discuss this in the next section.

possible realizers that vary so greatly in their (lower-level) causal powers that they could not belong to a scientifically respectable type when described strictly at that level, that is, independent of the higher-level type that they realize (Fodor 1974). So multiple lower-level instantiations will count as realizers of the same higher-level phenomenon if they have causal structures that, while different, are nonetheless sufficiently isomorphic (with respect to the higher-level entity being realized). Thus, to say that a mental kind is multiply realizable means that it can be realized by nomically different lower-level kinds (Davies 1996: 12-14; Polger and Shapiro 2016). The kind of mind-brain reduction that an identity theorist advocates requires what are known as *bridge laws*, that would enable the reduction of laws at the higher-level to laws at the lower-level by means of law-like generalizations that link the theoretical terms that feature in the higher level laws to the terms that feature in the lower-level laws. In other words, reduction requires entities at the higher level to be identified with the entities at the lower level. When something is multiply realized, lower-level realizers are so nomically different that no bi-conditional between higher-level mental types and lower-level types can be established thus blocking the possibility of any bridge laws (Lyre 2009).[68]

Mental kinds, according to the functionalist, could be instantiated by 'drastically heterogeneous' physical kinds (Lyre 2009). But furthermore, there is nothing in the functionalist framework which rules out the possibility of there being nonphysical realizers of mental kinds (as Putnam 1960 noted). Given this, most contemporary thinkers endorse a version of what Kim (2011: 130) calls the "realization physicalism" thesis which rules out this possibility by stipulating that if something has some mental property *m* then there is some physical thing *t* that has *m* in virtue of the fact that *t* has some physical property *p* that realizes *m* in *t*. Thus, for the realization physicalist all mental properties will be realized in physical systems in virtue of the physical properties of those systems, thereby ruling out the possibility of non-physical substances, such as a Cartesian soul, instantiating any mental properties. So for functionalists who endorse realization

---

[68] See Fodor (1974) for an early defense of multiple realizability in the special sciences and see Kim (1998) for a reply. For a recent comprehensive, but critical review of multiple realizability, see Polger and Shapiro (2016).

physicalism all mental properties are functional properties with physical realizers. But beyond this physical realization constraint, there are no further *a priori* constraints placed on exactly what the physical properties that realize or implement psychological states are, thereby allowing for the possibility of multiple realizers (Kim 2011: 131).

Accommodating the possibility of multiple realizers of mental kinds was Putnam's main motivation for advancing functionalism. But Putnam (1988) later criticized machine functionalism precisely because it could not accommodate multiple realizability to the extent that he thought necessary. States of a Turing machine table are all the states of a system. Thus, by characterizing mental kinds in terms of their role in a Turing machine table, machine functionalism requires that for any two subjects to be in a mental state of the same kind, they would have to realize the very same Turing machine table; that is, their total psychologies would have to be isomorphic. But it would be impossible for creatures with a different sensory apparatus from us to realize the same machine table that we realize. Thus, creatures with different physiology from us would be precluded from sharing the same types of mental states that we have.[69] Because the possibility of multiple realizers of mental kinds was the main objection levelled at mind-brain type identity theory and one of Putnam's motivations for advancing functionalism, it is a major concern that machine functionalism cannot accommodate it. Thus, as a response to this limitation of machine functionalism (among other reasons), there was a shift away from the characterization of mental kinds in terms of Turing machines.

2.5 The Ramsey-Lewis method

David Armstrong (1968) develops a different version of functionalism that moved away from Turing machine tables, instead defining mental states in terms of their causal role. David Lewis (1970, 1972) then developed this version of functionalism in more detail. Following Armstrong, Lewis advocates relying on a comprehensive psychological

---

[69] See Block (1978) and more recent discussions in Polger (2008) and Kim (2011).

theory to characterize definitions of mental kinds, rather than relying on machine characterizations. His proposal followed a methodology introduced by mathematician Frank P. Ramsey, who proposed that theoretical terms could be replaced by existentially quantified bound variables. The Ramsey method was developed by Rudolph Carnap (1947), who used it to define newly introduced theoretical terms (Lewis 1970: 427). Applying this strategy to a psychological theory, Lewis advocates transforming our chosen psychological theory into a 'Ramsey sentence' that existentially quantifies out all reference to mental states or properties. To do this we define the total set of functional states, where one functional state corresponds to each mental state (or property), by means of the Ramsey sentence of the psychological theory (Block 1978). The functional state corresponding to each mental state (or property) is known as the 'Ramsey functional correlate' of that mental state. The first step in transforming our chosen psychological theory, T, into a Ramsey sentence is reformulating it into a single sentence where all mental states terms appear as singular terms. It can be written as:

$$T (p, s_1 \ldots s_n, i_1 \ldots i_k, o_1 \ldots o_m)$$

Here p is the subject, s_1 to s_n are terms for mental states of different types, i_1 to i_k are terms for inputs, and o_1 to o_m are terms for outputs (Block 1978). The next step in 'ramsifying' this sentence is to replace reference to the subject, p, and to mental states, s_1 to s_n, with variables and to existentially generalize over each variable. If y is the variable that replaces p and $x_1 \ldots x_n$ are the variables that replace $s_1 \ldots s_n$, and x_i is a variable standing in for a particular mental state type, say pain, then an example definition of a Ramsey functional correlate for pain, with respect to T, is

$$\exists y \exists x_1 \ldots x_n [T (y, x_1 \ldots x_n, i_1 \ldots i_k, o_1 \ldots o_m) \ \& \ y \text{ is in } x_i]^{[70]}$$

---

[70] Here and throughout this section I am closely following Block's (1978: 269-271) examples of ramsification.

What is noteworthy about the Ramsey functional correlate for pain, or any other mental state (or property), is that it contains no mental terms, though it does contain input and output terms (Block 1978: 270). Thus, using this method, the relevant causal roles for each mental kind are describable as Ramsey sentences, which existentially generalize over each psychological predicate occurring in the law-like statements in our psychological theory so that the resulting sentence contains no psychological predicates.

This Ramsey-Lewis method, as it has come to be known, is central to Lewis's version of functionalism, known as causal-theoretical functionalism (Kim 2011). Like machine functionalism discussed earlier, causal-theoretical functionalism also reduces mental types to their input and output structures. But an advantage of Lewis's proposal is that, because Ramsey sentences contain no psychological predicates, they are able to avoid concerns of circularity that arise from functionalism's commitment to giving holistic definitions of mental kinds.[71]

2.6 Analytic vs. empirical functionalism

When it comes to choosing a psychological theory to specify different causal roles there are competing options. In moving away from machine functionalism and instead defining mental states in terms of their causal role, Armstrong (1968) views his project as a sort of conceptual analysis, which is why it is often known as *analytic functionalism* (Chalmers 1996). In developing Armstrong's view, Lewis (1966; along with Shoemaker 1975) advocates that the causal roles we use for defining mental states should be made up of the platitudes of our everyday views about the mind. His version of functionalism is thus known as *common-sense functionalism*. On the one hand, the platitudes of our common-sense psychology have a great deal of stability and, therefore, the functional roles that they would provide are not likely to change. On the other hand, common-sense

---

[71] A second advantage of causal-theoretical functionalism is that it avoids the chauvinism of machine functionalism, although some charge it of being too liberal, see Block (1978).

psychology is likely to be incomplete; it may contain errors or inconsistencies (see Kim 2011). Furthermore, it is likely to give us only coarse-grained functional roles.

A second option is to look to our best *a posteriori* scientific theory to serve as the ramsification base. In this case we begin with our common-sense understanding as a starting point for empirical investigation into how we should further identify mental properties with functional role properties (see Clark 2008b: 88). This brand of functionalism, known as 'empirical' or 'scientific' functionalism, is adopted by Putnam (1967), Fodor (1968, 1974), and Harman (1973). The primary example of this is 'psychofunctionalism', which looks to the science of psychology to define functional roles (Block 1978: 268).

But there are also concerns regarding empirical functionalism – most of all that both psychology and cognitive science, our main sciences of the mind, are (relatively) young and far less stable than common sense. It would not be an easy task to determine what parts of these sciences are well enough established to be taken as uncontroversially true. Critics point out that even fairly well established areas, such as vision theory, are not without controversy (e.g. Kim 2011: 174-5). Keep in mind what is at stake here: if the underlying psychological theory is false then all the mental concepts defined on its basis by the Ramsey-Lewis method are at risk of having null extensions (Kim 2011: 176). It might be possible to slice contemporary psychology into very fine parts, e.g. fundamental psychophysics, classical conditioning, but appealing to it wholesale is hardly an option. Another concern, which Braddon-Mitchell and Jackson (2007) raise, is that because empirical functionalism looks to the sciences to specify the causal roles, it is likely to provide much finer-grained descriptions of functional roles than common-sense psychology would. The concern is that these fine-grained descriptions will leave no room for multiple realizers of the same mental types—one of the main motivations for moving to a functionalist view of the mind in the first place (Clark 2008b: 240).[72] The question of how finely or coarsely grained descriptions of functional role properties should be, as we

---

[72] I return to this distinction in a later section. For more discussion on the differences between common-sense functionalism and psychofunctionalism see Braddon-Mitchell and Jackson (2007: Chap. 5 and 7) and Clark (2008b: 88-99).

will see, is a major issue for extended mind theorists.

2.7 Role vs. realizer functionalism

There is a further distinction that is orthogonal to the divide between analytic and empirical functionalists, which will come up in later discussions. This is the distinction between *realizer* and *role* functionalists. In order to appreciate this distinction we must distinguish between first-order and second-order properties. As an example, a banana has the *first-order property* of being yellow, or 'yellowness'. It also has the *second-order property* of having a property the linguistic representation of which, in English, starts with the letter 'b'. This second-order property quantifies over its first-order property of being yellow, or 'yellowness' (David 1997). But notice that both of these first-order and second-order properties are properties of the banana. Second-order properties are thus not the same as second-*level* properties, which are properties of properties (David 1997). For example, the property of being yellow, or 'yellowness', has the property of being a color property.

On the functionalist account, if a particular type of mental state, such as the state of being in pain, has a causal/functional role that is realized in humans by a particular type of (first-order) neural property, such as having C-fibres firing (under the right conditions), then for the functionalist a human is in pain in virtue of having C-fibres firing (under the right conditions). In this case the firing of C-fibres occupies, or instantiates, the pain role in humans. But it is still an open question as to whether the property of being in pain is identical to the firing of C-fibres that realizes it or whether it is identical to the second-order relational property of being in some state that plays the right causal role, in this case the pain role. Identity theorists *identify* mental states with the first-order properties that realize them. Realizer functionalism is a version of the identity theory as it makes token-token identity claims between mental states, which are individuated by their causal/functional roles, and the (first-order) neural properties that

realize them. But, it is importantly different from the type-type identity theory, which makes the stronger identity claims between mental types and physical types (these would require bridge laws, as discussed in the section on multiple realizability), not just the token identity claim that realizer functionalism makes.

For the realizer functionalist, therefore, pain is a functionally specified state and every token instance of a human-pain state just is a token instance of C-fibres firing. Thus, a human pain-state is the occupant of the causal role specified by the Ramsey functional correlate of pain with respect to the chosen psychological theory. Role functionalists, on the other hand, identify mental states with the second-order property of having some first-order property (or being in some state) that realizes the relevant causal role. In the case of pain that first-order property might be the neural properties of C-fibres firing but, for the role functionalist, mental states are not token identical with the realizer property; they are token identical with the second-order property (see Kim 2011: 187-8).

Lewis (1980) and Braddon-Mitchell and Jackson (2007) favour realizer functionalism in part because, as a version of the identity theory, realizer functionalism is compatible with the idea that mental causation is just a kind of physical causation and compatible with a physicalist solution to the mind-body problem (Block 1981: 119). Role functionalists can also be physicalists by adopting something like the realizer physicalism position mentioned earlier. But, because they do not identify mental states with their token physical realizers, this leaves them in a position of having to explain how mental events could causally influence physical events, since they are not themselves physical events. Role functionalists sometimes appeal to notions like *downward causation* to address this concern. However, by not positing token-token psychophysical identities, they are put in the position of trying to avoid epiphenomenalism (see McLaughlin 2006). Thus, realizer functionalism is widely thought to offer a better explanation of the causal efficacy of the mental than role functionalism (Levin 2013).[73]

---

[73] Of course the physicalist solution given by an identity theorist is not perfect either. One major problem that confronts it is the so-called 'causal exclusion of the mental' (Kim 1989, 1998). Put briefly, the Kim's causal exclusion argument maintains that if every physical event has a sufficient physical cause, and there is no double causation, i.e. no physical event has two

Many still favour role functionalism, however, as realizer functionalism faces the problem of looking overly chauvinistic in privileging features unique to human cognitive systems (Block 1980). The concern is that realizer functionalism risks collapsing into type-type identity between pain-in-humans and C-fibre-firing. Meanwhile there would be type-type identities between pain-in-other-species and X-fibre-firing. And, critics charge that any more general category of 'pain', which allows us to talk about pain across species, would not be a scientifically respectable category (see Churchland 1981). Thus, realizer functionalism is at risk of being unable to accommodate multiple realizers of mental kinds, which was one of motivations for preferring functionalism over the identity theory.

Realizer functionalists have offered various responses to this concern. One response is to point to a common lower-level *disjunctive* property that is shared by all creatures with lower-level states that occupy the causal role specified by the Ramsey functional correlate of a mental type, such as pain. But ultimately many think that role functionalism does a better job at preserving the intuition behind multiple realizability: that creatures with vastly different physical make-up can nonetheless enjoy the same mental states (Levin 2013).

2.8 Conclusion

This chapter provided an overview of functionalism in some detail and clarified some of the major differences among functionalists. This was necessary for several reasons. First of all, one must appreciate why Clark appeals to the version of functionalism that he does—a coarsely grained common-sense functionalism—as I will discuss in the next chapter. Second of all, it will also be helpful for understanding why Clark's appeal to this version of functionalism is problematic and why a supporter of the parity argument cannot easily appeal to another version. The discussion of functionalism

causes—one physical and one mental, then there cannot be any mental causes.

in this chapter will also help the reader appreciate why functionalism and the parity argument are widely thought to be closely associated.

CHAPTER THREE

FUNCTIONALISM AND THE EXTENDED MIND THESIS

3.1 Introduction

I now turn to the relationship between the parity principle and the parity argument on one hand and functionalism and multiple realizability on the other. I will consider several views about this relationship, beginning with Clark and Chalmers's, which most commentators agree with. They maintain that the parity principle requires functionalism. But, I will also consider several other perspectives. I will then turn to a set of objections that have been raised which assert that functionalism cannot in fact support the extended mind thesis. We will see that appealing to functionalism leads to several dilemmas, which explains why an extended mind theorist might not want to be reliant on functionalism. This will lead me to argue, in the next chapter, that the parity principle and a parity-driven argument does not in fact require functionalism.

3.2 Clark and Chalmers's view

In their original 1998 article Clark and Chalmers do not explicitly claim that the parity argument requires a functionalist view about the nature of mental properties. In later writings, however, they both acknowledge that there is a clear connection between the two, though they have slightly different positions on the nature of the relationship. Clark, as we will see, appeals quite explicitly to a form of common-sense functionalism, while Chalmers is more reluctant to say that the thesis requires functionalism. Thus, I will discuss their views separately.

Chalmers (2008) considers what commitments the extended mind thesis rests on

and quickly resists the idea that the thesis requires functionalism, if functionalism is understood as defining all mental states by their causal roles. He rejects this because he thinks it is "implausible" that conscious mental states can be characterized by their causal roles (xv).[74] He then considers an attenuated version of functionalism in which only certain mental states, such as non-conscious beliefs, are defined by their causal roles. But he says that not even this is needed for the parity argument to go through. His final position is that the extended mind thesis does require some kind of functionalism, what he describes as the "very weak functionalism captured by the parity principle" (xv). Chalmers then restates the parity principle. However, he does not say any more about exactly what kind of functionalism he intends or how this differs from the attenuated version of functionalism that he describes just before it. Chalmers goes on to say that the extended mind thesis does not require much theoretical presupposition at all. He contends that:

> [T]he extended mind thesis is compatible with both physicalism and dualism about the mental. It is compatible with connectionist and classical views, with computational and noncomputational approaches, and even with internalism and externalism in the traditional debates over mental content. (xvi)

I agree with Chalmers that the parity-driven extended mind thesis is compatible with all of these views. That is because, on my view, the parity principle that the argument rests on does not require even a very weak functionalism. I will argue for this in the next chapter, but for now it is enough to note that Chalmers' position is that the parity argument rests on the parity principle, which requires some kind of "very weak" functionalism.

Meanwhile, Clark (2008b) maintains that the parity argument appeals to our common-sense intuitions about mental properties. The argument is meant to show that the kinds of "course-grained functional poise" that common-sense associates with certain

---

[74] See also Chalmers (1996).

mental properties is, at times, instantiated by a nonstandard, or extended, physical realization, for example, by meaningful symbols stored in a notebook (Clark 2008b: 88) Thus, on Clark's view the parity argument requires a version of common-sense functionalism. He even describes the parity argument as "a simple argumentative extension" of Braddon-Mitchell and Jackson's (2007) common-sense functionalism (Clark 2008b: 88).

Clark argues that the underlying theory for functionalism that would best support the parity argument is folk psychology rather than an empirical science, which is why he appeals to common-sense functionalism over empirical functionalism. He cites the concern, raised by Braddon-Mitchell and Jackson (2007), that the sciences of the mind, especially psychology, are likely to provide much finer-grained descriptions of functional roles than the platitudes of folk psychology could provide. For Clark (2008b: 240), the concern is that these descriptions of the functional role properties will be too finely grained to allow for the drastic variation at the lower level that something like Otto's notebook would require.[75] Clark even labels his view 2008b: extended common-sense functionalism" (96; or "extended functionalism" for short; Wheeler 2010). He explains that the extended mind thesis first appeals to folk psychology for a coarse-grained role, and then searches for a more fine-grained description of the actual mechanisms that physically realize that coarse functional role, wherever they might be (88-9). Standard common-sense functionalism would appeal to folk psychology for the causal role and then, in contrast, would look to the biological system, particularly the brain, to find what mechanisms fulfil that functional role.

It is worth emphasizing that in the version of functionalism that Clark adopts, it is the *coarse-grained* common-sense functional role that dictates what is essential to any given mental property and the fine-grained descriptions matter only secondarily. (Recall that functional properties only have nominal essences, i.e. they are defined by their distinctive causal role and have no deeper essence beyond that.) And the parity argument

---

[75] Some disagree with this concern, e.g. Sprevak (2009; 2010). I will discuss this disagreement in a later section.

asserts that in certain cases the physical mechanisms that turn out to realize the coarse-grained role of a belief could be extended, or wide, in the sense that they include objects beyond the biological body, as is the case with Otto and his notebook.[76] Thus, extended common-sense functionalism does not restrict its search for the relevant physical mechanisms to the brain.

Finally, extended common-sense functionalism is likely to be a kind of *role* functionalism. So Otto's belief that the museum is located in 53rd street is identical to the second-order relational property of being in some state that plays the right causal/functional role—the same causal/functional role that Inga's belief about the location of the museum plays for her. However, extended common-sense functionalist might also be compatible with realizer functionalism. That is, an extended common-sense functionalist could make a token-token identity claim between Otto's belief, which would be individuated by its causal/functional role, and the (first-order) physical properties of the notebook that realize them. The risk of adopting realizer functionalism, as Block (1980) points out, is that it might be too chauvinistic for the extended functionalist. Realizer functionalism might privilege features unique to human neurobiological systems that would lead us to draw distinctions between beliefs-in-human-brains, such as Inga's beliefs, and other *types* of *extended* beliefs, such as Otto's, thereby preventing us from saying that Otto and Inga truly have the same type of belief.

---

[76] Farkas (2012: 441) disagrees with this interpretation. She argues that the extended mind thesis is really about metaphorically 'extending' the relevant functional *role*: "we extend what counts as a functional role or a dispositional profile that qualifies a state to be a certain kind of standing state…this is the real lesson of the Otto-Inga case." On this view the externalism of the vehicle is "inessential" to Otto's story and to the extended mind thesis (437). Thus, the version of the extended mind that Clark supports involves updating what are thought to be the traditional realizers of cognition, while the version that Farkas advocates involves modifying the functional role itself. I find Clark's version to be more plausible and will explain why in a later chapter where I discuss Farkas's view in more detail.

3.3 The relationship between functionalism and the extended mind

Consider what others have said about the relationship between functionalism and the extended mind thesis. Most commentators agree with Clark and Chalmers (1998) that their version of the parity argument, and indeed any parity-driven argument, rests on functionalism.[77] But does functionalism entail the extended mind thesis? Most stop short of this claim (as Drayson 2010 notes). Rupert (2004: 421), for example, says, "functionalist theorising alone does not resolve the issue of extended states." And Shapiro (2008: 14) argues that an argument for the extended mind thesis "is not going to follow *a priori* from a theory of mind."[78] But a few commentators do take this further step. Weiskopf (2008), for example, says not only that the extended mind thesis rests on a functionalist principle but also that the thesis is more or less implied by functionalism. The extended mind thesis "is simply an instance of an unusual realization of a mental state, and thus a fixture familiar to functionalists" (267). On his view, the thesis is not "especially radical" precisely because it follows from the commitments of functionalism, a view that has been popular for some time (267). Sprevak (2009) agrees that (certain versions of) functionalism entails that the mind is extended. Sprevak turns this into an argument against both the extended mind thesis and functionalism. I will discuss his objection in the next section.

---

[77] Explicit statements of the dependence of the extended mind on functionalism can be found in: Wheeler (2010a); Drayson (2010); Rupert (2004); Gertler (2007); Shapiro (2008); Weiskopf 2008; Sprevak (2009); Rowlands (2010); Farkas (2012); Miyazono (2015); Wadham (2016).

[78] In contrast to Sprevak, Shapiro (2008: 5) argues that functionalism does not imply the extended mind thesis; rather it is silent on the issue: "functionalism does not have the means by which to distinguish the realization of a mental state from causal influences on the realization." He concludes that functionalism cannot *a priori* adjudicate whether or not a mind is partially constituted by objects in the environment or merely causally influenced by those objects: both the externalist and the internalist picture is consistent with a functionalist theory of mind and only empirical investigation can settle the issue. Shapiro calls this the "boundary problem (10). That is to say, once a functionalist conception of the mind is agreed upon, the world could turn out to be consistent with either intracranialism or the extended mind thesis. So while the extended mind thesis is modally possible, we can only settle if minds are actually extended by looking at the current state of the world.

3.4 Problems with relying on functionalism

If the parity argument relies on functionalism it runs into several problems. First of all, although functionalism is a longstanding and popular view in philosophy of mind and philosophy of cognitive science, it is still a controversial theory that faces many equally longstanding objections. Any theorist who embraces functionalism must confront these objections, e.g. Searle's (1980) Chinese room argument, Block's absent qualia (1978), and the problem of mental causation.[79] Many of these objections are well known and much has been said in response to them. I will leave this aside, as an in-depth discussion will be too far afield for current purposes. But it is worth pointing out that if the parity argument for the extended mind thesis relies on functionalism (either common-sense functionalism *or* empirical functionalism) it would also confront these objections.[80]

When we consider how functionalism could support the parity argument for the extended mind thesis we reach a dilemma. While Clark opts for a coarsely grained common-sense functionalism to support the parity argument, several commentators have argued that this version of functionalism cannot support the extended mind thesis without leading to serious problems. I will consider these objections in the next section. This will lead me to consider whether the extended mind theorist could instead opt for another version of functionalism, such as empirical functionalism. But, as we will see, this too would lead to problems. Critics charge that, no matter which science of the mind one appeals to, empirical functionalism appears to be too finely grained to support cases of cognitive extension. Thus, there are several objections confronting extended mind theorists who appeal to functionalism.

---

[79] See also Block (1990) and Kim (1998, 2011).
[80] As noted by Rupert (2004: 424).

3.5 Problems with relying on common-sense functionalism

I will now turn to several objections that focus on the relation between common-sense functionalism and the extended mind thesis. The first concern, raised by Rupert (2004), is that the extended mind thesis would imply counterintuitive attributions of mental states that our common-sense conceptions would preclude rather than support. The second objection comes from Sprevak (2009), who argues that none of the glue-and-trust conditions are independently necessary and that without them common-sense role functionalism implies a radical version of the extended mind thesis. Finally, a third objection comes from Wadham (2016), who argues that having independently necessary glue-and-trust conditions is not in keeping with the doctrine of common-sense functionalism that Clark appeals to.

3.5.1 Rupert's objection

Clark appeals to common-sense functionalism to support the extended mind thesis. But Rupert (2004) argues that the extended mind thesis simply violates our common-sense views about where cognition occurs.[81] For example, he maintains that it is counterintuitive to attribute to Otto an extended belief about the location of the museum. Thus, the extended mind thesis cannot appeal to common-sense functionalism, as it would involve counterintuitive attributions of mental states that our common-sense conceptions would preclude rather than support. But, Rupert acknowledges that this criticism will be lost on those who do not share his intuitions. Thus, a stalemate looms (2004: 406). One way to resolve this is to appeal to other reasons, e.g. empirical fruitfulness, to prefer the one position to the other. There are several normative reasons, for example, which have since been offered in favour of the extended mind view. As a

---

[81] This is not the only objection that Rupert (2004) raises against the extended mind thesis. I will consider some of his other concerns in a later section.

few examples, it has been suggested that the extended mind thesis can better protects against harm to the mind (Levy 2007), that it better accounts for compensatory rehabilitation as a way of repairing the mind (Drayson & Clark forthcoming), and that better addresses concerns about the way we assess the capacities of learning disabled individuals (King 2016). I will not delve into these reasons in any more detail.

3.5.2 Sprevak's objection

Sprevak (2009) argues that the coarse-grained functionalism that Clark appeals to entails that the mind is *radically* extended. That is, it implies cases of extended mental states that are intuitively implausible—so implausible that Sprevak argues this entailment amounts to a *reductio* of functionalism. Recall that Clark and Chalmers offered the glue-and-trust conditions to characterize the functional role of the stored information in the Otto and Inga cases. In response to this, Sprevak offers a series of counterexamples to show that none of these conditions are independently necessary. He argues that, for each condition, we could imagine a Martian whose cognitive processing did not meet that condition and yet, intuitively, we would still want to accept the process as a genuinely *cognitive* process of the Martian. Sprevak (2009: 507-8) calls this the 'Martian intuition':

> The Martian intuition claims that it is possible for creatures with mental states to exist even if such creatures have a different physical and biological makeup than ourselves. …

> The Martian intuition applies to fine-grained psychology as well as physiology. There is no reason why an intelligent Martian should have exactly the same fine-grained psychology as ours. A Martian's pain response may not decay in exactly the same way as ours; its learning profiles and reaction times may not exactly match ours; the typical causes and effects of its mental states may not be exactly

the same as ours; even the large-scale functional relationships between the Martian's cognitive systems (e.g. between its memory and perception) may not match ours.

With respect to the first condition, the constancy condition, for example, Sprevak invokes the Martian intuition by arguing that one could imagine a Martian with internal cognitive resources that are not typically invoked:

> The Martian might have cognitive resources that are only available after it gets a good night's sleep, and it does not reliably or often get a good night's sleep. However, that does not stop, on those occasions when the Martian does get a good night's sleep, those resources from counting as genuinely cognitive. (514)

Sprevak raises similar objections against each of the other conditions. With this strategy, he argues that the glue-and-trust conditions must be dispensed with entirely, as none are independently necessary.

With the glue-and-trust conditions removed, Sprevak concludes that the coarse-grained functionalism which Clark appeals to implies a "rampant expansion" of the mind into the world (503). For Sprevak, this rampant expansion is indicative of a deeper problem with functionalism: it entails all sorts of implausible scenarios. For example, when a person steps into a library or uses the internet all the contents of these systems would be considered as part of that person's extended cognitive system—as their extended non-occurrent beliefs, just as the information in Otto's notebook would count as his extended non-occurrent belief. Thus, either our minds are currently *radically* extended or the (coarse-grained) functionalism it rests on is false.

There have been several responses to Sprevak's objection. One response comes from Drayson (2010) who agrees that certain versions of functionalism entail that minds extend, but disagrees that the extension would be so radical as to count as a *reductio* of

functionalism.[82] Drayson disagrees with Sprevak's portrayal of Clark and Chalmers's position. In addition to the Martian examples, Sprevak also argues that there are examples of internal cognitive processes in humans that do not meet the conditions and, thus, for this reason as well, the conditions cannot be accepted as independently necessary. For Drayson the problem with this line of argument is that it is not clear that Clark and Chalmers ever intended these as conditions for a cognitive process generally speaking; they only describe them as conditions on non-occurrent beliefs. Thus, Drayson argues that the conditions are only intended to highlight the "ways in which the contents of Otto's notebook have the same functional poise as our own (and Inga's) normal dispositional beliefs." (372). In this case it is irrelevant whether there are other forms of cognition (e.g. desires, visual processing, imagination) that do not meet these conditions. Sprevak would need to show that there are non-occurrent beliefs that do not display these features. Thus, Drayson's view is that common-sense functionalism implies a more moderate version of the extended mind than Sprevak defends.[83] Another response to Sprevak, which I will now turn to, comes from Wadham (2016). But Wadham's response is not a defense of Clark's position. It is, in effect, another objection to the extended mind theorist's reliance on functionalism.

3.5.3 Wadham's objection

Wadham argues that both Clark and Sprevak misunderstand Jackson and Braddon-Mitchell's common-sense functionalism. Clark and Sprevak treat the glue-and-trust conditions as independently necessary (and jointly sufficient) for functional equivalency, but Jackson and Braddon-Mitchell (2007) maintain that there are no individually necessary features that characterize a functional role. Jackson and Braddon-Mitchell hold that there may be some set of features that, satisfied together, are necessary for something

---

[82] For two other responses to Sprevak see Milojevic (2013) and Miyazono (2015).

[83] I will have a lengthier discussion of Drayson's position in a later section.

to be said to instantiate or play a given functional role, but that no feature is by itself *independently* necessary. Wadham argues that, because of this misunderstanding, both Clark and Sprevak's argumentative methodology is misguided. Instead of showing that some given condition is or is not independently necessary for a non-occurrent belief, an extended mind theorist should be trying to show that "the sum total of respects" in which Otto and his notebook are functionally different from Inga do not, even "when taken together, or in any combination" undermine the claim that the two are functionally equivalent (Wadham 2016: 148). And a critic should be arguing that, when taken together, the respects in which Otto and his notebook are functionally different from Inga, *do* undermine functional equivalency between the two cases. Wadham concludes that the common-sense functionalism that Clark relies on does not imply that our minds are extended *at all* and certainly not radically extended, as Sprevak tries to show. If Wadham is right, then either Clark needs to give a principled reason for altering the version of common-sense functionalism he appeals to or the extended mind thesis will need to be supported in a new way—without functionalism or by appealing to a different version of functionalism.

Along the lines of the response that Drayson gave to Sprevak, one response that Wadham considers is that the glue-and-trust conditions were never intended as independently necessary conditions. But, to respond to Wadham's objection, one would have to argue that the conditions are not intended as necessary even for a non-occurrent belief. So far as I can tell, Clark and Chalmers never explicitly describe their conditions as independently necessary. For example, in explaining the intent behind the conditions, Clark (2008b: 79) says, "In response to the concerns about availability and portability, we [i.e. Clark and Chalmers 1998] then offered a rough-and-ready set of additional criteria to be met by non-biological candidates for inclusion into an individual's cognitive system." And in a later paper Clark says: "The functional poise of the stored information was [in Otto's case] sufficiently similar (we argued) to warrant similarity of treatment" (2010a: 45). But if the conditions are not intended as independently necessary, then at least two other concerns rear their heads. First, if these conditions are only intended as "additional

criteria" on non-occurrent beliefs, then they begin to look ad hoc. They look ad hoc because they fall outside the doctrine of common-sense functionalism and yet serve an important role in the argument for a moderate version of the extended mind. And, secondly, these "additional criteria" would risk question-begging, since they have been carefully selected by those arguing for the extended mind.

3.6 The functionalist dilemma

So far we've seen that Clark (2008b) appeals to common-sense role functionalism to support the extended mind thesis to avoid an overly fine-grained description of the functional role. And he provides the glue-and-trust conditions to begin to characterize the coarse-grained functional poise of a non-occurrent belief (of the sort that Otto has). But we've just seen three objections to this strategy. Rupert (2004) argues the extended mind thesis implies counterintuitive attributions of mental states that conflict with our common-sense views. Sprevak's (2009) argues that the glue-and-trust conditions can be undermined and that without them common-sense role functionalism implies a radical version of the extended mind thesis. Finally, Wadham (2016) argues that having independently necessary glue-and-trust conditions is not in keeping with the doctrine of common-sense functionalism that Clark appeals to (that of Braddon-Mitchell and Jackson 2007).

Given these objections to Clark's appeal to common-sense functionalism, it is worth exploring the possibility of a parity argument that relies on a different version of functionalism. Here the question of how finely or coarsely grained descriptions of function role properties should be becomes a major issue for extended mind theorists. One possibility is to opt for empirical functionalism in place of common-sense functionalism. But what we will see is that by relying on empirical functionalism more problems arise for the extended mind theorist. The result is a "deadlock" between extended and non-extended functionalists. By relying on common-sense functionalism extended mind theorists risk being overly liberal and begging the question against their

opponents. But, by insisting on empirical functionalism and appealing to our latest psychological theories to give us fine-grained functional profiles of mental states, the non-extended functionalist risks being overly anthropocentric and bio-chauvinistic, thereby begging the question against extended mind theorists.

3.7 Trying on empirical functionalism

Let's begin by trying on empirical functionalism in place of common-sense functionalism. We saw that Clark's reason for appealing to common sense over empirical science was the concern that empirical functionalism will provide overly fine-grained descriptions of the functional role properties to allow for the variation at the lower-level that something like Otto's notebook would require.[84] Sprevak argues that the question of how finely or coarsely grained descriptions should be is one that confronts both the empirical functionalist and the common-sense functionalist alike.[85] For the empirical functionalist it will depend in part on which science one appeals to. One might appeal to modern psychology, cognitive science, or neuroscience, for example, and each field provides a variety of different levels and grains at which the relevant mechanisms and processes they study are cast. But Sprevak concedes that, given the available sciences, empirical functionalism is likely to be too fine-grained to support the extended mind thesis.[86]

Several critics of the extended mind actually appeal to empirical functionalism to refute the thesis. Rupert (2004), for example, argues that neither common-sense functionalism nor empirical functionalism would support the extended mind thesis. As mentioned, he rejects the possibility of relying on common-sense functionalism on the grounds that extended mental states would betray our folk psychological conceptions.[87]

---

[84] Clark (2008b: 240); discussion in Braddon-Mitchell and Jackson (2007: Chap. 5).

[85] Sprevak (2009: 510-1).

[86] Drayson (2010: 373) agrees.

[87] See Rupert (2004: 423; 2009: 91-4; 2010: 345).

He then considers an appeal to empirical functionalism but argues that the explanatory kinds that contemporary cognitive science currently describes are not likely to have extended realizers because they will be too finely grained.[88] Rupert points to empirical psychological data that shows how writing information down rather than simply reading it, for example, can help improve one's memory of the information. This is known as the "generation effect" as the act of generating the information in writing has an effect on one's memory of the information. Rupert argues that the generation effect is an example of a fine-grained feature of internal memory that would not be present in extended 'memory' systems.[89] He further argues that if the extended mind thesis is intended as a theoretical postulate of cognitive science then to evaluate it, one must consider how valuable the empirical work that depends on it is. But he charges that the extended mind theorist postulates cognitive kinds at a level that is so coarse that they fail to account for a range of phenomena studied by cognitive scientists. Furthermore, Rupert maintains that the theory fails to add any causal-explanatory value to standard cognitive science and thus it should not be accepted as a postulate of cognitive science. Rupert concludes that *neither* common-sense nor empirical functionalism can support the functionalist argument for the extended mind thesis.

Two other critics, Adams and Aizawa (2001, 2008) similarly appeal to contemporary cognitive science to show that the grain of functional roles is too fine to support the extended mind thesis. Adams and Aizawa insist that how mental types are to be discriminated is an issue of primary importance. And, according to them, cognitive science tells us that mental types must be differentiated on the basis of their underlying causal powers. In contrast, the common-sense functionalism that Clark appeals to considers this issue (of how specific mechanisms actually instantiate functional roles

---

[88] Although Rupert (2004:423) also says that some extended mind theorists do try to appeal to empirical functionalism over common-sense functionalism, but he does not indicate which philosopher(s) he has in mind. It seems clear that Clark (2008b) at least appeals to common-sense.

[89] Rupert (2004: 415-8); further discussion in Wheeler (2010a).

within a given physical system) to be a secondary matter.[90] Discriminating mental types on the basis of their underlying fine-grained causal properties causes a problem for the extended mind thesis because whichever empirical science one appeals to is likely to be focused on understanding terrestrial human psychology more than anything else. This focus is likely to lead to anthropocentric, and especially neurocentric, cognitive theories that will surely restrict the possibility of extended states.

I think Clark is, thus, right to be concerned that by focussing on (internal) human non-occurrent beliefs one will end up requiring that all beliefs of the same type exhibit all the same "idiosyncratic features of terrestrial neural activity" (2008b: 93). Again, this is why Clark opts for common-sense functionalism over empirical functionalism, and why his critics use an empirically informed version of functionalism to argue against the extended mind thesis. Much like Clark, Wheeler (2010a), a supporter of the extended mind thesis, argues that the cognitive science Rupert (2004: 10) appeals to is a "conventional human-oriented and inner-oriented cognitive psychology." He labels this a "chauvinistic" form of functionalism and instead advocates for a "liberal" form, of the sort Clark adopts, which gives a broader and more inclusive functional profile of mental states (Drayson 2010: 370). This exchange leads to what Wheeler (2010a: 11) calls "the Rowlands deadlock", which I will now turn to.

3.8 The functionalist deadlock

Rowlands (2010) argues that there is no straightforward way (without begging the question) to settle this dispute between extended and non-extended functionalists over how finely or coarsely grained our descriptions of the relevant causal roles of psychological states should be. Thus, by relying on a liberal form of functionalism the extended mind theorist is "vulnerable to charges of question-begging" (2010:105; similar remarks in Shapiro 2008: 10). But by appealing to a conservative form of functionalism

---

[90] Clark (2008b: 89).

83

critics of the extended mind thesis, such as Rupert, and Adams and Aizawa, are likewise vulnerable of begging the question in favour of intracranialism. For Wheeler the 'Rowlands deadlock', as he labels it, implies that "to the extent that [extended cognition] is allied to functionalism, the best it can achieve against its critics is a stalemate." (2010a: 11) Wheeler tries to overcome the deadlock by offering a reason (that does not beg the question) for preferring a liberal functionalism to a chauvinistic one. The generation effect, Wheeler argues, is merely an accidental feature of (human) memory, not an essential one, and thus the difference between exhibiting or failing to exhibit the generation effect, for example, does not mark the boundary between having a memory and not having one. This is confirmed by the fact that cognitive psychologists may find just such a functional difference worthy of investigation (Wheeler 2010a), thereby supporting the liberal version over the chauvinistic one. Thus, from the fact that some particular fine-grained feature has received attention from contemporary cognitive scientists one cannot infer that it should be included as a necessary feature of all non-occurrent beliefs.[91]

But, as Wheeler notes, this does not quite resolve the deadlock as the non-extended functionalist might just as well offer her own reasons, independent of any internalist commitment, for rejecting liberal functionalism in favour of chauvinistic functionalism at which point the deadlock would be reinstated. Thus, Wheeler further argues that even "microfunctionalist" connectionism, which "specifies a system only in terms of input-output profiles for individual units and thus is not crucially dependent on any particular biological substrate"[92], could support the extended mind thesis since it preserves the functionalist commitment to multiple realizability. This, Wheeler argues, indicates that the issue of the appropriate grain of functional analysis is orthogonal to the issue of cognitive extension (2010a: 16). What really matters, according to Wheeler, is the preservation of multiple realizability. I will return to Wheeler's view on the relationship between the parity argument and multiple realizability in a later section.

---

[91] See Rupert (2004: 418-21; 424) for a replies reply to the line of argument put forth by Wheeler.
[92] Clark (1999: 40), quoted in Wheeler (2010a: 16).

The upshot, for now, is that no matter how one slices it, relying on functionalism is problematic for the extended mind theorist. Extended mind theorists who rely on functionalism must first overcome Rupert's dilemma by finding either a folk psychological theory or a scientific one that will support their argument. Then they must get around the so-called 'Rowlands deadlock' between extended and non-extended functionalists by finding a non question-begging way of settling how finely or coarsely grained our descriptions of the relevant causal roles of psychological states should be. And, even if Wheeler is right that we have independent reasons for adopting a liberal over a chauvinistic form of functionalism, relying on a liberal brand of functionalism leads to a further obstacle for extended mind theorists. Liberal functionalism threatens the radicalization problem that Sprevak (2009) raises. And if Wadham (2016) is right about how to understand Braddon-Mitchell and Jackson's functionalism then Clark's glue-and-trust conditions, which might prevent this radicalization, are ad hoc add-ons, which beg the question in his favour. If the parity argument requires functionalism to get off the ground, the argument may be doomed.

3.9 Non-functionalist, non-parity arguments for an extended mind

Given the widely held view that a parity-driven argument must rely on some variety of functionalism, some extended mind theorists have responded to these functionalist-related concerns by developing non-functionalist arguments for the extended mind thesis that move away from considerations of parity entirely. In Chapter One I explained how, as a consequence of this move away from parity, some of these arguments end up supporting versions of cognitive extension that look quite different from Clark and Chalmers's view. Rowlands (2010), for example, cites the deadlock as one of his motivations for developing a non-functionalist version of the extended mind thesis. In other words, he is concerned that a parity-driven argument requires functionalism and that this weakens the case for cognitive extension. In attempting to

avoid functionalism, however, Rowlands' argument instead relies on Heidegger's notion of intentionality, which faces its own set of problems.[93]

As another example, DiPaolo (2009) appeals to enactivism, a view that rejects core commitments of traditional cognitive science including representationalism, computationalism, and functionalism (as discussed in Chapter One), to support cognitive extension. For some, this rocks the boat too much. One attractive feature of Clark and Chalmers's argument is that it is compatible with so much of traditional cognitive science. On my interpretation of their view, as we will see, it is primarily a challenge to intracranialism. The enactivist version of cognitive extension that DiPaolo supports also risks, at least by many people's standards, being too liberal as it is would count more external resources as genuine cognitive extensions than the parity-driven version, which already faces objections from over-extension (e.g. Gertler 2007).

More recently, Palermos (2014) looks to dynamical systems theory to provide two distinct arguments in support of a version of extended cognitive systems that treats continuous mutual interactions (or "loops") not just as sufficient but also as necessary for extension. In addition to avoiding functionalism, Palermos' *continuous reciprocal causation* argument also does not rely on the possibility of multiple realizers of the same mental kind. On his view, continuous mutual interactions allows for external cognitive processes that could never have occurred internally, given the limitations of our biological resources. So it does not *require* that the very same mental states can occur internally or externally only with different realizers. But, Palermos' argument is also more restrictive than Clark and Chalmers's version, at least in some respects. On his account, for example, Otto and his notebook would not count as a case of extension since Otto does not "mutually interact" with the notebook.[94]

---

[93] Rowlands is also motivated by an apparent tension between the embodied mind thesis and the liberal functionalism that the extended mind thesis relies on that Clark (2008a, 2008b: 198-217) points out, which I will not discuss here.

[94] Menary's (2007) cognitive integration approach also does not rely on multiple realizers. Clark expresses an openness to the possibility that (a non-parity driven version of) the extended mind thesis does not need to involve multiple realizability. He says that the parity argument is intended

As a final example, Sutton (2006, 2010a,b) avoids appealing to functional equivalence by instead requiring the appropriate complementarity of inner and outer processes. Sutton's complementarity argument strikes a balance (between being too liberal or too restrictive) by placing limitations on what kinds of thing count as genuine extensions: only those outer processes that sufficiently complement ongoing inner processes in the execution of some cognitive task count as extensions. Like Palermos' argument, Sutton's does not seem to rely on multiple realizers. His argument depends on outer processes being appropriately connected to ongoing processes so as to 'complement' them in their completion of some cognitive task. Thus, complementarity allows us to build on our internal capacities by using external resources to achieve a new range of cognitive capacities, which we could not achieve without the external processes. But again, the range of cases that the complementarity argument supports is still different from what Clark and Chalmers advocate. While the parity argument maintains that cognitive processes that we know can occur internally could also occur externally, the complementarity argument allows for an expansion, or a new range, of cognitive processes.

Thus, several others have developed non-functionalist arguments for the extended mind thesis, often in response to what they see as a problematic commitment between functionalism and parity. But these versions differ in important ways from the influential version of the thesis that Clark and Chalmers put forth. The central purpose of this thesis is to explore how Clark and Chalmers's parity-driven view interacts with, or is related to, functionalism. In particular I will argue that parity is not committed to functionalism as these second-wave extended mind theorists have thought. I therefore will not have further discussion of these other arguments for the extended mind. The focus hereafter will be on the parity argument.

---

to "provide the essential first means by which to begin to break the stranglehold of vehicle-internalist intuitions concerning cognition." (2008b: 99) Once the possibility of vehicle externalism is established, "all kinds of process that have no fully [internal] biological analog" should also count (99).

3.10 Conclusion

The main objective of this chapter has been to consider the relationship between the parity argument and functionalism. Clark and Chalmers, and many other commentators, have said that the parity argument requires functionalism. But when we consider how functionalism could support the extended mind thesis we reach a dilemma. We saw that while Clark opts for a coarsely-grained common-sense functionalism to support the parity argument, many commentators have argued that this version of functionalism cannot support the extended mind thesis without leading to serious problems. But, should an extended mind theorist opt for another version of functionalism, such as empirical functionalism, other problems arise. No matter which science of the mind one appeals to, empirical functionalism appears to be too finely grained to support cases of cognitive extension. In the next chapter I will argue that the parity principle does not require the truth of functionalism or even the multiple realizability thesis because they rest on theoretically distinct issues. This will be the first step in advancing a non-functionalist parity-driven argument for the extended mind thesis.

CHAPTER FOUR

PULLING FUNCTIONALISM AND PARITY APART

4.1 Introduction

We've seen that most commentators agree with Clark and Chalmers that their parity argument rests on functionalism, while others further argue that certain versions of functionalism actually entail the extended mind thesis. But there is a third position, defended by Wheeler (2010b) and Drayson (2010), that maintains that while the original version of the parity argument relies on some kind of functionalism, a non-functionalist parity-style argument could be given that instead relies only on the thesis of multiple realizability (operating with some other definition of the nature of mental kinds). This position relies on the fact that mental types, being multiply realizable, are compatible with a non-functionalist conception of mind. In this chapter I develop this position and consider how a non-functionalist parity-driven argument might go. Finally, I will outline a fourth position: that it is possible to extend the mind even without relying on multiple realizers. This will set the stage for the next chapter where I present an argument for this position.

4.2 Separating the multiple realizability thesis from functionalism

Multiple realizability and functionalism share the same founding father: both were developed by Hilary Putnam (1960, 1967). Their shared history may explain why they are often thought to go hand-in-hand. But the multiple realizability thesis does not require a commitment to a functionalist conception of mental types. While functionalism is a view about the *nature* of mental states, multiple realizability is a thesis about the *composition*

of mental states: it is a claim about the how types of one kind (e.g. mental states) relate to the physical substrates that constitute or instantiate them. The multiple realizability thesis about mental types says that the composition of higher-level mental types is not fixed—that is, according to multiple realizability there is a one-to-many mapping relation between mental types and the lower-level physical types that compose them.

We've seen that the functional characterization of mental types abstracts away from reference to physical properties such that particular mental types could turn out to have multiple kinds of physical instantiations (or realizers). This abstract functional definition, Putnam argued, allowed for the possibility that mental types could be multiply realized. But notice that it could still turn out that each type of mental state has only one type of lower-level physical instantiation, just as a matter of contingent fact. This is one sense in which functionalism and multiple realizability might be said to come apart, as functionalism could be true even if there are not in fact multiple realizers of any mental states. Thus, while functionalism may entail the possibility of multiple realizability (that is, understood modally) it does not entail that mental states are in fact multiply realized. This is the position defended by Bechtel and Mundale (1999), for example, who maintain that functionalism can be true even if mental types are not multiply realized. So, even if it turns out that all mental types are realized by the same lower-level kinds this would not render the functional characterization of mental states any less important.

To develop a non-functionalist parity-style argument that still appeals to multiple realizability, one needs to preserve multiple realizability without relying on functionalism. To this end, it is worth making clear that while functionalism nicely accommodates multiple realizability, the latter does not require the former. What is key for the possibility of multiple realizability is not a functional characterization of mental types; rather, it is having a definition of the nature of mental types that abstracts away from the details of their physical makeup in a way that allows for a one-to-many mapping. Thus, to move away from functionalism while still preserving multiple realizability, one must define mental kinds by something other than their functional properties or the physical properties of their realizers in a way that will allow for vastly

different lower-level kinds to instantiate the same mental kind. Some examples will help clarify this point. A phenomenalist, for example, defines mental states in terms of their phenomenal properties. Phenomenal property types could turn out to be realized by different types of physical stuff. Thus, a phenomenalist could accept that (at least some) mental types are multiply realizable, even while rejecting a functionalist characterization. As another example, Paul Churchland (2005) rejects functionalism but argues that the multiple realizability thesis is almost certainly true. For Churchland it is not a function that characterizes mental states of a particular type, but the broad kind of hardware, i.e. their computational mechanisms, that they have in common.[95] Thus, the multiple realizability of mental kinds does not require functionalism.

It is also worth considering how Putnam (1988) uses multiple realizability as an objection against (his own) Turing-inspired machine functionalism. Recall that machine functionalism identifies mental states with computational states of a suitably programmed Turing machine. And Putnam had argued that mental states (and, thus, computational states) are '*compositionally* plastic', in the sense that they can be realized by different kinds of physical stuff. This was his original claim that supported the multiple realizability of mental kinds. But Putnam later recognized that mental states can also be instantiated by Turing machines in different kinds of computational states, thus making mental states '*computationally* plastic'.[96] The computational plasticity of mental types means that mental kinds could not be identified with computational kinds and, thus, machine functionalism had to be rejected. And so, according to Putnam, the computational version of functionalism is faced with an objection from multiple realizability just as the theory it was meant to replace – the mind-brain identity theory – had been. Putnam's eventual rejection of functionalism shows why the multiple realizability thesis about mental kinds needs to be treated as distinct from functionalism: in order to appreciate his concern we have to understand multiple realizability as an

---

[95] Churchland (2005: 25-6) notes the irony of this, given that the multiple realizability thesis is often motivated by the diversity in hardware across individuals and across species.
[96] Discussion in Bickle 2013.

independent thesis that can be preserved without functionalism.[97]

4.3 Parity without functionalism

Before turning to multiple realizability, I argue that the parity principle does not rely on functionalism. This will be the first step in explaining why multiple realizability is really the more central idea. Recall that in defending the extended mind thesis Clark and Chalmers rely on the parity principle, which they state as follows:

> 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 part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process. (Clark and Chalmers 1998: 8; restated in Clark 2008: 77)

This statement of the principle clearly appeals to functional similarities between inner and outer resources so it is easy to see why one might think that the principle is committed to some form of functionalism. But there is another way of interpreting the parity principle that separates it from functionalism and emphasizes instead the 'fair treatment' aspect of the principle. Consider how Clark and Chalmers state their principle at a later point in their article, "What makes some information count as a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body." (1998: 14) The first conjunct is more or less a statement of the basic functionalist commitment that to be a belief is just to play the right functional role.[98] The second conjunct, in contrast, emphasizes that, for the functionalist, the location of some

---

[97] As we have seen, functionalists have branched into many different kinds, but the same style of argument could be raised against various kinds of functionalism. For example, against role functionalism, which identifies mental states with their functional roles, one could argue that different kinds of functional roles realize the same type of mental state (Bickle 2013). See Polger and Shapiro (2016) for a reply to this computational argument for multiple realizability.
[98] Here I agree with Weiskopf's discussion of this passage (2008: 266).

mechanism that realizes the causal role of a mental kind should not matter. That is, functionalism is locationally uncommitted. Advocating location neutrality, I argue, is the main contention of the parity principle and this does not require functionalism.

On this reading the parity principle states that we should treat equivalent processes with "the parity they deserve", irrespective of whether they are internal or external to the skull. Why should we do this, according to Clark and Chalmers? Because, they say, what is relevant in deciding the status of processes as mental or otherwise is the *functional role* that they play. But, one could appeal to some other unifying or defining feature of the mental, besides functional role similarities, by which to judge parity. One could appeal to equivalency in phenomenal properties, for example.[99] As another example, a connectionist could appeal to equivalency in activation patterns between entities with similar connectionist architectures. Thus, when Clark (2008b: 77) says we look to this principle to assess whether there should be "parity of treatment based on the significant commonalities rather than simple prejudices about skin and skull, inner and outer", the principle itself remains open on exactly what commonalities should count as "significant" (77). For the functionalist, they will be functional commonalities, no doubt; but a non-functionalist could substitute in for the features they take to be relevant.

This interpretation suggests that by maintaining that the parity principle requires functionalism Clark and Chalmers emphasize the wrong aspect of their principle. The emphasis should instead be on the *fair treatment* aspect of their principle. In other words, while all claims about fair treatment will rely on some characterization of the feature that merits fair treatment of different cases, what this characterization is can differ. On a functionalist account, it will be sameness of function that matters. But for an identity theorist what matters will be sameness of neurobiological kind. At one point Chalmers (2008: x) restates the parity principle as maintaining that, "if a process in the world works in a way that we should count as a cognitive process if it were done in the head, then we

---

[99] Of course, a parity principle based on equivalency in phenomenal properties could support the extended mind thesis only if conscious states can be externally realized, which Clark and Chalmers deny. But Clark and Chalmers deny extended consciousness due functionalist-related concerns. This issue will be the topic of Chapter Six.

should count it as a cognitive process all the same." This better captures the non-functionalist *fair treatment* interpretation that I am advocating, as does Clark's description of the parity principle as providing a "'veil of ignorance' style test meant to help avoid biochauvinistic prejudice" (2008b: 77). Clark (2008b: 114) further explains that the principle "is about equality of opportunity: avoiding a rush to judgment based on spatial location alone. [It] was meant to engage our rough sense of what we might intuitively judge to belong to the domain of cognition—rather than, say, that of digestion—but to do so without the pervasive distractions of skin and skull." This moral, of course, does not commit one to functionalism.

Some of the aforementioned critics of the extended mind also endorse this interpretation of the parity principle. Adams and Aizawa (2001, 2008) and Rupert (2004, 2009), for example, seem to endorse this interpretation (discussion in Parthemore 2011). Adams and Aizawa (2001: 46) say: "To us, [the parity principle] means that the skull does not constitute a theoretically significant boundary for cognitive science. More specifically, it means that being inside the brain cannot be the mark of the cognitive. This seems to us true and obvious." Meanwhile, Rupert (2009: 30) says, "I sympathize with the motivation behind the Parity Principle. After all, why should it matter where a process takes place? If that process instantiates cognitive or mental properties when it is over here, why should things change simply because it is now over there?"[100]

My interpretation of the parity principle as not requiring functionalism also aligns with Wheeler's. Wheeler (2010b: 9) maintains that Clark and Chalmers's original idea was that "in applying the parity principle, we should ask of some external process that plays a part in governing behaviour, 'Were this process done in the head, would we have any hesitation in recognizing it as part of a cognitive process?'" On this interpretation there is no mention of 'function' or 'functional roles'.[101] There is no reason why the

---

[100] Both quoted in Parthemore (2011: 84).

[101] At times Wheeler (2010b) seems to defend the position I attribute to him here, which is that the parity principle does not require functionalism, it only requires multiple realizability, but at certain points it is less obvious if this is his position. For example, at one point he says the parity

94

relevant feature of the processes we are considering for parity must be their function. Clark and Chalmers go the functionalist route, but Wheeler explains that the parity principle simply asks us to imagine that "certain external processes get shifted spatially, across the boundary of the skin, in an inwardly moving direction." (9) Thus, as he elaborates in a later passage, all the principle needs is a "locationally independent account of the cognitive that fixes the benchmark for parity." (20) If we wouldn't hesitate to characterize the processes (regardless of how they are type-identified) as cognitive when they are going on within the boundary of the skull, then consideration of parity demands that we give them equal cognitive credit even while they take place externally.

So Wheeler's position is that the parity principle does not need functionalism. But he maintains that it does require multiple realizers:

> The parity principle is based on the thought that it is possible for the very same type-identified cognitive state or process to be available in two different generic formats – one non-extended and one extended. Thus, in principle at least, that state or process must be realizable in either a purely organic medium or in one that involves an integrated combination of organic and non-organic structures. In other words, it must be multiply realizable. So, if we are to argue for cognitive extension by way of parity considerations, the idea that cognitive states and processes are multiply realizable must make sense. Now, as we have seen, functionalism provides one well-established platform for securing multiple realizability. (4)

Thus, an argument for the extended mind thesis by way of parity considerations could be given that does not commit one to functionalism (as quotes in the previous paragraph demonstrate). In place of functionalism, however, one would still require a locationally and compositionally uncommitted account of how mental kinds are type-identified. A

---

principle "forges a strong connection between functionalism and [the extended mind thesis]" (4; see similar comments on 9).

'locationally uncommitted' account is one that does not take a position on the location of mental kinds in its account of how these kinds are type-identified[102]; while a 'compositionally uncommitted' account is one that does not take a position on the composition of mental kinds in its account of how these kinds are type-identified.[103] On Wheeler's view an argument for the extended mind thesis by way of parity considerations needs an account of how mental kinds are type-identified that is compositionally uncommitted because it must allow for multiple realizers. The final sentence of the above quote suggests why so many think that a parity-driven argument requires functionalism: because of how closely it is tied to multiple realizability.

There is another line of reasoning that aims to show how the parity argument might drop a commitment to functionalism. Drayson (2010: 374) maintains that the parity principle "need not even involve weak functionalism."[104] Her argument rests on a distinction between 'mindedness' and 'cognition'. Some, such as Clark and Chalmers (1998), are careful to distinguish between extended cognition and extended minds, while others such as Wheeler (personal communication) draw no distinction and speak of extended mind and extended cognition interchangeably.[105] But Drayson maintains that the distinction is important. She explains:

Functionalism is the metaphysical view that mental states are individuated by

---

[102] In so far as a mind-brain identity theorists insists that all mental kinds must be located in the brain, for example, she would be committed to mental kinds have a particular location. As another example, a behaviorist who insists that all mental kinds are to be identified with one's manifest bodily behavior might also be considered locationally committed.

[103] Identity theorists, for example, are compositionally committed: their account of how mental kinds are type-identified commits them to a particular view on what mental kinds are composed of.

[104] Further support for this interpretation might be found in Gertler (2007: 9), who says Clark and Chalmers's central claim makes no commitment to functionalism. According to Gertler, Clark and Chalmers's central claim (1998: 14) is "when it comes to belief, there is nothing sacred about skull and skin."

[105] Nothing in the argument that follows hinges on this distinction, although I have and will continue to speak mostly of extended minds, or mental states/processes.

their functional relations with mental inputs, outputs, and other mental states; computationalism is the hypothesis that the functional relations between cognitive inputs, outputs, and internal states are computational. (376)

The functionalist view about mental states, Drayson argues, is often conflated with the computationalist view about cognition.[106] Drayson argues that Clark and Chalmers's parity principle is in fact committed to a computational account of cognition, not a functionalist account of the mental. Thus, one could reject functionalism (and the extended *mind* thesis) and still argue for extended *cognition* using the parity principle. Drayson argues that there is a general tendency to think that extended cognition requires functionalism because the 'parity principle requires multiple realizability and "functionalism is an obvious way to accommodate multiple realizability, so it is natural to interpret the 'parity principle' as a functionalist assumption." (376) But, computationalism can also accommodate multiple realizers of the same cognitive program, and thus, the parity principle can also be "understood as a claim about computational accounts of cognition." (376) On Drayson's view, however, the parity argument for the extended *mind* does require a commitment to a functionalism as a metaphysical theory about the mind. Thus, distinguishing between extended cognition and an extended mind allows Drayson to maintain that only the latter requires functionalism (understood as a metaphysical view about how mental kinds are individuated).

To sum up, on the Wheeler-Drayson interpretation the parity principle requires multiple realizability and not functionalism, and so, a parity-driven argument can be given without a commitment to functionalism (although for Drayson this could only support extended *cognition*). One consequence of this interpretation, which Drayson points out, is that even if Sprevak is right that functionalism entails that minds are

---

[106] Computational functionalism is a metaphysical view about the nature of mental kinds and the computational theory of mind is, as its name suggests, a theory about the nature of the mind, but these views should be distinguished from the computational program in psychology. For longer discussions of this issue see Fodor (2000) and Piccinini (2004), who Drayson agrees with.

radically extended to the point of absurdity, the extended mind theorist (or, on Drayson's account, the extended *cognition* theorist) need not fret as their argument can be independent of functionalism.

4.4 Parity without multiple realizability

We've seen that most commentators agree with Clark and Chalmers that a parity-driven argument for the extended mind thesis rests on functionalism, while others argue further that functionalism entails the extended mind thesis. Meanwhile a third position, held by Wheeler and Drayson, is that a parity-driven argument only requires multiple realizability. In what follows I begin to outline a fourth position: that a parity-driven argument does not even require multiple realizability.

It should be clear by now that there is a general tendency to think that in order for the mind to exist (partially) outside the skull one would have to make use of a different type of realizer, and therefore that the extended mind theorist must be committed to some variety of multiple realization of mental state types. But I maintain that this rests on a misunderstanding of the central insight of the parity principle and the parity-driven version of the extended mind thesis. All that is required for the parity principle is locational neutrality. Thus, just as Wheeler and Drayson asked whether functionalism was necessary for parity, and argued that it was not, so too can we ask whether multiple realizability is necessary. We've already seen that a parity-driven argument could appeal to a theory other than functionalism about the nature of mental states. A phenomenalist, for example, or a connectionist, could be motivated by considerations of parity to endorse the extended mind theory. But I will argue that even a psycho-neural identity theorist, who rejects multiple realizability and supports reductionism, could be motivated by considerations of parity to endorse the possibility of extended mental states. This is

because, on the fair-treatment interpretation, the parity principle applies just as well to a neurologically specified account of the mental.[107]

A parity-driven argument requires a commitment to some theory of mind (that is, some theory about the nature of mental kinds). In other words, one can only assess the equivalency of inner and outer processes or states given various assumptions about what minds are (Shapiro 2008). And, in particular, a parity-driven argument requires a locationally uncommitted account of what a mental state is. But, whatever skull-bound mechanisms one identifies as relevant for their theory of mind, whether these mechanisms are neurologically specified or specified in terms of a more abstract functional role, the demand for fair treatment made by the parity principle applies equally. Thus, as long as one's theory of mind is locationally uncommitted, the parity principle asks that one grant equal status to external mechanisms, however specified, that one would grant to those very same mechanisms that are skull-bound. This can be characterized as a positive commitment to a thesis about the location of mental types. I will explain and defend this thesis in the next chapter.

It is worth noting that, like functionalism, the multiple realizability thesis about mental types is not uncontroversial. Some are sceptical that mental types really are multiply realizable.[108] This is especially true with respect to the evidence from neuroplasticity that has long been cited in support of the multiple realization of mental types, e.g. by Block and Fodor (1972; see Appendix B). By showing that the parity principle does not rely on the truth of the multiple realizability thesis, one can avoid these various objections. If, on the other hand, one were to agree with the dominant

---

[107] Parthemore (2011: 83) says the extended mind thesis rejects any version of mind/brain identity (e.g., Churchland 1989) and instead maintains that "[m]ind is neither the same as nor reducible to brain; and in particular, they need not share the same boundary with respect to the world." I agree for the most part, but my account of the parity principle would allow for some cases of mental state occurring outside the skull, but still being reducible to neurological activity, and thus is compatible with the psycho-neural identity theory. Thus, my view might be in tension with Parthemore's (2011) interpretation.

[108] For example, Shagrir (1998); Bechtel and Mundale (1999); Shapiro (2004); Polger (2009); and Polger and Shapiro (2016).

interpretation of the parity principle—that it requires multiple realizers of mental types—then a supporter of the parity argument would need to address the objections that currently confront multiple realizability.

4.5 Conclusion

In this chapter I have argued that neither the extended mind thesis itself nor a parity-driven argument for the extended mind thesis needs to rely on functionalism. Clark and Chalmers emphasize the connection between functionalism and the parity argument, but in fact they could do without it. My objective is to make explicit what the core, or most central, theoretical commitments of a parity-supported extended mind thesis are and in this chapter I have argued that functionalism is not among them. In the next chapter I will argue for what I take to be the real content of the extended mind thesis, which I take to be a claim about the location of mental states. This will then allow me to show how a parity argument could support the extended mind without relying on a multiple realization of mental kinds.

CHAPTER FIVE

THE WHAT AND WHERE OF MENTAL STATES: ON WHAT IS DISTINCTIVE ABOUT THE EXTENDED MIND THESIS

5.1 Introduction

In this chapter I present a parity-driven argument for the extended mind thesis that does not rely on either multiple realizability or functionalism. The argument instead requires what I call the 'multiple localizability thesis', which says that particular kinds of mental states need not be uniquely located in any particular place, e.g. in the brain or in one of its regions. I begin by introducing and supporting this claim with evidence to do with neuroplasticity. In advancing this new argument I hope to clarify what is distinctive about the parity-driven extended mind thesis: namely that it offers insight into the location of mental types. My interpretation aligns with Clark's own view about his thesis. Nevertheless, some commentators, including both critics and supporters, have disagreed with this interpretation. I will review this debate and show that distinguishing between multiple realization and multiple localization helps to illuminate that the extended mind really is about the location of mental states, while functionalism and multiple realizability are about their nature and composition, respectively. And, I argue that because the matters of location and composition can come apart, an argument for the extended mind need not take any position on the composition of mental states. Thus, the view can remain uncommitted as to whether mental states are multiply realized or whether they can be characterized functionally.

5.2 Strict localization

Within neuroscience, there is a long-standing tradition of cerebral localizationism, the idea that each kind of mental function depends, in some way, on a particular region of the brain.[109] One of the earliest discoveries that helped establish this tradition was when Paul Broca studied the post-mortem brains of stroke patients with loss of speech articulation and found similar damage to a particular part of the left frontal lobe in all of them. Broca concluded that a lesion in the left anterior third frontal convolution resulted in a loss of speech and, thus, that this area was associated with speech production (Walker 1998). This region in the frontal lobe is now known as Broca's area.

Within the tradition of cerebral localizationism we can distinguish between two different claims. The weaker claim, which I call 'cerebral anatomical specialization', is the idea that certain regions of the brain are *typically* responsible for (or better suited to) bringing about certain mental activities. On this view, neural substrates located in Broca's area are typically involved in speech production. It is, perhaps, the area's proximity to other cortices, such as the auditory cortex and the visual cortex, that are associated with language processing functions that make it particularly well suited for the job; or perhaps for some other reason, e.g. the kind of neurons found in Broca's area.

The stronger claim within the tradition of cerebral localizationism is 'strict localizationism'. This view maintains that there is a one-to-one mapping between mental functions and brain regions. This is a type-to-type mapping claim: each type of mental function has a distinct type of location in the brain. Location types have typically been distinguished anatomically and referred to as 'brain regions' (more on this in the next section.) A strict localization, for example, requires that different regions of the brain (however finely grained[110]) are responsible for a different type of mental function

---

[109] Nowadays this tradition is familiar and even commonplace but it in fact had a tumultuous history and was long resisted. In Appendix A I discuss some of the key figures in establishing cerebral functional localizationism and respond to some of the evidence for this movement.

[110] There need to be some constraints on what differentiates one region from another, and I will discuss the leading views in the next section, but different versions of strict localization could

(however finely grained[111]), and each type of mental function has neural substrates in a particular region in the brain. The strict localizationist would maintain, for example, that substrates in Broca's area (a region of the brain) subserve speech production and no other types of mental function, and that speech production is uniquely localized to substrates in this area. This one-to-one mapping is true both across individuals. Strict localizationists do not maintain merely that each token mental state is identical to a token location: this would be the weak claim that every mental state is located *somewhere*.[112] When Paul Broca (1861) studied the post-mortem brains of stroke patients and found similar damage to a particular part of the left frontal lobe in all of them, he concluded that this area was associated with speech production not just in the individual patients he studied, but across all individuals. In other words, he took this to be a type-to-type mapping that held true across individuals. This is what the strict localization maintains: that the Broca's area of my brain subserves my speech production and that the Broca's area in my father's brain subserves his speech production. Furthermore, the strict localization maintains that within a single individual over time the same area is responsible for the same mental types: i.e. neural substrates in the Broca's area of my brain will always be necessary and sufficient for my speech production.

Strict localizationism is a stronger claim than cerebral anatomical specialization because the former entails the latter, whereas a belief in cerebral anatomical specialization does not entail that mental kinds are uniquely localized. In other words, cerebral anatomical specialization does not require that there is a one-to-one mapping

---

place different constraints. Some might maintain that regions are defined by local circuitry, while for others it might be the kind of neurons that constrain a region. What unities strict localizationists, despite differences in how they set these constraints on regions, is that they maintain that every mental kind has one particular region, and every region subserves one particular mental kind (there will also need to be clear constraints on what counts as a 'mental kind').

[111] There are also different views on what should count as a mental type and how finely grained these types should be. Ultimately, I will not take a position on this, although I will speak of 'pain' as an example of a type.

[112] A token mapping claim would be widely uncontroversial among modern philosophers. Although Descartes, for example, would have denied even this weaker claim, since he denied that mental functions were extended he would have denied hat they have any location.

between kinds of mental functions and brain regions. It merely maintains that certain regions are better suited to subserve certain cognitive functions. Many accept the weaker claim that the brain has regional specialization (Mogensen 2011), but there is a growing movement amongst those working in cognitive ontology against the possibility of mapping types of mental functions onto particular brain regions (Price and Friston 2005; Anderson 2010, 2014).

5.3 Brain regions

How exactly does the strict localizationist distinguish between different kinds of cerebral regions or locations? In contemporary literature there is not much said about how a brain region is differentiated. Early localizationists, such as Franz Joseph Gall (1822-1825), had limited ways of studying the brain. Gall examined the shapes of skulls and argued that mental faculties were localized to different areas on the surface of the cerebral cortex, which he delineated by based on properties such as concaves and convexes. The phrenological map that emerged from Gall's work drew strict, one-to-one, relations between types of mental functions and these areas on the skull. Phrenology can thus be seen as an early version of strict localizationism.

As technology has advanced, however, cortical maps have become more refined. In 1909 Korbinian Brodmann used the more advanced method of cell staining to develop a map that divided the brain into 52 different areas based on their cytoarchitectonic characteristics, i.e. cellular structure and composition. These 52 areas, which are now known as 'Brodmann areas', are still referred to today in discussions on cerebral localization. Broca's area, for example, is typically defined as Brodmann areas 44 (*par opercularis*) and 45 (*pars triangularis*) of the dominant hemisphere (Grodzinsky and Amunts 2006). Just like the earlier phrenological map that Gall had produced, Brodmann's map allows a way to differentiate between different *kinds* of cerebral locations. Each Brodmann area is a different type of cerebral location. Thus on this

classification system there are 52 types of cerebral locations. There is a token Brodmann area 44 within my skull and another token area 44 in yours—these are two tokens of the same kind of cerebral location, as defined by Brodmann.

Because different brains can vary substantially in both shape and size, by the mid-twentieth century, neurosurgeons needed a way to map the location of brain structures independent of these individual differences. To this end, in 1967, neurosurgeons Jean Talairach and Gabor Szikla (1967, 1980) developed a 3-dimensional coordinate system now known as the Talairach coordinates or 'atlas'. Their atlas was based on a single post-mortem dissection of a human brain, in which they used Brodmann areas as approximate labels for the different regions. With the development of neuroimaging techniques, however, neuroscientists needed a common space that was more representative of a wider population and that would allow them to combine data from across individuals.[113] For these reasons, researchers at the Montreal Neurological Institute (MNI) developed a new, more refined, coordinate system based on a large number of magnetic resonance images (MRIs). The MNI atlas is now the more commonly used system in the field (Poldrack, et al. 2011). The coordinates of the MNI atlas offer a way of type-classifying neurological regions independently of their functions.[114] In other words, it allows a more precise way to differentiate between different *kinds* of cerebral locations across individuals.

Hereafter, I will speak of brain regions, or neural regions, as types of cerebral locations.[115] While it is unclear how widespread a belief in strict localizationism is today[116], there have been growing discussion about the function of different brain regions

---

[113] Although this is also possible in Talairach space: see Lancaster and Fox (2009).

[114] Lancaster and Fox (2009) discuss some of difficulties with identifying anatomical or functional areas within the brain, an issue that Brodmann (1909/1994) himself wrote about.

[115] Ultimately, my position is that neuroanatomists should decide what counts as a neural region and I would defer to their expertise on the issue. In order to preserve equivalent talk about location types outside of the skull we could extend the 3 dimensional coordinates beyond the skull (i.e. extend the X, Y, and Z axes).

[116] Some supporters of localization efforts include Posner et al. (1988), who argue that elementary brain functions are strictly localized. Also Kandel, Schwartz, and Jessell (2000: 8-9) say "Gall's original idea that different regions are specialized for different functions is now accepted as one of the cornerstones of modern brain science." Though they acknowledge that some mental kinds

by cognitive neuroscientists who defend the method of network analysis (e.g. Price and Friston 2002, 2005; Friston 2002). In these discussions, the brain regions that get mentioned are often at a structural level that is appropriate for modern neuroimaging, using Talairach coordinates or the MNI atlas. Defending network analysis (against region analysis views such as strict localizationism), Price and Friston (2002: 416) for example discuss "cortical and subcortical regions of the brain with a spatial space of millimeters to centimeters." These narrowly circumscribed regions are then assembled into neuronal systems that, on the strict localizationist view, are both necessary and sufficient for particular mental functions. Price and Friston reject strict localization as overly simplistic and instead defend the method network analysis, which allows for a more distributed architecture.[117]

5.4 Multiple localization

Various kinds of empirical evidence now suggest that strict localizationism is false and that certain mental functions are instead 'multiply localizable' rather than strictly localizable.[118] That is, a given mental type need not always be realized by physical

---

are served by distinct areas, they also maintain that others have been precisely located. For some arguments against strict localization, see Vygotsky (1965); Uttal (2001); Marshall and Fink (2003); Price and Friston (2002, 2005); Friston (2002); and Anderson (2011, 2014). More recently, Uttal (2011: 256) says, "it is becoming more and more widely accepted that many parts of the brain are involved in even the simplest cognitive activity." I will discuss this topic in a bit more detail in Appendix B. There I will also consider Mundale's (2002) view: a 'heuristic' version of localization, which is compatible with multiple localization.

[117] I have a lengthier discussion of their view in Appendices A and B, where I make clear how my rejection of strict localization in the next section (5.3) is both compatible with and useful for network-based analysis (rather than region-based anaylsis, which underlies the tradition of cerebral localization).

[118] There are, of course, other options too. 'Holistic' or 'global' views that maintain that the whole brain contributes to mental functions, or that a collection of regions scattered around the brain might contribute to bringing about a mental function: distributed neural networks and dynamic systems theory may fall somewhere within these family of views. I believe at least some

structures located in the same kind of place. Instead, physical structures located in different regions of the brain can realize the same mental type. This can be true both across different individuals and in the same individual at different times. Some evidence in support of multiple locations across individuals is the fact that the lateralization of linguistic functions can differ between right and left-handed subjects.[119] So the areas responsible for language processing can be located in different hemispheres in right- and left-handers. A strict localizationist might insist that the right classification of linguistic functions should take into account whether an individual is left- or right-handed. But in fact right hemisphere language lateralization is observed in only about 20-40 percent of left-handers. The majority of left-handers also have a left-hemispheric brain specialization for language processing.[120] Thus, even amongst left-handers lateralization can differ. Furthermore, brain scans of ambidextrous individuals indicate higher hemispheric symmetry.[121] Presumably, there is less hemispheric differentiation because both sides can assume responsibility for language processing. Thus, ambidextrous individuals provide some evidence of (simultaneous) multiple locations of the same type of mental processing within the same individual.

More compelling support for the multiple localization of mental kinds within the same individual comes from evidence of brain plasticity that indicates the 'movement' of mental kinds from one time to another. For example, while linguistic processing is normally handled in the left hemisphere of right-handers, it has long been known that this can change: suffering a traumatic injury early in life can cause linguistic functions to relocate and be established in a different brain region.[122] This kind of neuroplasticity, known as 'homologous area adaptation' (Grafman and Litvan 1999, Grafman 2000a,b), shows that mental functions (or 'mental types' if one wants to avoid talk of 'functions')

---

varieties of holism are compatible with multiple localization, but this relationship merits a lengthier discussion then I can give here.

[119] Paul Broca (1865) was the first to suggest that a person's handedness was a specialization of the contralateral hemisphere.

[120] See Damasio and Damasio (1992); Springer and Deutsch (1993); and Pinker (1994).

[121] Szaflarski et al. (2002).

[122] See Block and Fodor (1972); Chugani et al. (1996); Rosen et al., (2000); and Stemmer (2006).

can shift locations and are, in this sense, not strictly localizable, but rather are 'multiply localizable'. [123] One can therefore formulate an argument from homologous area adaptation for the possibility of multiple locations of the same mental types:

(P1) A subject, S, has a mental state of a particular type, M, which is typically has a neural substrate located in a particular brain region, r_1.

(P2) Homologous area adaptation is possible, i.e. if S suffers an injury that damages r_1, M will be lost, but eventually M may be recovered by a neural substrate located in a homologous region in the opposite hemisphere: call it r_2.

(C1) Thus, M is a mental state of a particular type that could have neural substrates located in either region r_1 or r_2.

(C2) Therefore, at least some mental types are multiply localizable.

Thus, evidence of homologous area adaptation shows that strict localizationism is false, as is implied by C2. Notice that while the multiple localization of mental kinds is incompatible with strict localization, it is still compatible with the weaker claim about cerebral anatomical specialization that also emerged from the cerebral localization tradition, which does not require a one-to-one mapping of types of mental function to neural substrates in particular brain regions. It might still be true, for example, that the left anterior third convolution of the frontal lobe is well-suited to serve as the substrate of speech production because of the area's proximity to the auditory cortex and the visual cortex. Cerebral anatomical specialization is compatible with multiple localization. The multiple localizability thesis only rejects the claim that there is a strict one-to-one mapping of every mental kind to some particular type of region of the brain.

More evidence of multiple locations of mental kinds within an individual comes from another kind of brain plasticity known as 'cross-modal reassignment' (Grafman and

---

[123] I use the example of a linguistic function here because I think most would count this as a 'mental' or 'cognitive' function. Grafman (2000b) also discusses another case of homologous area adaptation involving a shift in 'visuospatial skill', studied by Levin et al. (1996).

Litvan 1999; Grafman 2000a,b). Cross-modal reassignment is characterized by sensory substitution where one cortical area previously committed to processing information in one sensory domain (e.g. audition) adapts to input from a different sensory domain (e.g. vision, Grafman and Litvan, 1999). Sharma et al. (2000) and von Melchner et al. (2000) rewired the visual systems of ferrets to project into the area that would be the location of the auditory cortex in non-manipulated ferrets. The rewired ferrets are able to process some visual stimuli by means of structures located in what is typically the auditory cortex. So their visual functions are able to shift locations and in this sense are multiply localizable. Here is an argument from cross-modal reassignment for the possibility of the multiple locations of mental types:

(P1) A subject, S, has a mental process of a particular type, M, which is typically assumed by neural substrates located in a particular brain region, r_1.

(P2) Cross-modal reassignment is possible, i.e. if S is rewired, then M can be assumed by substrates located in a different region, call it r_2.

(C1) Thus, M is a mental process of a particular type that could have neural substrates located in either region r_1 or r_2.

(C2) Therefore, at least some types of mental processes are multiply localizable.

Thus, cross-modal reassignment is a second type of brain plasticity that shows that strict localizationism is false, as is implied by C2.

Philosophers sometimes discuss cases of homologous area adaptation and rewired ferrets as offering evidence of multiple realizers of mental kinds (e.g. Block and Fodor 1972). Shapiro (2004) argues that cross-modal reassignment is not a case of multiple realization because the rewiring in the auditory cortex ends up not being a *distinct kind* from the wiring that was lost in the visual cortex. It is *the same kind* of wiring—the same kind of neural structures—only in a new region. He argues that this is evidenced by the fact that the auditory cortex of the rewired ferrets "develops the columns of orientation-

sensitive cells" that are typical of the visual cortex (64).[124] So Shapiro maintains that rewired ferrets do not provide conclusive evidence of multiple realizers of visual function. Whether or not this is right, it is clear that the visual functions in question are being instantiated in a spatially distinct region of the brain and, thus, that they are multiply localizable.[125] This example illustrates how multiple localization could be true even if multiple realization is not. Here the crucial distinction between the two is that in the first case the same mental type is 'moving' from one set of neurons to another set *of the same kind* of neurons (or neural activity) as opposed to a mental type moving from one set of neurons to another set *of a different kind* of neurons (or *neural activity*).

The same might be said of homologous area adaptation: linguistic functions may relocalize after a traumatic brain injury without being multiply realized[126] Block and Fodor, who were amongst the earliest to cite homologous area adaptation as evidence supporting multiple realizability, never consider this possibility. (I will respond to Block and Fodor's evidence in more detail in Appendix B of this thesis.) Ultimately the issue of whether neuroplasticity provides evidence of multiple realizability is an empirical matter. I maintain that more work must be done to settle whether functional recovery after brain injury relies on a novel kind of neural structure in a novel region, or on the same kind of neural structure in a novel region.[127]

This debate raises the question of what exactly the same kind of neural structure, or a difference in kind, amounts to—whether it is the kind of neurons present, the neural microstructure or class, the anatomy of the circuitry, a specific activity, or so on. Whether

---

[124] Further discussion in Polger (2009) and Polger and Shapiro (2016).

[125] Overgaard and Mogensen (2011) have a related discussion of these issues and seem to agree with Shapiro's analysis. They argue that this kind of rewiring shows that relocalization of mental function (either as a result of modified sensory input or from regional brain injury) is possible but that relocalization is not conclusive evidence of multiple realization precisely because it might be accompanied by a 'rewiring' of the local circuitry within the novel substrate that results in a neural circuitry resembling (or of the same kind as) the original, only in a new place (3). See Polger and Shapiro (2016) for more recent discussion of this issue.

[126] Shagrir (1998: 448), for example, makes this point against Fodor (1974); see also Mogensen (2011) and Polger (2009) for further discussion.

[127] This is also the position taken by Mogensen (2011) and Overgaard and Mogensen (2011).

or not these various forms of neuroplasticity are evidence of multiple realizability will ultimately hinge on what counts as 'sameness in kind' or a 'difference in kind' at the neurobiological level.[128] I will not take a position here on whether or not homologous area adaptation or cross-modal reassignment do involve multiple realizers. Nor will I settle the issue of what counts as a neurobiological kind. Instead I will stick to using what might fairly be described as toy examples: I will talk about 'C-fibres firing' versus 'O-fibres firing', or activity in a silicon-chip, as different lower-level kinds without taking a position on exactly what makes them different in kind. This was Putnam's strategy in his initial defense of multiple realization and I think it should suffice for my current defense of multiple localization as well.

5.5 How the matters of localization and realization (or composition) come apart

Notice that even a type-type identity theorist, who opposes multiple realizability, can accept the multiple localizability thesis: indeed, as long as one is willing to give up strict localization, one can accept multiple localization. If all pain states are numerically identical to C-fibres firing, for example, then pain must be located wherever C-fibres are firing. Notice that there is nothing that prevents C-fibres from being the sort of thing that could fire in the right way (under the right conditions) in numerous places – i.e. being multiply localized. Suppose we discover that C-fibres are only located in the left hemisphere of left-handed people and only located in the right hemisphere of right-handed people. For an identity theorist these differences in location, though perhaps interesting, would be completely irrelevant to the claim that pain is identical to C-fibres firing. Thus, pain could be located in different regions of the brain without being multiply realized.

---

[128] There is not a consensus on how to understand 'multiple realizability', and this issue merits more discussion that I can give here. For some discussions on this issue see Shagrir (1998), Shapiro (2004), Polger (2009), Lyre (2009), and Polger and Shapiro (2016).

A further point here is that there is nothing in the psycho-neural identity theory that seems to require a commitment to intracranialism, the view that the mind is located within the skull. This is because the identity theory does not seem to have the resources necessary to require that the skull is a relevant boundary or to make any strong commitment about the location of lower-level types at all.[129] If the relevant neurobiological type, e.g. C-fibres firing, were located outside of the skull then (assuming an appropriate connection to the wider relevant neural network) it would still be identical to some particular mental type, e.g. pain.

This example shows how the matter of location is orthogonal to the issue of composition, or realization: localization concerns *where* things are placed in space, while composition concerns *what* things are made of (or realized by). We cannot infer solely from a claim about the composition of a thing (or specifically, a part within a system) where that thing (or part) is located. Likewise, we cannot infer only from the location of a thing what that thing is composed of. In other words, assuming no other background information, we cannot derive a thing's location from knowledge about its composition, and we cannot derive its composition from knowledge about its location. Thus, taking a position on one of these issues—localization or composition—does not entail a position on the other and, in this sense, the two issues can come apart. Therefore, we can say that multiple localization and multiple realization come apart in that they are not extensionally equivalent: either one can be true in a world where the other is false.

Others have touched on the distinction between multiple realizers and multiple locations. In a paper that challenges arguments for multiple realizability, Shagrir (1998: 448) argues "psychological states and processes could be realized in different areas of the brain without being multiply realized." He thus recognizes how the multiple localization

---

[129] For a mind-*brain* identity theorist, of course, the claim is that all mental states are identical to brain states and thus the view is internalist because the brain is (typically) skull-bound. But one would need to give an independent reason for embracing this internalist commitment, otherwise it cannot be assumed without begging the question against the extended mind theorist. So in this section I focus on the *psychoneural* identity theory which says that the mental states types are identical to neural states types, and that mental states are thus located wherever the neurons that they are identical to are located, whether that is within the skull or not.

of mental types is a distinct issue from their multiple realization, and that multiple locations are not evidence of multiple realizers. Aizawa (2009) also acknowledges the distinction. He argues that lesion studies, brain imaging, and brain stimulation studies are all methods for localizing cognitive functions in the brain, that is, for determining where in the brain particular mental functions are taking place. They are not methods for determining which kinds of neural structures realize particular mental functions. And localization, Aizawa argues, "is orthogonal to the issue of multiple realization." (501) Thus, even if these methods were successful in localizing a particular mental function to a particular region of the brain, this would merely show that that mental function is uniquely localized in that region of the brain, not that it is uniquely realized. Strict localizationism, if true, does not logically entail that multiple realization is false (502). The same brain region might be able to re-wire its circuitry such that different neurobiological kinds, however they turn out to be classified, could occur within the same region. This would allow for strict localization with multiple realizers. It may seem unlikely that this is possible in human brains, but there is nothing conceptually incoherent about the idea. Aizawa thus agrees that the matters of localization and realization 'come apart' in the relevant sense,

> But what does unique or multiple localization of a psychological function have to do with unique or multiple realization? Very roughly speaking, unique localization means something like always occurring in the same place, unique realization means something like always constructed in the same manner. They are entirely separate distinctions. (501)

Thus Aizawa argues that whether strict localization or multiple localization is true, the multiple realizability thesis could still be true (or false). Aizawa brings out this distinction with the example of automobile engines. In the actual world, automobile engines are multiply realized and multiply localized: automobile engines are multiply realized in that some "use fuel injection where others use carburetors, some are water cooled where

others are air cooled, some have variable valve timing where others do not", and they are multiple localized as "some cars have engines in the front and some have them in the back." (501) But we can imagine a world where multiply realized engines might all be found only in the front of the car, "giving rise to the unique localization of multiply realized engines." (501) Or we could equally imagine a world where all engines are of the same kind, that is, built the same way and of the same materials, and yet these engines are sometimes located in the front of the car and sometimes in the back. And finally, we could imagine a world in which automobile engines are both uniquely localized and uniquely realized. It is because the location of the engine within the automobile is orthogonal to the issue of its composition that we can imagine these possible worlds. Aizawa and I agree that the same can be said of mental functions.

Aizawa defends this claim in a paper that does not discuss the extended mind thesis. But from his perspective as a contingent intracranialist, it is significant that neither multiple realizability nor multiple realization, by itself, can support the extended mind thesis because this conclusion challenges a popular view in the extended cognition literature.[130] We saw in the previous chapter that Wheeler (2010b), for example, a prominent defender of this extended mind thesis, argued that the parity principle requires multiple realizability and he seems to think this is all that it needs.[131]

---

[130] In a recent commentary on an earlier version of this chapter, Aizawa (2016) suggests that a stronger conclusion is justified: "[M]ultiple realization is stronger than multiple realizability. The former entails the latter, but not vice versa. Moreover, multiple localizability is weaker than multiple localization. The latter entails the former; but not vice versa. But, if one follows Vold's earlier argumentation, one can also see that even the stronger assumption of the multiple realization does not entail multiple localization or even multiple localizability." Available online: <http://mindsonline.philosophyofbrains.com/2016/2016-2/the-what-and-where-of-mental-states-on-what-is-distinctive-about-the-extended-mind-thesis/>
[131] I will say more on Wheeler's view in Chapter Seven.

5.6 The extended mind thesis: on the location of mental types

I maintain that the extended mind thesis is primarily a view about the location of the physical realizers of mental states. It is intended to provide an answer to the question *'where is the mind?'* And the response it offers is that the location of (the physical realizers of) mental types is not fixed to any one place within the skull or to a particular region of the brain. There is not a definite or fixed boundary of the mind, either at the skull, skin, or elsewhere. This is why it is possible for objects located in an agent's environment to partially constitute an agent's mental state. For example, Otto and his notebook realize externally the very same belief (that the museum is located on 53rd street) that Inga realizes internally. Thus the example requires that the same belief can be located in two different places and so the belief cannot be strictly localized. And any case of cognitive extension that can be supported by Clark and Chalmers's parity argument will similarly require that mental states are multiply localizable.[132]

Recall the Tetris example that Clark and Chalmers (1998) discuss. We are presented with three scenarios of human-problem solving. In each case the person is asked to answer questions concerning the potential fit of a two-dimensional geometric shape into depicted "sockets." To assess fit, the person rotates the shapes (in some manner) to align them with the sockets. In the first case, the person mentally rotates the shape. In the second case, the person uses a button that physically rotates the shapes on a computer screen. And, in the third case, the person has a neural implant that rotates the shapes at the same speed as in the second case.

Clark and Chalmers argue that all three cases have the same computational structure and therefore all three equally merit recognition as instantiating the same kind of cognitive process, a kind of "human problem-solving." They are cognitive processes *of*

---

[132] One might argue that not all versions of extended cognition, e.g. those defended by Rowlands (2010), Chemero (2009), or Menary (2007), require that mental types are multiply localizable. This might be true, but theirs are not parity-driven arguments and furthermore, I believe all their arguments offer their own insight into the *location* of mental states, e.g. that the boundary of mental states is not fixed.

the same kind but they are physically instantiated in different materials. And, in the second case, the process takes place in a distinct location, namely, outside of the skull. Thus, the second case alone challenges the doctrine of intracranialism. But to do so it relies on the possibility of a distinct localization of the relevant cognitive process as well as a different realization. In other words, the second case rests on the thesis that cognitive processes of the same type are multiply localizable, not just multiply realizable. Case (3) does not challenge intracranialism because the cognitive process takes place in the skull. Nor does it require that the cognitive process in question be multiply localizable as, for all we know, the neural implant could be in the same brain region where the process in case (1) takes place.

The parity-driven version of the extended mind thesis depends on the possibility of multiple locations. This makes sense given that Clark and Chalmers say that their thesis is about *where* the mind is, whereas functionalism is typically understood as a view about *what* the mind is. Multiple realizability is about the composition of mental states, and so, because the matters of location and composition come apart, we ought to treat them as distinct. Thus, we can understand Clark and Chalmers's parity-driven extended mind thesis as resulting from a conjunction of two claims: the thesis that mental states are multiply realizable with the thesis that they are multiply localizable. Their thesis can be broken down into these two distinct commitments, both of which are needed to support cases such as Otto and his notebook.

The 'location interpretation' is also the interpretation that Clark and Chalmers endorse (see Clark 2009b and Chalmers 2008). Another prominent supporter of the extended mind thesis, Wheeler (2010a), seems to agree with the location interpretation: he says that extended cognition "is a view about the *whereabouts* of thinking and thoughts" (2). On Wheeler's model, "parity is conceived not as parity with the inner simpliciter, but rather as parity with the inner with respect to a locationally uncommitted account of the cognitive." (8) Parity demands a locationally uncommitted account of the cognitive rather than a compositionally uncommitted account because a locationally committed account would beg the question at hand—*where is the mind?* Presumably, this

is because the key issue for the extended mind theorist is just this: an issue about the *location* of the mind.[133]

5.7 The extended mind thesis: on the composition of mental types?

There is disagreement over what the content of the extended mind thesis really is. Some, e.g. Rowlands (2010), interpret the thesis as being primarily about composition, while others, e.g. Farkas (2012), have argued that the thesis is primarily a claim about the nature of mental states. I will discuss both of these interpretations in turn. Rowlands maintains that the extended mind thesis should be understood as primarily a claim about the composition of mental states, and not a claim about location. He writes (2010: 83):

> The idea that cognitive processes are extended can easily conjure up the wrong
> sorts of images. The root of the difficulty is that extension is a spatial concept and,
> so, is closely tied to that of location. And the issue of the location of cognitive
> processes can easily side track us into concerns that we should not have.[134]

On his interpretation, the important claim is that cognition can be *realized* not just in the brain, but also by various instruments, like notebooks, which just happen to be located outside the brain. Rowlands says "what is of primary importance to this thesis is the issue of composition, not location." (84)[135] But we encounter a problem if we adopt Rowlands'

---

[133] Similarly, while none saya so explicitly, there is evidence that Barker (2010); Menary (2010); Rupert (2010a,b); Theiner (2011); and Jarvis (2014) might endorse the 'location interpretation' as well.

[134] Still, even Rowlands acknowledges that the extended mind thesis is, "among other things, a thesis about the location of mental items, and it is not clear where, if anywhere, mental types are located." (p. 64)

[135] Rowlands is discussing what he calls the 'amalgamated mind' thesis, which is meant to subsume both the embodied and the extended mind theses (2010: 84). It is worth noting that Rowlands seems to agree that the matters of location and composition come apart (2010: 83): "What is important for the new science is the composition or constitution of cognitive processes

'composition interpretation', namely that the extended mind thesis so construed does not appear to have contributed any major new insight to contemporary discussions: it would just be a consequence of multiple realizability and functionalism, ideas that have been on the table for decades in philosophy of mind. The important insight that multiple realizability represents is that other types of stuff, such as silicon, can realize mental types. It is not clear, therefore, that the extended mind thesis provides any novel insight into either the composition or the nature of mental states. Shapiro and Spaulding (2009) question the novelty of the extended mind thesis for just this reason:

> If all that matters to minds are computational processes, or if minds just *are* particular sorts of computational processes, one must begin to wonder about the novelty of [Clark's view]. Functionalists, as will be known by those who remember Ned Block's example of a nation that realizes a mind, would have no difficulty accepting the possibility that minds extend beyond heads. Is Clark simply recounting in greater detail how this might be true?[136]

I think that Shapiro and Spaulding misunderstand both Clark's extended mind thesis and Block's Chinese nation example. Block (1978) argues that as a physical system the people of China can be functionally equivalent to an individual's brain and yet intuitively lack any qualitative mental states. Clark, on the other hand, appeals to the functionalist reasoning to argue that objects located outside of the traditional boundaries of a cognitive system can in fact be a constitutive part of that cognitive system. But, Shapiro and Spaulding are right that the composition interpretation of the extended mind thesis struggles to account for what is novel about the extended mind thesis. On the composition

---

and not, in the first instance, their location. Claims about the precise location of cognitive processes do not follow, in any straightforward manner, from claims about their composition. That would depend on whether the things that compose cognitive processes have precise spatial location. However, claims about where such processes are not located do follow."

[136] Available online at: <http://ndpr.nd.edu/news/supersizing-the-mind-embodiment-action-and-cognitive-extension/> See relevant discussions in Shapiro (2011), and a similar analysis of Clark's view in Weiskopf (2008).

interpretation it seems Clark and Chalmers's thesis is merely an implication of a long-standing idea in philosophy: that mental types can have multiple realizers. On this view, the extended mind thesis has offered no new insight about the mind.

A more charitable interpretation of the thesis, therefore, must interpret it as asserting something more than just an implication of multiple realizability. On my interpretation the important claim that Clark and Chalmers's view establish is that cognition can be realized not just in the brain, but also in one's use of various instruments, like notebooks and computers, which are *located outside the brain*. I think the 'location interpretation' does a better job of explaining what is new about the extended mind thesis. The insight the thesis gives us is that location doesn't matter: mental types can be located in different types of places: both inside and outside of the brain.

5.8 The extended mind thesis: on the nature of mental types?

Interestingly, Farkas (2012) agrees that Clark endorses the location interpretation but argues that he "misrepresent[s] the essence" of his thesis (439). Farkas rejects the location interpretation on the grounds that it would not be philosophically significant because it does not conflict with any current philosophical view. She thus reasons that there must be another interpretation of the extended mind thesis that can account for why some "hail it as a significant insight" (439). She does not adopt the 'composition interpretation' but rather a third option: the extended mind thesis "has to do with how we conceive the *nature* of [mental] states" (435). Farkas argues that the "literal spatial extension" of the *realizer* of the functional role is not the central insight of the extended mind thesis (this is more or less the interpretation that I support). Instead she argues that the thesis is really about metaphorically 'extending' the relevant functional *role*: "we extend what counts as a functional role or a dispositional profile that qualifies a state to be a certain kind of standing state…this is the real lesson of the Otto-Inga case." (441)

Farkas maintains that the information in Otto's notebook figures differently into his life than how the information in Inga's brain (that constitutes her belief about the location in the museum) figures into her life. The two will have different phenomenological experiences during the retrieval process, for example. Furthermore, information in Otto's notebook will not integrate with newly formed beliefs.[137] Yet, Farkas notes that "despite these differences" defenders of the extended mind thesis argue that Otto should be regarded as having the same belief as Inga. Thus, she reasons that Clark is really advocating that we should extend what counts as the relevant functional role (or dispositional profile) for standing beliefs to include atypical profiles such as the information in Otto's notebook.

Farkas and I agree on how Clark interprets his own thesis but I support his interpretation as correct. First of all, Clark's parity argument, as I understand it, appeals to role functionalism and then makes the case that Otto's and Inga's beliefs have the same dispositional profiles. Farkas's suggestion would change the structure of the parity argument: it would begin with the assertion that Otto and Inga both have the same type of belief, followed by the claim that what makes something count as a belief is the role that it plays, and from this the conclusion would be that we should extend what counts as the relevant functional profile for standing beliefs. Thus, the extended mind theorist advocates 'extending' or 'expanding' the relevant functional role profile to include atypical realizers, such as Otto and his notebook (which, as critics point out, do not display all of the idiosyncratic features of our typical (biologically constituted) beliefs). As I see it, this no longer looks like an argument *for* the extended mind thesis as it *assumes* at the outset that Otto and Inga have the same type of belief thereby begging the question in favour of the extended mind thesis.

But, moreover, Farkas is wrong in saying that the location interpretation does not conflict with any current philosophical view and that it does not offer significant insight.

---

[137] Although Farkas does not cite Weiskopf (2008), she seems to have in mind something like the 'information integration' concern that he raises, which I discuss in Chapter One, section 1.7.2.

She does not consider the intracranialist orthodoxy of traditional cognitive science, which the extended mind thesis does conflict with. It is possible that Farkas does not consider this a philosophical view because it is not about the composition or the nature of mental kinds. But, if the extended mind thesis is about the location of mental states and if their location is a separate issue from their composition, as I have argued, then it makes good sense that the extended mind should conflict with other views about their location, such as intracranialism and strict localizationism, rather than views about composition. In this respect, even on the location interpretation, the extended mind thesis does offer some significant insight. As a further point, a predominant goal in contemporary cognitive neuroscience is to give a mechanistic explanation of psychological functions. A mechanistic account involves two essential steps: localization and decomposition (Bechtel and Richardson 2010; Craver 2007; Silberstein and Chemero 2013). The first step involves the identification of some components of the system as relevant for producing the effect of the mechanism. But this step leaves aside the issue of how those components produce the effect. This is handled during decomposition in which one breaks down the components into smaller parts based on the operations that they perform (Bechtel and Richardson 2010). The focus in contemporary philosophy of mind and cognition has long been on the *composition* of psychological functions, while the extended mind thesis, and perhaps other situated views of cognition as well, have shifted focus off of the issue of the composition and on to the issue of their location. This shift may have been so subtle that it went unnoticed by some, but the extended mind thesis has nonetheless made a relevant contribution to our understanding of how the mind works.[138]

5.9 A parity argument without multiple realizability

At the end of the last chapter I explained that a parity-driven argument requires a locationally uncommitted account of what a mental state is. The commitment to

---

[138] See Appendix A for a longer discussion of mechanistic explanations.

functionalism can thus be dispensed with and another locationally uncommitted view can be substituted in. In this chapter, we've seen that for Otto and his notebook to count as Otto's having an extended mental state both multiple realizability and multiple localizability are needed. I have argued that these two theses are conceptually distinct and that the essence, or core metaphysical commitment, of the extended mind thesis is about the location of mental kinds. To make this point clear I will show how the commitment to multiple realizability can also be dispensed with, at least in some cases of extended mental states. In other words, in its most basic form the parity argument does not need multiple realizability.

Consider the 'silicon brain chip' thought experiment (Pylyshyn 1980). Imagine that neurons in your brain are deteriorating and affecting your mental capacities, so doctors begin to replace some of your natural neurons with silicon chips. These silicon-chips are programmed to perform all of the same functions of the neurons they are replacing. They work as "artificial neurons." We can imagine that in an extreme case this replacement process continues until no part of your biological brain remains. Philosophers have different views about what will happen to your mental life while this goes on (see, e.g., Chalmers 1996). Role functionalists would say that so long as the chips really do perform all of the same functions as the neurons they replace then your mental life would not be impacted. If role functionalism is right then simply changing the physical realizers should not matter, as long as they are responsible for realizing the relevant functional role. The example of the silicon brain chips requires that mental states are multiply realizable. But it does not require that they are multiply localizable, since (we can imagine) the silicon chips are replacing neurons where they are within the skull.

Now consider the case of Diva that Clark (2009b: 1) describes in a reply to Fodor:

[I]magine a case in which a person (call her Diva) suffers minor brain damage and loses the ability to perform a simple task of arithmetic division using only her neural resources. An external silicon circuit is added that restores the previous functionality. Diva can now divide just as before, only some small part of the

work is distributed across the brain and the silicon circuit: a genuinely mental process (division) is supported by a hybrid bio-technological system.

Thus, imagine that, as in the case of Diva, instead of entering your skull to replace your deteriorating neurons where they are, doctors consider it preferable to do as much as they can externally. So they decide to attach your remaining well-functioning neurons through tiny electrical nodes to an implant that threads a wire out of your ear. They then attach this wire to an external device that contains your programed silicon chips and which attaches to your person. We can imagine that after the operation, you come to with a small wire now reaching out of one of your ears. Attached to this wire is a small device that hugs your ear, much like a hearing aid. Doctors call it an 'iCog'. Of course the iCog is not currently available and may never be, but it is conceivable and both logically and metaphysically possible. Furthermore, iCogs would be objects in an agent's environment—beyond her biological brain-and-body—that function just like the objects constituting her ordinary mental states (in this case, the neurons in her brain). Thus, iCogs are possible cases of an extended mind. About the Diva case, Clark (2009b: 1) says:

> That alone, if you accept it, establishes the key principle of *Supersizing the Mind*. It is that non-biological resources, if hooked appropriately into processes running in the human brain, can form parts of larger circuits that count as genuinely cognitive in their own right.

The idea here, as Clark says, is to "start small" with the simplest case that is alone sufficient to establish the key principle of the extended mind thesis: that "the machinery that makes minds can outrun the bounds of skin and skull" (2009b: 1). We might then consider more extravagant cases of extension, like the Otto-and-notebook case.

But even the seemingly simple case of Diva, involving hybrid biotechnology such as a silicon-based iCog, requires a commitment to both multiple localization and multiple realization. One could further imagine, however, that the iCog is itself made of neurons.

Either artificially made or naturally occurring biological neurons could be connected through one's ear by a long axon, also naturally occurring. In this case the iCog is just a (protected) clump of neurons attached to one's ear, either by an axon or by a string of other neurons. In this case multiple realizability is not required, since all mental states remain constituted by the same kind of neural states. All that is required is multiple locations. But notice that the neuron-based iCog is still an example of the extended mind thesis. In fact, when considering the possibility of extended consciousness Chalmers (2008: xiv) describes a similar case: "one could imagine that some of the neural correlates of consciousness are replaced by a module on one's belt, for example."[139] Thus, the neuron-based iCog case does not require functionalism or multiple realizability. One could be a psycho-neural identity theorist and still subscribe to the possibility that the neural correlates of consciousness could be moved to a module outside one's ear or on one's belt. Yet, considerations of parity—in the 'fair-treatment' sense—would still apply.[140] Thus, we have an example of a parity-supported case of the extended mind thesis that relies neither on functionalism nor multiple realizability. We can use this thought experiment to articulate a new parity-driven argument for the extended mind thesis:

> (P1) (At least some) mental types are multiply localizable.
>
> (P2) An object that surely counts as (at least partially) constitutive of a mental state (e.g. a neural structure in the brain) could just as well be located in the agent's environment (e.g. in the iCog).
>
> (C) Therefore, objects in an agent's environment can count as partially constitutive of an agent's mental state.

[139] I will consider Clark and Chalmers's reasons for denying that consciousness extends in the next chapter, but here it is enough to note that Chalmers thinks there is "no principled reason why the physical basis of consciousness could not be extended" (xiv).

[140] I am aware that Clark clearly opposes the identification of the neural with the cognitive, calling such identity claims "simple" and "surely naïve" (2008b: 91), but my objective here is to show that such a view is still compatible with his parity principle.

The first premise was supported earlier with several different kinds of neuroscientific evidence of multiple locations of mental types within the brain. The second premise is a claim that Clark and Chalmers (1998) articulate; to deny this outright would beg the question at hand—'*where is the mind?*'—unless one can give an independent reason for thinking the skull is theoretically significant (see Wheeler 2010a). The second premise is supported by the example of the iCog. This premise also relies on the intuition that neurons in the brain surely count as (at least partially) constitutive of (some) mental states. I take this to be uncontroversial and widely accepted by contemporary philosophers of mind, cognitive scientists, and neuroscientists. None of these premises commits one to functionalism or multiple realizability. Indeed even an identity theorist could accept the argument.

5.10 Conclusion

In this chapter I have advanced a non-functionalist version of the parity argument for the extended mind thesis. My objective was to make explicit what the core, or most central, contention of a parity-inspired extended mind thesis is, and in particular to argue that it does not require functionalism or multiple realizability. My main objective was to clarify what is really distinctive about a parity-based extended mind thesis: it is a thesis about the *location* of our mental states, not a claim about their nature or their composition, which is why it can remain uncommitted on the latter issues.

CHAPTER SIX

THE PARITY ARGUMENT FOR EXTENDED CONSCIOUSNESS

6.1 Introduction

Having argued that the extended mind thesis is really about the *location* of our mental states and, thus, that it does not require functionalism or multiple realizability, I turn to the possibility of extended consciousness. Citing concerns deriving from functionalism, Clark and Chalmers maintain that it is 'far from plausible' that conscious mental states extend. According to them only non-conscious mental states can extend, while consciousness supervenes only on internal brain states. In this chapter I will argue that extended consciousness is in principle possible and can be supported by Clark and Chalmers's original functionalist parity argument as well as by the non-functionalist version of the parity argument that I advanced in the previous chapter.

Clark and Chalmers maintain that processes that occur beyond the skull are not able to instantiate conscious mental kinds because they cannot play the necessary functional role. On their view, consciousness relies on processing that is too high in speed and bandwidth to exist beyond the brain (Chalmers 2008; Clark 2009a). In response to these concerns I will argue (a) that this claim is likely already false; (b) that it does not hold up against foreseeable advances in technology; and (c) that even if it is true, Clark and Chalmers must explain why only conscious states, and not unconscious ones, require a high degree of speed and bandwidth at the neural level. In the end I maintain that Clark and Chalmers's parity argument supports extended consciousness just as well as extended non-conscious states. Thus Clark and Chalmers's current position needs re-evaluation or a new line of defense.

6.2 Why not extended consciousness?

The first glimpse at Chalmers's reasoning for denying extended consciousness comes from his foreword to Clark's book:

> An argument for extended consciousness would require twins with different states of consciousness: Olga and Twin Olga are internal duplicates, but what it is like to be Olga differs from what it is like to be Twin Olga. But no matter how hard one tries to construct an Otto-style story that works like this, the story does not seem to succeed. Perhaps part of the reason is that the physical basis of consciousness requires direct access to information on an extremely high bandwidth. Perhaps some future extended system, with high-bandwidth sensitivity to environmental information, might be able to do the job. But our low-bandwidth conscious connection to the environment seems to have the wrong form as it stands. (2008: xiv-xv)

Here Chalmers reasons that consciousness does not extend because of some crucial difference in information-processing features — speed, access, and bandwidth — between neural and extra-neural processes. He suggests that consciousness requires bandwidth (i.e. speed of information flow) and processing speeds that are in excess of what can possibly be met by any extra-neural resources. And although Chalmers adds, "there is no principled reason why the physical basis of consciousness could not be extended", he tentatively concludes that the extended mind thesis is compatible with a denial of extended consciousness (xiv).

Clark (2009a) subsequently endorses Chalmers's suggested distinction.[141] Clark explains that consciousness may require 'certain information-accessing and information-

---

[141] Clark (2009a: 983) writes: "Chalmers (2008) does not develop this suggestion, but the direction seems promising." It's worth noting that in this article Clark does not consider a parity argument for extended consciousness. Instead he considers arguments for extended consciousness based on 'dynamic processing loops' and enactivist views of mind. I don't discuss these because I think none is as strong as Clark and Chalmers's own parity argument.

integrating operations whose temporal scale makes neural processes (just as a matter of contingent fact, in us humans) the only adequate "vehicle"' (983). Clark cites Eliasmith (2008: 150), who argues that the essential difference between neural and extra-neural dynamics is the "speed of information flow (i.e. bandwidth), and the degree and kind of coupling." He continues: "Because bodies have mass, they tend to slow down the transfer of information to the world from the brain (i.e. they effectively act as a low-pass filter). However, no such impediment to information flow exists between brain areas" (Eliasmith 2008: 150; quoted in Clark, 2009a: 984). The idea is that our extra-neural body is slower at transferring information to and from the world to the brain and so acts as a 'low-pass filter' — a physical medium that allows certain low-frequency signals to pass through, while blocking higher-frequency signals (Clark 2009a: 985).

Ultimately what is essential to Clark and Chalmers's position is their commitment to the claim that there are certain properties of neural processing that are necessary for consciousness and unlikely to be found in extra-neural processes. And, these crucial differences in properties between neural and extra-neural functioning block the possibility of parity. In other words, they provide reason to reject the possibility of an external process ever functioning sufficiently similarly to internal brain processes to merit equal recognition as constitutive parts of the physical realization of consciousness. This explains why neural processes alone constitute consciousness while all other kinds of processes in the body and environment do not, though they may provide crucial causal support for neural processes.

Clark and Chalmers's position of supporting non-conscious mental state extension while denying that conscious states could extend leaves them open to an objection. Instead of challenging their parity argument for the claim that non-conscious states can extend, it is possible for one to deny that non-conscious states really are *mental* states. One could insist that phenomenal consciousness is the defining mark of the mental. Thus, all and only mental states and processes must have phenomenal consciousness. On this view, any state that does not have phenomenal consciousness should not be considered a mental state. This view of the mark of the mental has not always been popular because

many feel that there are a wide array of intentional but non-phenomenal states that should genuinely count as *mental* states. Clark and Chalmers (1998) consider this objection and describe it as the most consistent way to resist their extended mind thesis. It is a strategy that several commentators have since advocated (e.g. Gertler 2007; Stanciu 2014).[142] My argument in this chapter will attempt to block this strategy. By arguing that extended consciousness is in principle possible and can be supported by Clark and Chalmers's original functionalist parity argument (as well as by the non-functionalist version of the parity argument that I advanced in the previous chapter) this objection will no longer be available to opponents of the extended mind—at best, they could adopt the contingent intracranialism.

6.3 Interpretations of and responses to Clark and Chalmers's claim

Clark and Chalmers's claim is a descriptive one: that as a matter of contingent empirical fact, the physical realizers of conscious states must be able to transmit and receive information with direct access and at a high speed and bandwidth. And this is why neural processes alone are able to perform the job: because they alone are able to instantiate these unique information processing features which consciousness requires. I will develop three interpretations of this claim and argue that each is likely false.

(a) The first interpretation is that neurons operate with direct access to information *from other neurons* on an extremely high bandwidth but that the same speed of information flow is not present between extra-neural processes. That is, the rate of information sharing that can take place between two neurons can never be matched between any two extra-neural objects. But this is surely false. This would amount to the claim that no extra-neural system could ever process information at the speeds of our brains; that connections between two silicon chips, for example, could never reach the connections between two neurons. But extra-neural computation can already occur at

---

[142] Gertler (2007) and Stanciu (2014), for example.

speeds that exceed brain operations. For example, computers are capable of executing more operations per second than the biological brain can.

(b) Recall that Chalmers claims that "our low-bandwidth conscious connection to the environment seems to have the wrong form as it stands" (2008: xv). So perhaps what he means is that neurons operate with direct access to information from other neurons on an extremely high bandwidth, but the same is not true of the interactions between neurons and the environment. On this interpretation, a neuron can never transfer information to any kind of extra-neural object at a high enough rate for the object to play the same role as neurons do. So the concern is not whether connections between two silicon chips, for example, could reach the speed of neurons, but rather whether a silicon chip and a neuron could ever connect in the same way (or with the same speed and bandwidth) as two neurons connected to each other. This interpretation better accords with Chalmers' description of our low-bandwidth conscious connection to the environment and with Clark's description of the body as a low-pass filter.

But, at least in one sense, extra-neural to neural-based information processing already happens at sufficiently high speeds to make this claim false. For example, one form of information exchange is visual. We know that our conscious states represent visual information that comes in from beyond our brain in a rapidly changing manner. Information about the surfaces of objects is transferred when light hits the eye, which is subsequently transmitted to the brain. But the brain, which cannot transfer or receive information at the speed of light, slows this information processing down. So extra-neural processes must be constantly sending information to the brain, through the low-pass filter Clark describes, at least as quickly as neural processes can operate.

At one point Clark (2009a: 985) acknowledges the fact that extra-neural bodily 'goings-on', such as muscular processes, must be a source of input in order for us to experience things like muscular action — an occurrent bodily process that has real-time impact on our conscious experience. But he never explicitly considers the timescale at which these operations must occur in order for this to happen. He never addresses why, if the connections between neural and extra-neural processes are sufficient to cause real-

time changes in the character of our phenomenal experience, they should not also be sufficient to *constitute* our conscious experience. This suggests that differences in processing speeds between neurons and extra-neural processes may not be relevant in determining their contributions to bringing about higher-level phenomenal conscious experiences. That is, even if extra-neural connections are relatively slower than inter-neural connections, they nonetheless suffice to realize a stream of consciousness that represents information from the outside in approximately real time.

(c) Finally, it is possible that the key difference between the neural and the extra-neural for Clark and Chalmers really has to do with having "direct access to information", as Chalmers mentions (2008: xv). Part of the difficulty in responding to this is that it is not entirely clear what 'direct' is supposed to mean. It could be that neural processes do not have direct access to information from extra-neural processes, but the information sharing between neurons is direct, in the sense of unmediated. But if this is the case then the same kind of response I gave before will do: even if extra-neural connections are relatively less direct than intra-neural connections, they nonetheless suffice to realize a stream of consciousness that processes information from the outside in approximately real time. Alternatively, if the claim is that our biological body has direct (unmediated) access to environmental information, but any add-on or 'extension' would not, this too seems false. After all, glasses can be seen as extensions of a person's visual system and there is no reason to think that these have any less direct access to visual information coming in from the environment than one's biological perceptual system has.

So Clark and Chalmers's claim can't come down to (a) a difference in speed and bandwidth between neural processing and extra-neural processing, since computers can already process information at greater speeds than our brains. It can't come down to (b) a difference in speed and bandwidth connection between the neural and extra-neural processes, since we know information from the environment must flow in at sufficiently high speeds. And finally, it can't be about a difference in (c) the 'direct access' of information, since the access that a biological perceptual system has to information from the environment is no more direct than an extended perceptual system, including a pair of

glasses or corrective lenses. At least these three plausible interpretations of Clark and Chalmers's claim appear false.

6.4 Speed and bandwidth don't matter

Even if there is some interpretation that makes Clark and Chalmers's claim true, it's not obvious why lower, implementation-level features such as speed and bandwidth would matter to states at higher levels. In a well-functioning system, parts that realize that system may move at entirely different speeds. So long as each part plays its role at the *right* time, the system can function seamlessly. As an example, all the bits of a mechanical wristwatch may move around in different ways — as one winds the spring, moving the gears, which turns the wheels forward — but the parts work together, albeit at different times and with varying sorts of connection, to realize the whole functioning system. Likewise, processes external to the brain do not in principle have to be as fast as neural processes to function with them in a larger system. They just have to play their particular role at the right time. So merely citing the bandwidth differences between the neural and extra-neural is not relevant precisely because the speeds of neural and extra-neural activity are manifestly appropriate to allow for the necessary interactions required. Thus even if Clark and Chalmers's empirical premise is true they still need to explain why these lower-level features are essential for bringing about conscious mental states.

6.5 Why isn't high bandwidth also necessary for non-conscious states?

There is one further problem with Clark and Chalmers's strategy to reject extended consciousness. Suppose that there are grounds for asserting that there is an important distinction between neural and extra-neural processes. In order to use this as an objection against a parity argument for extended consciousness, Clark and Chalmers still need to explain why differences in speed and bandwidth are relevant only to bringing

about conscious states and not unconscious ones. Otherwise Clark and Chalmers risk conceding a strong objection against their own original argument for the extended mind thesis. Thus, even if they are right that there is a crucial difference between neural and extra-neural processing, they must explain why high bandwidth does not equally pose an objection to their claim that non-conscious states can extend, otherwise the objection to extended consciousness will work just as well against their argument for extended non-conscious states.

Aware of this, Clark (2009a: 984) asks: "When does such a difference [in bandwidth] make a difference?" He replies: "Not, we can reasonably assume, in the case of non-occurrent states such as dispositional believings." Clark maintains that differences in speed and bandwidth only concern the substrates of conscious states, and not unconscious states. He argues that the contents of non-occurrent mental states do not require the high-speed, 'moment-by-moment' construction that the rapidly changing phenomenal characters of conscious states do. Only conscious states depend on the very precise (down to the millisecond) temporal synchronies of the brain, as a means of binding together and transferring distinct bodies of neurologically represented information.[143] But it is not obvious why conscious and unconscious processes must rely on different kinds of processing. The content of unconscious but occurrent mental states could be changing just as rapidly as our conscious states, for all we know, only not at the level of our awareness. For example, expert behavior, e.g. the motor skills used in driving a car, happens quickly and unconsciously. The processes involved here are not phenomenally conscious but still work fast enough to control our real-time behaviors. There is no evidence that this speed and high bandwidth are not also necessary for rapid changes in our unconscious processing.

It is not clear what Clark and Chalmers would say about these cases. In their 1998 article, Clark and Chalmers argue explicitly for the possibility of non-occurrent mental state extension, as in the Otto example, and they deny that conscious states could extend.

---

[143] Clark cites Singer (2003); Thiele and Stoner (2003); and Lamme and Spekreijse (1998) in support of this claim.

133

But they also offer some examples that suggest they accept the possibility extended unconscious occurrent processes. Their Tetris example, for instance, is an example of an extended occurrent mental process. Since they deny extended consciousness, they must have in mind an extended unconscious occurrent process. They also discuss an example of a blind man who relies on his walking stick to navigate his way around the world. Again, this example involves some kind of occurrent processing and, given their explicit denial that consciousness can extend, it seems reasonable to interpret this as an example of an extended occurrent unconscious mental process. The difficulty is that there is as yet no decisive evidence that conscious and unconscious occurrent mental processes require on different speed and bandwidth processing.

I conclude that the principled distinction that Clark and Chalmers offer to reject extended consciousness does not hold up under close scrutiny. Moreover, if their principled distinction did hold up it would support rejecting the general extended mind thesis, along with the thesis of extended consciousness. Thus, even if Clark and Chalmers can offer an interpretation of their claim that picks out a real distinction between neural and extra-neural processes, they must still find a principled distinction between conscious and unconscious processes that can explain why only the former and not the latter require the unique features of neurons to be constituted. Only then will Clark and Chalmers be able to consistently maintain that unconscious states can extend, while at the same time denying extended consciousness.

6.6 The parity argument for extended consciousness

It looks as though Clark and Chalmers's original functionalist version of the parity argument should support the view that conscious mental states can extend just as well as it supports their original thesis that non-conscious states can extend. Recall that their argument looks like this:

(P1) A physical state (or content-bearing structure) *p* is constitutive of a mental state of type *M* when *p* plays the causal role characteristic of *M* in the system.

(P2) A physical state (or content-bearing structure) located beyond (or partially beyond) an agent's biological body can play the same causal role as physical states of the biological body that surely constitute an ordinary mental state of type *M*.

(C) Therefore, physical states located beyond (or partially beyond) the biological body can be constitutive of an agent's mental state.

To support the second premise Clark and Chalmers offer some examples, including Otto and his notebook. To help make sense of what kind of external object might play the same role as the neural correlates of consciousness found in one's brain, we can recall the example of the neuron-based iCog discussed earlier. I have argued that iCogs are conceivable and that they are logical and metaphysical possibilities. So even if Clark and Chalmers's claim about the uniqueness of neural processes turns out to be contingently true, which I've given reasons to doubt, it seems at least premature to rule this out as a possible advance in technology. Furthermore, iCogs would be objects in an agent's environment — beyond the skull — that function just like the objects constituting the agent's ordinary conscious mental states (i.e. the neurons in her brain). Thus, these devices are possible cases of extended consciousness, even if speed and bandwidth considerations prove relevant to bringing about phenomenally conscious states.

Finally, notice that the example of an iCog partially constituting a person's conscious mental states can be supported using either a parity argument that endorses functionalism, as above, or a non-functionalist version of the parity argument, of the kind I presented in the last chapter. Thus, if a phenomenalist denies that we can give a functionalist profile of phenomenal aspects of consciousness, this in itself is no reason for

her to deny the possibility of extended consciousness. Furthermore, the iCog would not require multiple realizers and thus a reductive view, such as the psycho-neural identity theory, is also compatible with extended consciousness. This is because the neuron-based iCog is essentially just a (protected) clump of neurons attached to one's ear, either by an axon or by a string of other neurons. Thus, all the neuron-based iCog requires is multiple locations for the same type of conscious states. It does not challenge any of the type-type reductionist's claims. So one could be a psycho-neural identity theorist and still subscribe to the possibility that the neural correlates of consciousness could be moved to a module outside one's ear or on one's belt. These would count as constitutive of mental states because considerations of parity—in the 'fair-treatment' sense—would still apply.

When discussing the possibility of extended consciousness Chalmers (2008: xiv) describes a case very similar to the iCog: "[O]ne could imagine that some of the neural correlates of consciousness are replaced by a module on one's belt, for example." What Chalmers has in mind is essentially a neuron-based iCog worn on one's belt, instead of on one's ear. So although Chalmers's position is that it is "unlikely" that an everyday process (xiv), such as Otto's interaction with his notebook, will yield extended consciousness, he seems receptive to the possibility of a device like the iCog. It is less clear if Clark would accept the neuron-based iCog as possible, at least in principle. But to deny it, one would have to come up with some principled reason why only neurons located within the skull could instantiate phenomenal consciousness and not neurons outside of the skull.

6.7 Conclusion

In this chapter I have argued for the possibility of extended consciousness. I first argued that Clark and Chalmers's position of defending the extended mind thesis, while denying that consciousness extends is not tenable. The reasons they offer to distinguish between the contributions of neural and extra-neural processes — differences of speed and bandwidth — do not do the job of distinguishing between conscious and non-

136

conscious states. Nor do they hold up against the test of foreseeable advances in technology. This means that if their original (functionalist) parity argument works as a defense of the extended mind thesis, it is not clear why it should not also work as a defense of extended consciousness. I ended by showing that the non-functionalist parity argument that I presented in the last chapter also supports the possibility of extended consciousness.

Finally, it is worth emphasizing the importance of defending extended consciousness in the ongoing debate around the extended mind thesis. One way to object to the extended mind thesis has been to deny that unconscious states are really a part of the mind at all, and to instead contend that all (and only) our mental states are phenomenally conscious (Gertler 2007; Stanciu 2014). Clark and Chalmers (1998) say this is the most consistent way to deny their thesis. In this case it would continue to be true that the information in Otto's notebook counts as his standing beliefs, but since standing beliefs are not a part of one's mind, in no sense would the mind extend. But if all versions of the parity argument support conscious mental state extension just as well as unconscious state extension then this strategy to reject the extended mind thesis is off the table. Even if one denies that unconscious states are really a part of the mind (a claim that, for many, is radical), one still needs to find a reason to deny that conscious states can extend.

CHAPTER SEVEN

CONCLUSION

> The point of philosophy is to start with something so simple as to seem not worth stating, and to end with something so paradoxical that no one will believe it.
>
>   - Bertrand Russell, *The Philosophy of Logical Atomism, 1918*

7.1 Concluding remarks

The main purpose of this thesis has been to clarify what Clark and Chalmers's extended mind thesis is really about. There has been disagreement over how the thesis should be understood and what metaphysical commitments it has. I argue that the emphasis on the relevance of functionalism in the extended mind debate has been misplaced. In the first chapter I discussed several objections to the functionalist version of the parity argument. This included debates over the role of the parity conditions, as well as over exactly which conditions should be included. I spent Chapter Two introducing the functionalist theory of mind. In the third chapter I discussed a second round of objections, this time directed at the parity argument's reliance on functionalism. The second round ended in a deadlock between functionalists who supported the extended mind thesis and those who felt that functionalism could not support the parity argument. In Chapter Four I began to develop an argument that the thesis can be stripped of its commitment to functionalism. This, in my view, shows that there has been a misunderstanding of the parity principle and of the central import of the extended mind thesis.

To show this I began to dismantle the theoretical commitments of the original parity argument. The first step was to show how the parity principle could do without functionalism. The second step was to show how it could do without multiple realizability. At this point it became clear that the extended mind thesis could survive even without these other metaphysical commitments. What its survival revealed is that the view is essentially a thesis about the *location* of the mind: that mental kinds can have multiple locations. The multiple localizability thesis is less controversial than multiple realizability; it may also be (to some) less interesting. Perhaps this is why it is not as often discussed. In Chapter Five, for example, I discussed some of the ongoing debate over whether or not neuroplasticity involves multiple realizers of mental kinds. There we saw that even opponents of multiple realization in this debate accept the possibility of multiple locations.

After removing functionalism and multiple realizability, I argued that multiple localization is what is left at the core of the parity argument for the extended mind thesis, this apparently very simple idea that mental types can be located in different kinds of places is able to take our beliefs very far—literally outside of their skull helmets and into the world. In the introduction to this dissertation I characterized the extended mind thesis as primarily a rejection of intracranialism, a central commitment of traditional cognitive science.[144] Intracranialism is also a view about the location of mental kinds: it restricts all mental states and operations to occurring within the skull. Thus, the extended mind thesis is, at its heart, a rejection of this idea.

Once we understand that many of the interesting cases (and much of the controversy) fall out of the commitment to multiple realizability or functional equivalence it seems the extended mind thesis may not be so controversial after all. But the view is still interesting, as it shifts the focus from the issues of the composition or nature of mental types and on to the issue of their location. The extended mind thesis is

---

[144] I think the extended mind thesis also pushes back against the scientific reductionism of traditional cognitive science, but of central importance is the claim that it makes against intracranialism: as Clark and Chalmers say, there is nothing special about the skin and skull.

not, therefore, simply an implication or extension of the functionalist view that has been around since the 1960s. Instead, it offers new insight into mentality.

To end, I will consider several objections to the arguments presented in the previous chapters. I will begin by considering six objections directed at the non-functionalist version of the parity argument (or 'multiple localization' argument) more generally, and to my claim that the extended mind thesis is primarily a view about the location of mental states. The final two objections are really rejoinders to the fourth objection I consider, raised by Wheeler and Gertler in comments on an earlier draft of Chapter Five. Finally, I end by considering two objections specifically directed at the possibility of extended consciousness as defended in Chapter Six.

7.2 Objections to non-functionalist (or the 'multiple localization') version of the parity argument

I will consider six objections directed at the non-functionalist version of the parity argument (or 'multiple localization' argument) and to my claim that the extended mind thesis is primarily a view about the location of mental states. I will also offer replies to each of the six objections. Here is a reproduction of the argument from Chapter Five:

> (P1) (At least some) mental types are multiply localizable.

> (P2) An object that surely counts as (at least partially) constitutive of a mental state (e.g. a neural structure in the brain) could just as well be located in the agent's environment (e.g. in the iCog).

> (C) Therefore, objects in an agent's environment can count as partially constitutive of an agent's mental state.

7.2.1 The functionalist objection

Objection: The multiple localization argument still relies on functionalism or, at the very least, on a functionalist individuation of mental kinds. The story of the iCog began with the silicon brain chip thought experiment that involves the replacement of functionally equivalent parts. Then using the intuition that this would preserve mental state types, the scenario was 'weakened' so that that the realizers are (arguably[145]) of the same type. The argument for the preservation of mental state types in the iCog case, however, still relies on accepting functionalism as a premise. So even if we accept that the final neuron-based version of the case does not involve multiple realization, the reason for thinking that there is anything mental there at all must still be that there is no functional difference in realizer. To avoid any commitment to functionalism one needs (a) to rely on a way of individuating mental kinds without making reference to their function, and (b) to rely on a way of individuating different kinds of locations without referencing their function.

Reply: The multiple localization argument and the iCog example are perfectly compatible with functionalism, but neither require functionalism. It is true that the story I tell to motivate the iCog case begins with Pylyshyn's thought-experiment intended to support functionalism, but an identity theorist could tell their own version of the same story; so too could a connectionist, or a phenomenalist, and so on. For the identity theorist, the reason for thinking that there is some mental function being partially realized by the iCog is not that there is no functional difference in the role played by the realizer, but rather because the realizers are of the same physical types, and mental states are identical to these physical types. So the identity theorist would see no difference between the C-fibres in one's brain and the C-fibres in one's iCog, as long as both sets of fibres can fire in the appropriate ways and circumstances.

To respond to the first concern, (a), that the multiple localization argument needs a way of individuating mental kinds that does not make reference to their function, I

---

[145] There may be questions about the possibility of neuron-based iCogs without multiple realizability, e.g. as to whether natural neurons could function appropriately outside the skull.

suggest that we consider how an identity theorist individuates a mental kind such as pain. They might appeal to the functional role of pain as a starting point: that is, they might use functional individuation heuristically. But once the relevant physical kind is found, an identification would be made between that higher-level type and the lower-level physical type. This is distinct from what functionalism, as a theory of mind, would advocate. Thus, there is a need to distinguish between functionalism, understood as a view about the nature of mental kinds, and merely relying on a functional individuation of mental kinds.[146] Common-sense functionalism, which Clark appeals to, identifies mental kinds with functional kinds; it does not make any identification between mental kinds and physical kinds. The type-type identity theorist does the latter not the former. Furthermore, the identity theorist need not appeal to functional roles to individuate mental kinds. Recall that functionalism hadn't entered the scene at the time that Place and Smart first advanced their theory. Thus, they might have appealed to dispositional behavioural properties or to phenomenal properties as a way of individuating mental kinds. Place (1956), for example, discusses mental concepts that come out of folk psychology, such as 'knowing' and 'wanting' and maintains that these can be analysed in terms of dispositions to behave. He then considers mental concepts that concern our conscious experiences and sensations, which he thinks must be analysed by some sort of "inner process story." He argues, famously, that these conscious experiences will turn out to be identical with physiological processes of the brain.

A phenomenalist could also endorse the multiple localization argument for the extended mind thesis. By a 'phenomenalist' I intend someone who maintains that mental states are identical with their phenomenal properties (as an alternative to a functional individuation, she might also attempt to individuate mental states by their phenomenal properties).[147] She would, of course, have to leave out any non-conscious intentional states that lack phenomenal properties. To some this would be unsatisfying and a reason

---

[146] Searle and Block both reject functionalism, for example, but they might still accept a functional individuation of mental states.

[147] Phenomenalism in this sense should be distinguished from other views that have taken on the same name in contemporary philosophical discussions.

to reject phenomenalism because it dramatically limits the scope of the mind. Not everyone, however, finds this untenable.[148] My main concern in this thesis has been to show that the parity-driven argument is compatible with a range of views about the nature of the mind, not that all of those views are independently plausible.

I now turn to the second concern, (b), that the multiple localization argument needs a way of individuating different kinds of locations without referencing their function. I maintain that while one could distinguish different kinds of locations functionally (since the argument is meant to be compatible with functionalism), this is not required (as the argument does not rely on functionalism). Once identity claims have been made between mental kinds, such as pain, and lower-level kinds such as C-fibres firing, for example, the identity theorist could then look to physical properties as a way of distinguishing between different kinds of locations. But this cannot be how the localization program begins. Thus, a defender of multiple localization requires a way of individuating different kinds of locations without referencing their function or their physical properties. In Chapter Five, I explained how those involved in brain mapping have traditionally differentiated location kinds by reference to a Cartesian coordinate system, such as the Talairach atlas or the MNI atlas, which both rely on Brodmann's cytologically defined areas of the cortex. Neither of these coordinate systems presupposes a functionalist view about the nature of the mind.

7.2.2 Objection to the iCog

Objection: iCogs might *expand* the brain, but they do not extend the mind and, therefore, the iCog thought-experiment does not support the second premise, which states:

A physical state (or content-bearing structure) *located beyond* (or partially beyond)

---

[148] There is ongoing debate about a defining mark of the mental and some contend that this mark is phenomenal consciousness. I cannot arbitrate this issue here, but it comes up Chapter Six and it is one reason that I have defended the possibility of extended consciousness.

an agent's biological body can play the same causal role as physical states of the biological body that surely constitute an ordinary mental state of type *M*.

The iCog is not located *beyond* the agent's biological body: it expands the biological body. One reason it only *expand*s but does not extend it is that there continues to be a physical continuity between the internal brain and the iCog, whereas there is no such physical continuity between Otto and his notebook. For this reason, an intracranialist might be able to accept the neural iCog example.[149]

Reply: iCogs are objects that are not located within the boundaries of the skin. They could be made in a petri dish and never get inserted into the organism, so there would not always have been a physical continuity. Thus, iCogs are objects in the environment external to the organism that become appropriately coupled with the agent's previously existing internal resources. In a sense they do expand the brain, just as a Universal Serial Bus (USB) key 'expands' the hardware, or the physical realizer, of a computer. But we do not say that USB keys are located 'inside' the computer, even when they are plugged into it. Thus, iCogs support the second premise of the non-functionalist parity argument because they are examples of objects beyond the natural boundary of the organism that, were they to be located in the head of the agent, would surely count as constitutive of the agent's mental states. As we've seen, both Clark and Chalmers discuss some similar examples themselves. They do so together in their original paper on the topic as well:

> In the distant future we may be able to plug various modules into our brain to help us out: a module for short-term memory when we need it, for example. When a module is plugged in, the processes involving it are just as cognitive as if they had been there all along. (11)

---

[149] I thank an anonymous reviewer for raising this concern.

This suggests that they would consider a plug-in module, something like an iCog, a genuine case of cognitive extension. Thus, an iCog should establish the key principle of their thesis, just as Clark (2009b) says his Diva case would.

7.2.3 The modal objection

Objection: The localization argument ends with a modal conclusion: that objects in an agent's environment could count as partially constitutive of an agent's mental state. But a critic might argue that no objects currently satisfy the parity conditions, and thus, no objects currently count as partially constitutive of an agent's mental states (or processes). Furthermore, a critic could argue that no object will *ever* meet the conditions for parity. So the multiple localization argument does not show that cognitive extension is actually happening or that it ever will happen. At best it shows that it *could* happen. And this is a less controversial claim than the claim that cognition *does* extend. We've already seen that critics of the extended mind thesis, including Adams and Aizawa (2008), accept this weaker claim, declaring themselves 'contingent intracranialists'.

Reply: The multiple localization argument shows how it is in principle possible for the extended mind thesis to be true even if functionalism and multiple realizability turn out to be false. The neural-based iCog can be understood as a heuristic device: it is intended to clarify what is distinctive about the extended mind given that the iCog does not require the truth of functionalism and multiple realizability. Thus it is meant to support the second premise of the argument, according to which an object that clearly counts as constitutive of a mental state type *could* just as well be located in the agent's environment. All that is needed to this end is the modal claim that iCogs could one day instantiate an agent's mental states. Whether or not iCogs will ever be realities is not important for this purpose. The argument shows how that the mind *could* extend without multiple realizability or functionalism. In doing so, it also shows us what is left of the

parity-driven extended mind thesis when it is stripped of its commitment to these other long-standing views: views — which it is often thought to require

But the iCog serves a further purpose as well. In a recent commentary, Aizawa (2016) raises a version of what I am calling 'the modal objection'. There he says:

> [T]o my knowledge no one has made the case that extended cognition is impossible. But, suppose that there are those who take extended cognition to be, in some sense, impossible. Why is the iCogs thought experiment any more illuminating that the infamous Inga-Otto thought experiment? And why is this any more illuminating than simply supposing that cognition would take place outside the skull, if someone were simply to have one's skullcap surgically removed? The … interest in the possibility of extended cognition, would be enhanced if a) there were some plausible reason to think that extended cognition is impossible, but b) this plausible reason was shown to be untenable by the iCogs example.

Aizawa's comments are on an earlier version of Chapter Five that did not handle the issue of extended consciousness. But it is worth pointing out that in Chapter Six I went on to use the iCog example in just the way that Aizawa suggests. There I used the iCog to show that the 'plausible' reason Clark and Chalmers cite for thinking that extended consciousness is not possible is not tenable. This, again, has been part of the main objective of this thesis: to highlight how the central commitment of the parity-driven extended mind thesis is not so radical. It is fully captured by the case of the neural-based iCog.

7.2.4 The multiple localization argument only supports a weak and uninteresting variety of the extended mind

Objection: The extended mind thesis is interesting because it claims that external objects in the most robust sense (e.g. notebooks, cell phones, or other people) can be partially constitutive of mental states. The multiple localization argument only supports a weak and uninteresting variety of the extended mind. The argument cannot support any of the interesting examples of cognitive extension that Clark and Chalmers give. Thus, even if it one grants that the neural iCog is an example of mental extension, this is only in the strictest (or *weakest*) sense of the term 'extension'. The example is not interesting, and so it cannot explain what is so interesting about the extended mind thesis.[150]

Reply: It is true that the iCog example is compositionally conservative, in that it is a case of the extended mind that preserves 'unique realization', when compared to more compositionally extravagant cases such as Otto and the notebook. The result is that there can be a spectrum of cases: the neuron-based iCog only requires multiple locations, but without multiple realizers the resulting version of the extended mind thesis is conservative when compared with the more extravagant range of cases that Clark and Chalmers's functionalist parity argument supports. So one must accept both multiple localizability and multiple realizability for Otto's notebook to count as a case of extension.

The goal of stripping the extended mind thesis of functionalism and multiple realization has been to show what is really distinctive about the thesis. Some might assume that in order for the mind to extend beyond the skull one would have to make use of a different type of realizer, and therefore that one has to have a commitment to some variety of multiple realization. I would venture to say that most of the arguments for the

---

[150] This concern was initially raised by Steven James, who I thank for very helpful comments and discussion at the 2016 Eastern Division Meeting of the American Philosophical Association.

extended mind make use of the possibility of multiple realization.[151] But iCogs show how we could preserve 'unique realization' and still move outside of the skull. Thus the example pulls apart two issues that were thought to go hand in hand. And in doing so it shows what is distinctive about the extended mind thesis: it is, in essence, a thesis that has given us original insight about the *location* of mental states. For this reason the extended mind thesis may not be so controversial after all, once we understand that many of the interesting cases (and much of the controversy) fall out of the commitment to multiple realizability or functional equivalence. But the extended mind thesis itself is still interesting, as it shifts focus off of the issues of the composition or nature of mental types and on to the issue of their location.

7.2.5 Wheeler's objection

Objection: Wheeler (2016) maintains that my response to the previous objection falls short. He agrees that one can have the extended mind thesis without multiple realizability, but he argues that the sorts of case that have been central to the appeal of the extended mind hypothesis involve *artifacts*. And these cases do have to appeal to multiple realization, if they rely on the parity principle. Thus, a limitation of the multiple localization argument is that it cannot support cases of artifact-involving extended cognition. So he argues that my non-functional parity argument cannot support the most interesting cases that make the extended mind thesis so compelling.[152]

---

[151] As mentioned, some notable exceptions might include the 'continuous reciprocal causation' argument (see, for example, Palermos 2014) as well as Menary's (2007) cognitive integration approach.

[152] Wheeler says: "[A]s an argument for the extended mind, Vold's new argument, based as it is wholly on multiple localizability, doesn't take us as far as we would like. Getting us the rest of the way depends on the very notion – multiple realizability – that Vold wishes to banish to the conceptual periphery. Indeed, it's that very notion that delivers the interesting and distinctive result about the spatial location of mental states, in the cases we care most about." This commentary appeared as part of the MindsOnline Conference on BrainsBlog. Available online:

Reply: Wheeler is wrong in thinking that multiple realizability delivers the interesting and distinctive result about the spatial location of mental states. Recall that multiple realization is possible even while multiple localization is false. Thus, there being multiple realizers of mental kinds, does not tell us anything conclusive about whether they can also have multiple locations. Wheeler seems to agree on this point—that a parity-driven argument can rely only on multiple locations, without multiple realizers—and even agrees with the 'location interpretation' according to which it is the claim about location that makes the extended mind thesis interesting and distinct.

Wheeler seems to be objecting to my view by saying, "For *that* (the kind of extended cognition we really care about), one needs multiple localization *plus* multiple realization." But in my view this is exactly right: it is only the conjunction of multiple localization with multiple realization that will yield the kind of parity-driven cases that Wheeler finds most interesting. Neither on its own will do the job. Thus, his only disagreement, as far as I can see, is one of emphasis. Wheeler thinks the only parity-argument of interest will be one that appeals to both multiple realization and multiple localization. I agree that this is needed for the cases he is interested in (e.g. Otto and the notebook). But my main aim has been to distinguish what is distinctive about the extended mind thesis and what the core metaphysical commitments of the parity argument are, and on this I maintain that any version of the parity argument needs multiple localization, while not all versions of the argument require multiple realizers.

7.2.6 Gertler's objection

Objection: Gertler (2016) objects that the location interpretation is undermined by the fact that multiple localization alone can only support 'conservative' cases, such as the iCog:

> The iCog argument supports a thesis of multiple localizability only: it shows that mentality can be partly constituted by factors outside the skull. The notion that this thesis is more "conservative" than the thesis supported by Clark and Chalmers' arguments (which I presume is the Extended Mind thesis) seems to undermine Vold's argument for the location interpretation of the Extended Mind thesis. After all, the comparative conservatism of the multiple localizability thesis implies that the location interpretation does not capture the sense that the Extended Mind thesis is, as Vold says, a "novel" and "important insight."

Gertler's concern here is similar to Wheeler's objection. I phrase it as follows: how can a relatively uninteresting example, such as the iCog, reveal to us what is novel and interesting about a thesis that has been used to support other quite interesting cases, such as Otto and his notebook?

Reply: Although I acknowledge that the iCog case may be (in some sense) less interesting than the notebook case, I do not think that this undermines the location interpretation of the extended mind thesis. The strength of the iCog example comes from its simplicity. It presents a direct challenge to the intracranialist position. The iCog example, in a sense, represents the purest form of cognitive extension—one that is not convoluted by other controversial commitments. The extended mind thesis makes no claim about the nature or the composition of mental states that has not already been stated. Instead, it appeals to two long-standing views—functionalism and multiple

realizability—coupled with a unique claim about the locations of mental kinds. It is, at its core, a rejection of intracranialism that relies on multiple localization.

7.3 Objections to a parity argument for extended consciousness

In this section I will consider and reply to two objections specifically directed at the possibility of extended consciousness as defended in Chapter Six.

7.3.1 The iCog example begs the question

Objection: One might charge that the iCog begs the question by assuming that parity will one day be possible. One could argue that the iCog will never be able to reach the high speeds and bandwidth connections of the brain. Electrical connections from neurons to a silicon iCog, for example, may turn out to be too slow. Thus, true parity between inner and outer processes will never be possible.

Reply: The iCog example no more begs the question than Otto's notebook would: both are plausible scenarios that merit the application of the parity principle—that is, fair treatment. The iCog, of course, meets the tentative functional conditions that were set out for an external resource, just as Otto's notebook is argued to: it is a constant in the agent's life, the agent accesses information from the iCog regularly, and relies on this information, endorsing it as true without hesitation. But furthermore even if silicon-processing turns out to be incapable of achieving the processing features of natural neurons iCog doesn't have to be made of silicon and the connection doesn't have to be an electrical one. What is important about the neuron-based iCog is that there will be no differences in speed, bandwidth, or access. So in this case, one can no longer appeal to a distinction between neural and extra-neural processing. But since the brain and the iCog

would both be composed of neurons, the entire realization-base of consciousness would still be purely neural, so one would be hard pressed to find a principled distinction between the neuron-based iCog and the neurons in the brain that doesn't just come down to a difference between inner and outer. And this is just what their parity principle is meant to work against.

7.3.2 Rejecting a functionalist account of consciousness

Objection: One could reject the first premise of the functionalist version of the parity argument for extended consciousness by maintaining that conscious states, unlike unconscious ones, cannot be individuated by their functional or causal roles. One might think this would be a natural strategy for Chalmers, who argues that a reductive explanation of consciousness fails precisely because it cannot be analyzed functionally (1996). In other words, a complete functional analysis of the mind would not suffice to capture consciousness. Thus, an argument that rests on a functional account of consciousness will not work.

Reply: We saw in Chapter Six that this is not Chalmers's main strategy: both he and Clark reject extended consciousness by arguing that an object located beyond an agent's biological body is never causally equivalent with neurons in the brain. This amounts to a rejection of the second premise of the parity argument, not the first one. But Chalmers also maintains that the parity argument does not commit the extended mind thesis to a functionalist account of consciousness:

> I think that functionalism about consciousness is implausible, for example, but this implausibility does not affect the arguments for the extended mind thesis. One might support the view by invoking an attenuated functionalism: say, one where certain mental states (such as dispositional beliefs) are defined by their causal

relations to conscious states, to behavior, and to other elements of the cognitive network. (2008: 7)

In other words, although they cite functionalist related concerns for rejecting extended consciousness, Clark and Chalmers also restrict the functionalism that they appeal to so that it only applies to 'certain mental states', namely non-conscious ones. This means that even if Clark and Chalmers concede my arguments against the functional differences they cite—differences of speed and bandwidth—they could maintain that the 'attenuated' version of functionalism they appeal to precludes the possibility of extended consciousness.

The problem with this strategy is that I have now argued for the possibility of a non-functionalist version of the parity argument, which also supports extended consciousness. This argument supports the modal claim that consciousness could extend into something like a neural-based iCog device without appealing to any kind of functionalism. Thus, rejecting premise one of the functionalist version of the parity argument for phenomenal consciousness will not allow one to avoid the possibility of extended consciousness.

7.4 Closing remarks

In my view, the most important objection that the multiple localization argument confronts is that it cannot support the case of Otto and his notebook, or any other cases beyond the neuron-based iCog. This is the fourth objection that I considered (in section 7.2.4), which is elaborated on in Wheeler's and Gertler's objections. In response to this concern I have accepted that multiple localization alone will not support the case of Otto and his notebook. But I maintain that multiple localization (alone) will give us cases of the extended mind, albeit very basic ones. And this fact reveals what the parity-driven version of the extended mind thesis is all about: it is foremost a rejection of

intracranialism, rather than an implication of functionalism. One future direction I will pursue in response to this concern is to consider other possible cases of extension, beyond the neuron-based iCog, that could be supported by multiple localization alone (i.e. without multiple realization). One possibility is to consider cases of socially extended cognition. If the neural substrates of a belief stored in Inga's brain could play the same role in Otto's mental life that neural substrates of the same kind of belief located in his own brain once did, then Otto's socially extended belief (stored in Inga's brain) would be in a dramatically different location (with respect to Otto, as the central agent). And yet, plausibly, the belief in question could still be uniquely realized by the same kind of neural substrates. In this way, cases of socially extended cognition might offer a way to respond to what seems to be the most pressing objection to the central argument in this thesis.

APPENDIX A

On the Whereabouts of Mental Types

A1 Introduction

The idea that different regions of the brain are responsible for different kinds of mental function has been around since antiquity.[153] But it wasn't until the 19th century that this view was popularized. The modern tradition of cerebral localizationism is often traced back to Franz Joseph Gall's theory of organology developed in the 18th century, and this tradition lives on in contemporary cognitive neuroscience. Most, perhaps all, view the use of brain imaging technologies, such as positron emission tomography (PET), magnetic resonance imaging (MRI), and magnetoencephalography (MEG), as modern attempts to localize cognitive functions.[154] Perhaps adding to this picture is in cognitive neuroscience localization is seen as an essential step, along with decomposition, in providing a mechanistic explanation of cognitive function.[155] In Chapter Five I argued against strict localizationism, the view that there is a one-to-one mapping of cognitive function (or mental types) to brain regions. Instead, I advance a thesis according to which mental state types can be realized by physical structures located in vastly different kinds of places. In this appendix I introduce some of the key figures and discoveries that established the tradition of cerebral localizationism. I then question how these discoveries have been used to support strict localizationism. Finally, because giving up the possibility of 'strictly localizing' mental kinds might appear to be in tension with mechanistic

---

[153] Herophilus of Alexandria (fl. C. 300 BCE) and Hippocrates (460?–377 BCE) reportedly both held similar views: Gross (1998); Marshall and Fink (2003); Clarke and Jacyna (1987) .

[154] For example, Uttal (2001); Marshall and Fink (2003); and Poldrack (2010).

[155] Salmon (1984); Craver (2006, 2007); Bechtel and Richardson (2010); Silberstein and Chemero (2013).

explanations, I will end by arguing that the possibility of multiple locations of mental types poses no threat to this explanatory goal.

A2 A brief history of the tradition of cerebral functional localizationism

In neuroscience, cerebral functional localizationism is the view that each type of mental function depends on a particular region of the brain. Nowadays this tradition is familiar but it in fact had a tumultuous history and was long resisted.[156] In this section I will briefly recount some of the key figures in establishing this view.[157]

In the Introduction of this thesis I discussed some of the features of Descartes's doctrine of the mind that shaped traditional cognitive science. For centuries after Descartes, his doctrine also shaped how scholars thought about the issue of localization. In the *Meditations*, Descartes (1641) conceived of the soul as a unitary, indivisible, non-physical substance. This was in harmony with the prevailing Christian religious doctrine of the time, which considered unity and indivisibility to be essential features of the nature of the soul. Thus, at that time any attempts to localize mental function to a region of the brain would have been seen as invalid because they contradicted church doctrine (Clarke and Jacyna 1987). Thus, Descartes likely reasoned that unity and indivisibility excluded the mental from the material world: matter is both extended (it takes up space) and divisible, and thus the mind had to be immaterial or else it would also be divisible and decomposable (Clarke and Jacyna 1987). Although Descartes eventually located the interaction of the mind and body to a particular part of the brain, the pineal gland, this was only in response to concerns (raised by Princess Elisabeth of Bohemia in

---

[156] Tracing the history of neuroscientific concepts, Clarke and Jacyna (1987: 212) contend that "[p]erhaps the most important, and certainly one of the most contentious, episodes in the history of the neurosciences has been the attempts made to localize functions in the various parts of the brain."

[157] For more in depth discussions see Clarke and Jacyna (1987); Walker (1998); Mundale (2002); Bennett and Hacker (2008); Gross (2009); Sheperd (2010); and Acharya et al. (2012).

correspondence) that interaction between material and immaterial substances is not possible. And, in the *Passions*, he (1649) maintained that this interaction occurred at a single point in space—likely in an attempt to avoid compromising the essential unity of the mind.

One of the earliest theories about how the brain acts and how mental functions are localized within the brain came from Albrecht von Haller in the 18th century, who proposed a theory of *action commune*, or unitary brain action. Von Haller maintained that the brain acts as a whole, with each of its parts contributing an equal functional significance, thus precluding the possibility of localizing any specific function to any particular region (Clarke and Jacyna 1987). Von Haller's doctrine prevailed until the early decades of the 19th century.[158] But by the 1820s there were several groups who opposed Hallerian brain equipotentiality in favour of something more akin to the modern view—that the various subdivisions of the brain are made of many discrete areas each of which is responsible for a particular function.[159] One of these groups was the supporters of Franz Joseph Gall's theory.[160] Gall (1822-1825) was one of the first to attempt to deny the Cartesian principle of unity by arguing that mental functions could be precisely localized in different areas on the surface of the brain. His theory of 'organology', which was later known as 'phrenology,'[161] is widely recognized as the earliest precursor to the modern view of localization (e.g. Finger 2010; Marshall and Fink 2003; Clarke and Jacyna 1987).

But localization efforts have long had their critics. Gall's supporters never had a chance to establish their view before Pierre Flourens came out against it. Flourens (1846) introduced experimental evidence in support of *actions propres*, which honoured the

---

[158] And his theory (more or less) was later revived by Pierre Flourens (1846) and Karl Spencer Lashley (1929), whose view I will discuss in Appendix B. See discussions in Clarke and Jacyna (1987); Gross (2009); Mundale (2002).

[159] Clarke and Jacyna (1987); Finger (2010).

[160] Clarke and Jacyna (1987); Gross (2009).

[161] Named by one of Gall's students, J. G. Spurzheim (Stemmer 2006). According to Clarke and Jacyna (1987: 222) Gall himself never used the title 'phrenology' for his view, though it is now widely known under this name.

Hallerian doctrine of equipotentiality. Furthermore, he attacked Gall's idea of cerebral localization charging that it violated Church doctrine. And, although Flourens' theory was eventually proven wrong, his view prevailed until experimental evidence began to confirm Gall's hypothesis (Clarke and Jacyna 1987; Gross 2009).

Paul Broca, for example, studied the post-mortem brains of stroke patients with loss of speech articulation and found similar damage to a particular part of the left frontal lobe in all of them. Broca published these findings in 1861. In his paper he concluded that a lesion in the left anterior third frontal convolution resulted in a loss of speech, "but not a paralysis of the muscles involved in phonation", and thus that this area was associated with language (Walker 1998: 91). This was different than what Gall had postulated. Gall maintained that mental faculties would be localized to different areas on the surface of the cerebral cortex, but Broca located the faculty of speech on its convexity, not on the brains surface (Clarke and Jacyna 1987). A decade later Carl Wernicke localized another language function to the left posterior, superior temporal gyrus and inferior parietal lobule (Acharya et al. 2012). Adding to the localizationist picture was evidence from Fritsch and Hitzig. They showed that the frontal cortex of dogs is responsible for motor control by stimulating it to produce discrete movements of the limbs. When excised on one side a paralysis on the opposite side would result. In 1870, they published their classical work *Uber die elektrische Erregbarkeit des Grosshirns*, in which they described the results of their experiments (Bennett and Hacker 2008). Despite these discoveries, by the mid-19th century the theory of discrete localization of cerebral function was still contested: some physicians favoured Flourens' concept of unitary action, while others believed in the localizationist picture (Bennett and Hacker 2008; Walker 1998; Clarke and Jacyna 1987).

A3 Evaluating the evidence for strict localizationism

Despite this long tradition of trying to localize mental functions I have argued in an earlier chapter that neuroscientific evidence shows that mental states types are not strictly localizable to any particular kind of place. Thus, it is worth being clear about what the error is in the interpretation of evidence in favour of strict localizationism.

Early localizationists, as we have seen had to rely on crude methodologies to determined locations: Gall examined the shapes of skulls; Fritsch and Hitzig relied on both cortical stimulation and ablation on the brains of canines; while Broca studied the post-mortem brains of stroke patients. For brevity, I will focus on just one type of methodology: lesion studies. Lesion studies would have led researchers to notice that every time a certain area of the brain, r_1, is damaged, a type of mental process ceases to be possible, M. And from this they would have likely reasoned that M depends upon r_1. And M's dependence on structures in r_1 suggested that r_1 was the unique locus of M. But early lesion studies (and stimulation studies) left open the possibility that the lesioned structures in area, r_1, were only one component responsible for bringing about M. In other words, it may be the case that every time structures in r_1 are lesioned M suddenly ceases, and so M does depend on structures in r_1. But it does not follow from this that m is located, or has its seat, in r_1. All that follows is that M depended on the structures in r_1. But while the structures in r_1 may be partially responsible for bringing about, or sustaining, M, they may not be the only, structures that M depends on. For all we know M could depend crucially on structures located in both r_1 and r_2, another discrete portion region of the brain, and the removal of either portion of the brain would cause m to cease, or malfunction (perhaps even in similar ways). Another concern with the line of reasoning is that even if M depends on the structures located in r_1, these relevant type of structures could just as well be located somewhere else. The evidence for this, which I discuss in Chapter Five, comes from cases of post-traumatic recovery, for example, when homologous areas in the opposite assume the lost functions (Grafman and Litvan 1999, Grafman 2000a,b).

159

Nowadays we have moved past deficit and lesion studies to brain imaging technologies, including PET, MRI, and MEG. And, while it is difficult to gauge just how widespread a belief in strict localizationism is today, some argue that associating functions with particular brain regions continues to be one of the main projects of modern cognitive neuroscience (Mogensen 2011; Mogensen and Malá 2009; Polger 2009).[162] But the evidence for localization that these new technologies provide is also being challenged. One recent and important objection comes from Anderson (2010, 2014) who argues that these brain-imaging techniques show that the very same region can be responsible for multiple different cognitive functions. His theory of neural reuse (also known as massive neural redeployment) maintains that even after an initial function is established, neural circuits continue to acquire new functional uses throughout life, such that the same brain region can be constitutive of different mental functions. On Anderson's view, the acquisition of new function does not need to involve any unusual circumstances, such as injury, nor does it need to involve any changes to local circuit structure. Neural reuse challenges strict localizationism from the bottom up (i.e. looking at the multiple functions of the same region), whereas multiple localization challenges strict localizationism from the top down. Thus, neither lesion studies nor modern brain imaging lend conclusive support for strict localizationism.[163]

---

[162] Mogensen (2011) further says that there is "an overall agreement" that functional localization is possible.

[163] Notice, however, that both of these challenges to strict localization are still compatible with the view that certain regions of the brain are *typically* responsible (or better suited) for bringing about certain mental activities—the view known as 'cerebral anatomical specialization'.

A4 Mechanistic explanations

Providing a mechanistic explanation of cognitive function is one goal of cognitive neuroscience.[164] A mechanistic explanation begins with a provisional understanding of the explanandum and then localizes the relevant mechanisms within the system before then decomposing the cognitive function to be explained into distinct sub-functions (Craver 2007; Bechtel and Richardson 2010; Silberstein and Chemero 2013). Thus, it is thought that one must first localize—that is, identifying which components of the system perform which of the sub-functions of the system—before taking the further step of decomposing these functions—determining how those parts produce or bring about the effect in question (Bechtel and Richardson 2010; Silberstein and Chemero 2013).

Given the prominence of localization in this methodology one might think that giving up the possibility of 'strictly' localizing a cognitive kind is in tension with the possibility of giving a mechanistic explanation of that kind. Mogensen (2011: 1), for example, argues that there is an "apparent contradiction" between the effort to localize function to specific brain regions and the post-traumatic recovery of these functions—precisely the evidence I use to support multiple localization. But multiple localization in fact pose no threat to mechanistic explanations within network analysis. Advocates of this methodology (e.g. Bechtel and Richardson 2010; Kiverstein and Miller 2015; Price and Friston 2002, 2005; Anderson 2010, 2014) generally acknowledge that mechanisms are not strictly localizable, thereby permitting the possibility that distributed networks are responsible for at least some, perhaps many, cognitive functions.[165] How then is the localization step to be understood?

Mundale (2002) defends a 'heuristic' understanding of localization: it is a strategy employed in the search of the relevant mechanisms, but when strict localization fails, it can be abandoned for a more 'complex' localization strategy. Knowing that cars can have

---

[164] See Bechtel (2009); Craver (2006, 2007). Some argue that there is no viable alternative, for example, Kaplan and Bechtel (2011) and Kaplan and Craver (2011). See response in Silberstein and Chemero (2013).

[165] I have further discussion of multiple localization and network analysis in Appendix B.

engines in the front or the back of the vehicle, for example, doesn't prevent us from giving mechanistic explanations of automobiles.[166] But it does prevent us from making any kind of one-to-one mapping between engines and their locations that could be generalized across all vehicles. The same is true with respect to mental kinds, although the localization strategy may be much more complex. Thus, while multiple localization is not compatible with the region-based analysis that underlies the strict localization tradition, it is compatible with the goal of mechanistic explanations, which continue to be employed in network analysis, for example. Cognitive neuroscientists can thus continue the project of *heuristically* localizing cognitive functions for the purpose of giving a mechanistic account without 'strictly' localizing them.

A5 Conclusion

I began this Appendix by considering some of the historical developments and scientific discoveries that played a role in establishing the tradition of cerebral localizationism. But I have questioned how these discoveries have been used to support strict localizationism. The localization of the underlying mechanisms of mental function, however, continues to be an important goal in contemporary cognitive neuroscience. Thus, I ended by making clear how giving up strict localization in favor of multiple locations of mental kinds poses no threat to the explanatory goal of giving mechanistic explanations of mental functions.

---

[166] This example is adapted from Aizawa (2009).

APPENDIX B

NEUROPLASTICITY AND THE MULTIPLE REALIZABILITY THESIS

B1 Introduction

Some philosophers e.g. Block and Fodor (1972), cite neuroplasticity as evidence of the multiple realization of mental types, while others e.g. Polger and Shapiro (2016) argue that neuroplasticity fails to support multiple realization. In particular some of the evidence of neuroplasticity, e.g. homologous area adaptation, that I draw on in Chapter Five to support the multiple localization of mental types is cited as evidence of multiple realization. Block and Fodor were perhaps the earliest to cite this kind of neuroplasticity. Therefore, in this appendix I will take a critical look at their evidence. I argue that it remains an open question whether this kind of neuroplasticity requires multiple realizers.[167] My reason for addressing this issue is not to deny that mental states are multiply realizable (I suspect they are), but rather to emphasize that mental kinds having multiple locations does not depend on them also having multiple realizers. Finally, I will end by briefly discussing an ambiguity in current discussions about 'degeneracy' in cognitive neuroscience that is highlighted by the distinction between multiple realizability and multiple localizability.

---

[167] As there are many different kinds of neuroplasticity and different researchers mean widely different things by the term, I will not survey all the various forms to evaluate whether any provide conclusive evidence for multiple realization.

B2 Block and Fodor's evidence

Following Putnam (1960, 1967) and Fodor (1974) the multiple realizablity of mental types would require that the same mental type could be realized or instantiated by heterogeneous lower-level (in this case, physical) types (Lyre 2009). Putnam is credited with first expressing the intuition that mental types are multiply realizable and the view is now widely accepted in cognitive science. The multiple realizability of mental types continues to be perhaps the strongest objection to reductionist theories of mind, such as the mind-brain identity theory. Block and Fodor (1972) were amongst the earliest to cite neuroplasticity as evidence of the multiple realization of mental types. The thought is that neuroplasticity shows that different kinds of neural substrates can subserve the same mental function. Block and Fodor discuss neuroplasticity in just one paragraph, which I quote in full:

> The Lashleyan doctrine of neurological equipotentiality holds that any of a wide variety of psychological functions can be served by any of a wide variety of neural structures. While the generality of this doctrine may be disputed, it does seem clear that the central nervous system is highly labile and that a given type of psychological process is in fact often associated with a variety of distinct neurological structures. (For example, it is a well-known fact that early trauma can lead to the establishment of linguistic functions in the right hemisphere of right-handed subjects). (160-1) [168]

---

[168] Along with equipotentiality, Block and Fodor also provide two other "empirical considerations" in support of the multiple realizability thesis. The first is the possibility that there could be "organisms whose psychology is homologous to our own but whose physiology is quite different"; the second is "the conceptual possibility that psychological predicates could apply to artifacts" (161). But I agree with Polger (2009) who argues that the identity theorist would likely not find either of these reasons compelling and would instead require evidence of multiple realizers to be swayed.

In the first sentence, Block and Fodor cite Karl Lashley's doctrine of equipotentiality. But they misrepresent Lashley's (1929) doctrine. The doctrine does not maintain, as Block and Fodor claim, that a wide variety of psychological functions can be served by a wide variety of neural structures. Lashley's (1929) views are seen as a revival of von Haller's doctrine of equipotentiality.[169] Recall that von Haller had thought that all areas of the brain are equally important in bringing about mental phenomena and thus, there was no regional specialisation. Likewise, Lashley maintained that there is just one neurological type responsible for realizing all mental types. In performing ablations on the cerebral cortex of animals he noticed that the subsequent learning deficits were proportional to the amount that was removed irrespective of the location (Sheperd 2010). It was through this discovery that he came to believe that there was only one neurological state kind (or type) in the cortex (Polger 2009) and that the entire cortex participates in all cognitive functions.

Equipotentiality, however, was eventually disproved by several discoveries (Huttenlocher 2002). Perhaps most important was the discovery that not all neurons are alike, or of the same type, as Lashley had maintained (Jabr 2012a,b; Sheperd 2010).[170] But what is worse for Block and Fodor is that, as Polger (2009) argues, even if Lashleyian equipotentiality were true it would not be evidence of multiple realizers. By maintaining that there is just one neurological type responsible for all mental types, the doctrine precludes the possibility of the same mental type being realized by distinct neurological types, which is what multiple realizability requires. This is what Block and Fodor themselves say in their next sentence, where they make a claim about the potential of plasticity in the nervous system: "a given type of psychological process is in fact often associated with a variety of distinct neurological structures."

---

[169] Precursors of Lashley's view can be found in von Haller's theory of equipotentiality and Flourens's concept of *action commune* (Clarke and Jacyna 1987). I discuss these views in Appendix A.

[170] Block and Fodor (1972: 160) do acknowledge that the "generality" of Lashley's views may be disputed.

Finally, in the last parenthetical sentence Block and Fodor give the example of homologous area adaptation that I have adopted in support of multiple localization.[171] As a reminder, we do not know enough about homologous area adaptation to know whether the neural structures that underlie language functions in the new hemisphere are of a distinct kind than those implicated in the original hemisphere. Certainly, for all Block and Fodor say, it is possible that the linguistic function established post trauma is realized by exactly the same kind of neural structure, only in a different region.[172] If this were the case it would not be evidence of multiple realizers, though there is evidence of multiple locations. What is needed to support multiple realizability (as Block and Fodor themselves say) is an example of a type of mental process (or psychological function) that can be realized by a variety of different neural (or lower-level) types. Thus, Block and Fodor cite two kinds of evidence—equipotentiality and homologous area adaptation—neither of which supports multiple realizability.

B3 Degeneracy and Pluripotency

Block and Fodor's misunderstanding of Lashley's views can be clarified by a distinction that is now made in modern cognitive ontology between two senses in which the brain can be said to be plastic: degeneracy and pluripotency. And I will argue that the concept of degeneracy can itself be clarified by the distinction between multiple realizability and multiple localizability.

---

[171] See Chapter Five of this thesis.

[172] As mentioned in Chapter Five, this raises the issue of what exactly the same kind of neural structure, or a difference in kind, amounts to—whether it is the type of neurons, the circuitry, a specific activity, or so on. Whether or not these homologous area adaptation relies on multiple realizers will hinge, in part, on what counts as 'sameness in kind' or a 'difference in kind' at the neurobiological level. I will not take a position in this thesis on whether or not homologous area adaptation does rely on multiple realizers, nor will I attempt to settle the issue of what should count as a neurobiological kind.

Lashley is sometimes considered an early supporter of neural plasticity (e.g. Block and Fodor 1972), while others describe him as an opponent (e.g. Belucchi and Buchtel 2009). In my view, the disagreement lies in what one means by neural plasticity. A relevant distinction here is between degeneracy, which refers to a one-to-many mapping between higher-level cognitive functions and lower-level neural structures, and pluripotency, which refers to a many-to-one mapping between functions and structures (Kiverstein and Miller 2015; Friston and Price 2003; Price and Friston 2002, 2005; Anderson 2010, 2014). The two are often seen as counterparts, although neither implies the truth of the other. Furthermore, while they both might be described as forms of neural plasticity, they refer to importantly different scenarios. Because Lashley maintained that there is just one neurological type, he thought that this single type had to take on multiple different functions. This view is thus compatible with pluripotency, but incompatible with degeneracy because it denies that there are distinct neurological types.[173]

Pluripotency is not evidence of the multiple realizers of mental kinds, but its counterpart, degeneracy, might be. Those working on degeneracy, however, make no distinction between multiple realizers and multiple locations (e.g. Kiverstein and Miller 2015, Friston and Price 2003, Price and Friston 2002, 2005). Price and Friston (2002: 416), for example, sometimes describe degeneracy as a one-to-many mapping between function and structures located in "different areas" and at other times as involving "structurally different mechanisms." To illustrate degeneracy they offer the example of how the same type of higher-level function of pressing a button can be done by different structures, such as the ten fingers (Price and Friston 2002); or how waving goodbye can be done by either one's left hand or one's right hand (Friston and Price 2003). Notice that in both of these examples, the different structures at the lower level are not different in kind, but rather tokens of the same kind, located in different places. In the case of cognitive functions, if the underlying neural structures are merely in different locations but the same in kind, as Price and Friston's examples suggest, then degeneracy is merely

---

[173] Actually, this depends on how 'degeneracy' is understood, as there is an ambiguity in how it is described by leading writers on this topic, including Kiverstein and Miller (2015), Friston and Price (2003), Price and Friston (2002, 2005). I will discuss this in the next section.

an example of multiple locations. In other words, multiple realizability would not be implied by degeneracy. Thus, there is a kind of ambiguity in how degeneracy has been described such that it is unclear whether it necessarily involves distinct *kinds* of neural structures or merely tokens of the same structural kind, only in different locations. It seems that degeneracy at least requires multiple locations.[174]

B4 Conclusion

We've seen that while some philosophers cite neuroplasticity, e.g. homologous area adaptation, as evidence of multiple realizers of mental types, others argue that this evidence fails to support multiple realization. But, I argue that this evidence does support mental kinds having multiple locations. This distinction is relevant for my non-functionalist parity argument for the extended mind thesis, which I advanced in Chapter Five, but it also brings out an ambiguity in current explanations of 'degeneracy' in cognitive ontology. This later point was meant to show how multiple localization is both compatible with and useful for network-based analysis.

---

[174] This ambiguity may merit a lengthier discussion, but because it is beyond the scope of the thesis, I will take it up elsewhere.

WORKS CITED

Acharya, S., S. Shukla, S.N. Mahjan, and S.J. Diwan. (2012). Localizationism to Neuroplasticity—The Evolution of Metaphysical Neuroscience. *The Journal of the Associations of Physicians of India, 60*: 38-46.

Adams, F. and Aizawa, K.. (2001). The Bounds of Cognition. *Philosophical Psychology, 14*(1): 43-64.

——— (2005). Defending non-derived content, *Philosophical Psychology, 18*(6): 661–669.

——— (2008). *The Bounds of Cognition*. Blackwell Publishing.

Aizawa, K. (2008). Why the Coupling-Constitution Fallacy is Really Still a Problem. 1-4. Available online: <https://studylib.net/doc/5874688/why-the-coupling-constitution-fallacy-is-really-still-a-p...>

——— (2009). Neuroscience and Multiple Realization: A Reply to Bechtel and Mundale. *Synthese, 167*: 493–510.

——— (2016). Commentary on "The what and the where of mental types" by Karina Vold. MindsOnline Conference. Brains Blog. Available online: <http://mindsonline.philosophyofbrains.com/2016/2016-2/the-what-and-where-of-mental-states-on-what-is-distinctive-about-the-extended-mind-thesis/>

American Honda Motor Co. (2017). ASIMO: The World's Most Advanced Humanoid Robot. Available online: <http://asimo.honda.com>

Anderson, M.L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences, 33*(3): 245-313.

——— (2014). *After Phrenology: Neural reuse and the interactive brain*. (MIT Press).

Armstrong, D. M. (1968). *A Materialist Theory of Mind*. Routledge and Kegan Paul.

——— (1981). *The nature of the mind*. University of Queensland Press.

Barker, M. J. (2010). From cognition's location to the epistemology of its nature. *Cognitive Systems Research, 11*: 357–366.

Bechtel, W. (2009). Looking Down, Around, and Up: Mechanistic Explanation in Psychology. *Philosophical Psychology, 22*(5): 543-564.

Bechtel, W. and Graham, G. (eds). (1999). *A Companion to Cognitive Science*. (Wiley-Blackwell).

Bechtel, W. and Mundale, J. (1999). Multiple Realizability Revisited: Linking Cognitive and Neural States. *Philosophy of Science, 66*(2): 175-207.

Bechtel, W. and Richardson, R. C.. (2010). *Discovering Complexity. Decomposition and Localisation as Strategies in Scientific Research*. MIT Press.

Bennett, M.R. and Hacker, P. M. S. (2008). *History of Cognitive Neuroscience*. Wiley-Blackwell Publishing.

Bergson, H. (1911). *Matter and Memory*. Translated by N. M. Paul and W. S. Palmer. (2004). Courier Corporation.

Berlucchi, G. and Buchtel, H.A. (2009). Neuronal plasticity: Historical roots and evolution of meaning. *Experimental Brain Research, 192*(3): 307–319.

Bickle, J., (2013). Multiple Realizability, *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/spr2013/entries/multiple-realizability/>.

Block, N. (1978). Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science, 9*: 261-325.

——— (1990). Can the Mind Change the World? In G. Boolos (ed.), *Meaning and Method: Essays in Honor of Hilary Putnam*. New York: Cambridge: 137-70.

Block, N. and Fodor, J.A.. (1972). What Psychological States Are Not. *Philosophical Review, 81*: 159-81.

Boag, Z. Interview with John Searle, titled "Searle: it upsets me when I read the nonsense written by my contemporaries." *New Philosopher*, January 25, 2014. Available online: <http://www.newphilosopher.com/articles/john-searle-it-upsets-me-when-i-read-the-nonsense-written-by-my-contemporaries/>

Braddon-Mitchell, D and F. Jackson. (2007). *Philosophy of Mind and Cognition: And Introduction*, 2nd ed. Oxford: Blackwell Publishing.

Broca, P. (1861). Sur le principe des localisations cérébrales. *Bulletin de la Société*

d'Anthropologie, 2: 190–204.

——— (1865). Du siège de la faculté du langage articulé. *Bulletins de la Société d'Anthropologie, 6*: 377-393.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues* ('Comparative Localization Studies in the Brain Cortex, its Fundamentals Represented on the Basis of its Cellular Architecture'). Trans. by L. J. Garey. (1994) In *Brodmann's 'Localisation in the cerebral cortex'*. London, UK: Smith-Gordon.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy, 4*(1): 73-122.

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

——— (2008). Foreword to Andy Clark's *Supersizing the Mind*. In Clark, A. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*: ix-xvi.

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology, 15*(2): 181-195.

——— (2009). *Radical embodied cognitive science*. Cambridge: MIT Press.

Cherniak, C. (1984) Computational Complexity and the Universal Acceptance of Logic. *Journal of Philosophy, 81*(12), 739-758.

——— (1986) *Minimal Rationality*. Cambridge, MA: Bradford.

Chomsky, N. (1959). A Review of B. F. Skinner's *Verbal Behavior*. *Language, 35*(1): 26–58

Chugani, H.T., R. A. Muller, and D. C. Chugani. (1996). Functional brain reorganization in children. *Brain Development, 18*(5): 347-56.

Churchland, P. M., (1981). Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy, 78*: 67–90.

——— (2005). Functionalism at Forty. *Journal of Philosophy, 102*: 33-50.

Clancey, W. J. (2009). Scientific Antecedents of Situated Cognition. In *The Cambridge Handbook of Situated Cognition*. P. Robbins and M. Aydede (eds.). Cambridge University Press: 11-34.

Clark, A. (1996). *Being There: Putting Brain, Body, and World Together Again*. Cambridge: MIT Press.

——— (1998). Twisted tales: Causal complexity and cognitive scientific explanation. *Minds and Machines, 8*: 79-99.

——— (2008a). Pressing the flesh: exploring a tension in the study of the embodied, embedded mind. *Philosophy and Phenomenological Research, 76*(1): 37-59.

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

——— (2009a). Spreading the joy? Why the machinery of consciousness is (probably) still in the head. *Mind, 118*(472): 963–993.

——— (2009b). Letters: Reply to Fodor. *London Review of Books, 31*(6). Available online: http://www.lrb.co.uk/v31/n03/jerry-fodor/where-is-my-mind

——— (2010). Coupling, constitution, and the cognitive kind: A reply to Adams and Aizawa. In R. Menary (ed.), *The Extended Mind*. Cambridge: MIT Press: 81-100.

Clark, A. and D. Chalmers. (1998). The Extended Mind. *Analysis*, *58*: 7-19.

Clarke, E. and L. S. Jacyna. (1987). *Nineteenth-Century Origins of Neuroscientific Concepts*. University of California Press.

Collins, S., Wisse, M., and A. Ruina. (2001). A three-dimensional passive-dynamic walking robot with two legs and knees. *The International Journal of Robotics Research, 20*.7: 607-615.

Collins, S. and A. Ruina. (2005). A bipedal walking robot with efficient and human-like gait. *Proceedings IEEE International Conference on Robotics and Automation*, Barcelona, Spain.

Craver, C. (2006). When mechanistic models explain. *Synthese, 153*(3): 355-376.

——— (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Clarendon Press: Oxford.

Damasio, A. R. and H. Damasio. (1992). Brain and Language. *Scientific American, 267*(5): 88-95.

David, M. (1997). Kim's Functionalism. *Philosophical Perspectives*, *11, Mind, Causation, and World*: 133-148.

Davidson, D. (1985). Deception and Division. In J. Elster (ed.) *The Multiple Self*, Cambridge University Press: 79-92.

Davies, D. (1996). Explanatory disunities and the unity of science. *International Studies in the Philosophy of Science, 10*(1): 5-21.

Dawkins, R. (1982). *The Extended Phenotype*. New York: Oxford Press.

Dennett, D. (1986). The myth of original intentionality. In *Modelling the Mind*. W. Newtown Smith and R. Viale (eds). Oxford: Oxford University Press.

Descartes, R. (1996/1641). *Meditations on First Philosophy*. Translated by J. Cottingham. Cambridge: Cambridge University Press.

——— (1649). Passions of the Soul (Excerpts), In *Philosophy of Mind Classical and Contemporary Readings*, Chalmers, D. (ed.) Oxford Press, (2002): 21-23.

Di Paolo, E. (2009). Extended life. *Topoi, 28*: 9-21.

Di Paolo, E and E. Thompson. (2014). The Enactive Approach. In *The Routledge Handbook of Embodied Cognition*. L. Shapiro (ed.). New York, NY: Routledge.

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge: Harvard Press.

Drayson, Z. (2010). Extended Cognition and the Metaphysics of Mind. *Cognitive Systems Research, 11*(4): 367-77.

Drayson, Z., and A. Clark. (Forthcoming). Augmentation, Agency, and the Spreading of the Mental State. *<https://sites.google.com/site/zdrayson/papers>*

Egan, A., (2008). Seeing and believing: perception, belief formation and the divided mind. *Philosophical Studies, 140*(1): 47–63.

——— (2011). Comments on Gendler's 'The epistemic costs of implicit bias'. *Philosophical Studies, 156*: 65–79.

Eliasmith, C. (2009). Dynamics, control, and cognition, In Robbins, P. and Aydede, M. (eds.) *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press: 134–154.

Esfeld, M. (2001). *Holism in philosophy of mind and philosophy of physics*. Boston: Kulwer Academic Publishers.

Farkas, K. (2012). Two Versions of the Extended Mind Thesis. *Philosophia, 40*(3): 435-47.

Feigl, H. (1958). The "Mental" and the "Physical": The essay and a postscript. In *Minnesota Studies in the Philosophy of Science: Concepts, Theories, and the Mind-Body Problem* 2. H. Feigl, M. Scriven, and G. Maxwell (eds.) University of Minnesota Press: 370-497.

Finger, S. (2010). Chapter 10: The birth of localization theory. *Handbook of Clinical Neurology, 95*: 117-128.

Flourens, P. (1846). *Phrenology examined*. Hogan and Thompson, Philadelphia.

Fodor, J.A. (1974). Special Sciences, or the Disunity of Science as a Working Hypothesis. *Synthese, 28*: 97-115.

——— (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences, 3*(1): 63-73.

——— (2000). *The mind doesn't work that way: The scopes and limits of computational psychology*. Cambridge, MA: MIT Press.

——— (2009). Where is my mind? *London Review of Books, 31*(3): 13-15.

Friston, K. (2002). Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience, 25*: 221–250.

Friston, K. J. and Price, C. J. (2003). Degeneracy and redundancy in cognitive anatomy. *Trends in Cognitive Science, 7*: 151–152.

Fritsch, G.T. and Hitzig, E. (1870). Uber die elektrische Erregbarkeit des Grosshirns. *Arch. Anat. Physiology Wissen. Med, 37:* 300-32.

Gall, F.J. (1822-1825). *Sur les Fonctions du Cerveau et sur celles de chacune de ses Parties*. Bailliere, Paris.

Gallagher, S. (2005). *How the Body shapes the Mind*. Oxford University Press.

——— (2009). Philosophical Antecedents of Situated Cognition. In The Cambridge Handbook of Situated Cognition. P. Robbins and M. Aydede (eds.). Cambridge University Press. 35-51.

——— (2013). The socially extended mind. Cognitive Systems Research. 25-26: 4-12.

Garey, L. J. (1994). *Brodmann's 'Localisation in the cerebral cortex'*. London, UK: Smith-Gordon.

Garrett, R. K. (2011). Troubling consequences of online political rumoring. *Human Communication Research* 37.2: 255-274.

Garrett, R. K., E. C. Nisbet, and E. K. Lynch. (2013). Undermining the corrective effects of media-based political fact checking? The role of contextual cues and naïve theory. *Journal of Communication* 63.4: 617-637.

Garrett, R. K., E. C. Weeks, and R. L. Neo. (2016). Driving a wedge between evidence and beliefs: How online ideological news exposure promotes political misperceptions. *Journal of Computer-Mediated Communication* 21.5: 331-348.

Gertler, B. (2007). Overextending the Mind? In *Arguing about the mind*. B. Gertler and L. Shapiro, eds. (Routledge). Citations refer to version available online: *<people.virginia.edu/~bg8y/ExMind4AAM.doc>*. pp. 1-13.

——— (2016). Commentary on "The what and the where of mental types" by Karina Vold. MindsOnline Conference. Brains Blog. Available online: <http://mindsonline.philosophyofbrains.com/2016/2016-2/the-what-and-where-of-mental-states-on-what-is-distinctive-about-the-extended-mind-thesis/>

Gibson, J. J. (1977). The theory of affordances. In *Perceiving, acting, and knowing*, R.E. Shaw and J. Bransford (eds.) Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.: 67-82.

——— (1979). *Ecological Approach to Visual Perception*. Houghton Mifflin Publishing.

Grafman, J. (2000a). Picking Two Scientific Roses for the Next Century. *Brain and Cognition* 42: 10-12.

——— (2000b). Conceptualizing functional neuroplasticity. *Journal of Communication Disorders* 33.4: 355-6.

Grafman, J. and I. Litvan. (1999). Evidence for Four Forms of Neuroplasticity. In *Neuronal Plasticity: Building a Bridge from the Laboratory to the Clinic*. J. Grafman and Y. Christen (eds.). Springer-Verlag Lublishers.

Grodzinsky, Y. and K. Amunts (eds.). (2006). *Broca's Region*. Oxford University Press.

Gross, C.G. (1998). Brain, Vision, Memory. MIT Press, Cambridge, MA.

——— (2009). *A Hole in the Head: More Tales in the History of Neuroscience*. Cambridge: MIT press.

Heidegger, M. (1927). *Being and Time*. Translated by John Macquarrie and Edward Robinson. (1962). Harper and Row Publishers.

Hurley, S. L. (1998a). Vehicles, Contents, Conceptual Structure and Externalism. *Analysis* 58: 1-6.

——— (1998b). *Consciousness in Action*. Harvard University Press.

——— (2010). Varieties of Externalism. R. Menary (ed.), *The Extended Mind*. MIT Press: 101-153

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

Huttenlocher, P. R. (2002). *Neural Plasticity: The Effects of Environment on the Development of the Cerebral Cortex*. Harvard University Press.

Hutto, D. D., and Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content*, Cambridge, MA: MIT Press.

Jabr, Ferris. (2012a). Know Your Neurons: The Discovery and Naming of the Neuron. *Scientific American: Brainwaves*. Available online: <https://blogs.scientificamerican.com/brainwaves/know-your-neurons-the-discovery-and-naming-of-the-neuron/>

——— (2012b). Know Your Neurons: How to Classify Different Types of Neurons in the Brain's Forest. *Scientific American: Brainwaves*. Available online: <https://blogs.scientificamerican.com/brainwaves/know-your-neurons-classifying-the-many-types-of-cells-in-the-neuron-forest/#>

Jarvis, B. (2014). Evaluating the Extended Mind. *Philosophical Issues* 24.1: 209-229.

Kandel, E., J. Schwartz, and T. Jessell. (2000). *Principles of Neural Science*, Fourth Edition. McGraw-Hill Medical.

Kaplan D., and W. Bechtel. (2011). Dynamical Models: An Alternative or Complement to Mechanistic Explanations? *Topics in Cognitive Science* 3: 438-44.

Kaplan, D. and C. Craver. (2011). The explanatory force of dynamical and mathematical models in neuroscience: a mechanistic perspective. *Philosophy of Science* 78.4: 601-28.

King, C. (2016). Learning Disability and the Extended Mind. *Essays in Philosophy, 17*(2): 38-68.

Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge: MIT Press.

——— (2011). *Philosophy of Mind*. Third Edition. Boulder, Col.: Westview.

Kirsh, D. (1995). The Intelligent use of space. *Artificial Intelligence, 73*(1-2): 31-68.

Kirsh, D. and P. Maglio. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science 18*: 513-449.

Kiverstein, J., and M. Miller. (2015). The Embodied Brain: Towards a Radical Embodied Cognitive Neuroscience. *Frontiers in Human Neuroscience, 9*(237): 1–11.

Lancaster, J. L. and P. T. Fox. (2009). Chapter 38: Talairach Space as a Tool for Intersubject Standardization in the Brain. *Handbook of Medical Image Processing and Analysis*, Second Edition. Elsevier: 629–641.

Lamme, V. and Spekreijse, H. (1998). Neural synchrony does not represent texture segregation, *Nature, 396*: 362–366.

Lashley, K. S. (1929). *Brain mechanisms and intelligence: A quantitative study of injuries to the brain*. Chicago: University of Chicago Press.

Levin, H.S., J. Scheller, T. Rickard, J. Grafman, K. Martinkowski, M. Winslow, and S. Mirvis. (1996). Dyscalculia and dyslexia after right hemisphere injury in infancy. *Archives of Neurology, 53*: 88–96.

Levy, N. (2007). *Neuroethics: Challenges for the 21ˢᵗ Century*. Cambridge: Cambridge University Press.

Lewis, D. (1970). How to Define Theoretical Terms. *Journal of Philosophy, LXVII*(13): 427-46.

——— (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy, 50*(3): 249-58.

Lyre, H. (2009). The "multirealization" of multiple realizability. In *Reduction and Elimination in Philosophy and the Sciences: Proceedings of the 31st Wittgenstein-Symposium*. A. Hieke and H. Leitgeb (eds.). Ontos verlag, Frankfurt.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman.

Marshall, J. C. and Fink, G. R. (2003). Cerebral localization, then and now. *NeuroImage, 20*: S2–S7.

von Melchner, L., S. L. Pallas, and M. Sur. (2000). Visual behavior induced by retinal projections directed to the auditory pathway. *Nature, 404*: 871-5.

Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. Palgrave Macmillan.

——— (ed.) (2010). *The Extended Mind*. MIT Press.

Merleau-Ponty, M. (1962). *Phenomenology of Perception*. London: Routledge and K. Paul; New York: Humanities Press.

Milojevic, M. (2013). Functionally Extended Cognition. *Prolegomena, 12*(2): 315-336.

Mindell, D. (2002). *Between human and machine: feedback, control, and computing before cybernetics*. Baltimore: Johns Hopkins University Press.

Miyazono, K. (2015). Does Functionalism entail extended mind? *Synthese*: 1-19.

Mogensen, J. (2011). Reorganization of the injured brain: implications for studies of the neural substrate of cognition. *Frontiers in Psychology, 2*(7): 1-10.

Mogensen, J. and Malá, H. (2009). Post-traumatic functional recovery and reorganization in animal models. A Theoretical and methodological challenge. *Scandinavian Journal of Psychology, 50*: 561-573.

Mundale, J. (2002). Concepts of Localization: Balkanization in the Brain. *Brain and Mind, 3*: 313-30.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

——— (2009). *Out of our heads*. Hill and Wang.

Overgaard, M. and J. Mogensen. (2011). A Framework for the Study of Multiple Realizations: the Importance of Levels of Analysis. *Frontiers in Psychology, 2*(39): 1-10.

Palermos, S.O. (2014). Loops, Constitution and Cognitive Extension. *Cognitive Systems Research, 27*: 25-41.

Parthemore, J. (2011). Of Boundaries and Metaphysical Starting Points: Why the Extended Mind Canët Be So Lightly Dismissed. *Teorema, 30*(2): 79-94.

Pears, D. F., and Pugmire, D. (1982). Motivated Irrationality. *Proceedings of the Aristotelian Society*, Supplementary Volumes, *56*: 157-196.

Piccinini, G. (2004). Functionalism, computationalism, and mental states. *Studies in the History and Philosophy of Science, 35*: 811-833.

Pinker, S. (1994). *The Language Instinct*. New York, NY: HarperCollins.

Pitt, D. (2017). Mental Representation. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.). Available Online: <https://plato.stanford.edu/archives/spr2017/entries/mental-representation/>.

Place, U. T. (1956). Is consciousness a brain process? *British Journal of Psychology, 47*: 44-50.

Poldrack, R. (2010). Mapping mental function to brain structure: How can cognitive neuroimaging succeed? *Perspectives in Psychological Science, 5*(5): 753-761.

Poldrack, R., J.A. Mumford, and T.E. Nichols. (2011). *Handbook of Functional MRI Data Analysis*. Cambridge University Press.

Polger, T. W. (2008). Computational functionalism. In P. Calvo and J. Symons (eds.), *The routledge companion to the philosophy of psychology*. London: Routledge.

――― (2009). Evaluating the Evidence for Multiple Realization. *Synthese, 167:* 457-472.

Polger, T. W. and L. A. Shapiro. (2016). *The Multiple Realization Book*. Oxford University Press.

Posner, M. I., S. E. Petersen, P. T. Fox, and M. E. Raichle. (1988). Localization of Cognitive Operations in the Human Brain. *Science, 240*(4859): 1627-31.

Price, C. J., and Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences, 6*(10): 416-421.

Price, C. J., and Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology, 22*(3): 262–275.

Putnam, H. (1960). Minds and Machines. In S. Hook, ed., *Dimensions of Mind*. New

York: New York University Press: 148-79.

——— (1963). Brains and Behavior. In Ronald J. Butler (ed.), *Analytical Philosophy: Second Series*. Blackwell.

——— (1967). The Nature of Mental States. (Originally published as 'Psychological Predicates'. In *Art, Mind, and Religion* (eds.) W. H. Capitan and D. D. Merrill. Pittsburgh: University of Pittsburgh Press: 37-48.) Reprinted in Rosenthal (1971): 150-61.

——— (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science, 7*:131-193.

——— (1988). *Representation and Reality*. Cambridge, MA: MIT Press.

Pylyshyn, Z.W. (1980). Computation and cognition: issues in the foundations of cognitive science. *The Behavioral and Brain Sciences, 3*: 111-69.

Robbins, P. and Aydede, M. (eds.) (2009). *The Cambridge Handbook of Situated Cognition*. Cambridge University Press.

Rorty, A. O. (1988). The Self-Deceptive Self: Liars, layers and Liars. In B. P. McLaughlin and A.O. Rorty (eds), *Perspectives on Self-Deception*, University of California Press, Berkeley: 11-28.

Rosen H J, Petersen S E, Linenweber M R, Snyder A Z, White D A, Chaoman L, Dromerick A W, Fiez J A & Corbetta M D (2000). Neural correlates of recovery from aphasia after damage to left inferior frontal cortex. *Neurology, 55*: 1883-1894.

Rosenthal, D. M. (1971). *Materialism and the Mind-Body Problem*. Englewood Cliffs, NJ: Prentice-Hall.

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

——— (2009a). Enactivism and the Extended Mind. *Topoi, 28*(1): 56-62.

——— (2009b). The extended mind. *Zygon: Journal of Religion & Science 44*(3): 628-641.

——— (2010). *The New Science of the Mind*. Cambridge, MA: MIT Press.

Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy, 101*(8): 389-428.

——— (2009). *Cognitive Systems and the Extended Mind*. Oxford University Press.

——— (2010a). Extended Cognition and the Priority of Cognitive Systems, *Cognitive Systems Research, 11*: 343-56.

——— (2010b). Systems, functions, and intrinsic natures: On Adams and Aizawa's The Bounds of Cognition. *Philosophical Psychology, 23*(1): 113–123.

Russell, B. (1918). The Philosophy of Logical Atomism. *The Monist, 28-29:* 495–527.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences, 3*: 417-457.

——— (1982) The Chinese Room revisited: Response to further commentaries on 'Minds, Brains, and Programs', *Behavioral and Brain Sciences, 5* (2):345–348.

——— (1992) *The Rediscovery of the Mind*, Cambridge, MA: MIT Press.

——— (1997). Breaking the hold: Silicon brains, conscious robots, and other minds, in Flanagan, O.J., Block, N. and Guzeldere, G. (eds.) *The Nature of Consciousness*, Cambridge, MA: MIT Press: 451–559.

Schwitzgebel, E. (2010). Acting contrary to our professed beliefs, or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly, 91*: 531–553.

Shagrir, O. (1998). Multiple realization, computation and the taxonomy of psychological states. *Synthese, 114*: 445-461.

Shapiro, L. A. (2004). *The Mind Incarnate*. MIT Press.

——— (2008). Functionalism and the boundaries of the mind. *Cognitive Systems Research, 9*(5): 5-14.

——— (2011). *Embodied Cognition*. Taylor and Francis.

——— (2013). Dynamics and Cognition. *Minds and Machines, 23*: 353–375.

——— (2014). The Routledge Handbook of Embodied Cognition. New York, NY: Routledge.

Shapiro, L. A. and S. Spaulding. (2009). Review of Andy Clark: *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Notre Dame Philosophical Reviews. Available online at: <http://ndpr.nd.edu/news/supersizing-the-mind-embodiment-action-and-cognitive-extension/>

Sharma, J., A. Angelucci, and M. Sur. (2000). Induction of visual orientation modules in auditory cortex. *Nature, 404*: 841-847.

Sheperd, G.M. (2010). *Creating Modern Neuroscience: The Revolutionary 1950s*. Oxford University Press.

Silberstein, M., and A. Chemero. (2013). Constraints on Localization and Decomposition as Explanatory Strategies in the Biological Sciences, *Philosophy of Science, 80*: 958-970.

Simon, H. (1969). *The Sciences of the Artificial*. MIT Press.

Singer, W. (2003). Synchronization, binding and expectancy. In Arbib, M. (ed.) *The Handbook of Brain Theory and Neural Networks*, 2nd ed., Cambridge, MA: MIT Press: 1136–1143.

Skinner, B. F. (1953). *Science and Human Behavior*, New York: Macmillan.

——— (1957). *Verbal Behavior*. Harvard University Press.

——— (1977). Why I am not a Cognitive Psychologist. *Behaviorism, 5*(2): 1-10.

Smart, J. J. C. (1959). Sensations and Brain Processes. *Philosophical Review*, *68*: 141–156.

Sprevak, M. (2009). Extended Cognition and Functionalism. *Journal of Philosophy, 106*: 503-27.

Springer, S. and G. Deutsch. (1993). *Left brain, Right brain*. Freeman, New York.

Stanciu, M. (2014). "Personal Identity, Functionalism, and the Extended Mind", *Social and Behavioral Sciences, 127*: 297-301.

Stemmer, B. (2006). Imaging Brain Lateralization: Words, Sentences, and Influencing Factors in Healthy, Pathological, and Special Populations. In *Encyclopedia of Language & Linguistics,* Second Edition. Elsevier: 545-553.

Sterelny, K. (2004). Externalism, epistemic artefacts and the extended mind. In Richard Schantz (ed.), *The Externalist Challenge*. Walter de Gruyter, Berlin: 239-254.

Sutton, J. (2006). Distributed Cognition: Domains and dimensions. *Pragmatics and Cognition, 14*(2): 235-247.

——— (2010a). Exograms and interdisciplinarity: history, the extended mind, and the civilizing process. In Richard Menary (ed.), *The Extended Mind*. Cambridge, MA: MIT Press: 189-225.

Sutton, J. C. B. Harris, P. G. Keil, and A. J. Barnier. (2010b). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the Cognitive Sciences, 9*(4), 521-560.

Szaflarski, J. P., Binder, J. R., Possing, E.T., McKiernan K. A., Ward, B.D., and T.A. Hammeke. (2002). Language lateralization in left-handed and ambidextrous people: fMRI data. *Neurology, 59*(2): 238-44.

Talairach, J. and Szikla, G. (1967). Atlas of stereotactic concepts to the surgery of epilepsy. Paris: Masson and Cie.

——— (1980). Application of stereotactic concepts to the surgery of epilepsy. *Acta Neurochirurgica Supplement, 30*: 35-54.

Thagard, P. (1992). *Conceptual Revolutions*. Princeton University Press.

——— (1996). *Mind: Introduction to Cognitive Science*. MIT Press.

Theiner, G. (2011). *Res cogitans extensa : a philosophical defense of the extended mind thesis*. Frankfurt, New York: P. Lang, Internationaler Verlag der Wissenschaften.

Thelen, E., Schoner, G., Scheier, C., and Smith, L.B. (2001). "The Dynamics of Embodiment: A Field Theory of Infant Preservative Reaching." *Behavioral and Brain Sciences, 24*: 1-86.

Thiele, A. and G. Stoner. (2003). Neural synchrony does not correlate with motion coherence in cortical area MT. *Nature, 421*: 366–370.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: MIT Press.

Turing, A. (1936). On computable numbers, with an application to the Entscheidungsproblem. In *Proceedings of the London Mathematical Society, Series 2*(42): 230-265.

Uttal, W. (2001). The New Phrenology: The limits of localizing cognitive processes in the brain. Cambridge, MA: MIT Press.

——— (2011). Mind and Brain: A critical appraisal of cognitive neuroscience. Cambridge, MA: MIT Press.

Varela, F., E. Thompson, and E. Rosch. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Vold, K. (2015). The Parity Argument for Extended Consciousness. *Journal of Consciousness Studies, 22*(3-4): 16-33.

Vygotsky, L. S. (1930). *Mind in Society: he development of higher psychological processes*. M. Cole, V. John-Steiner, S. Scribner, and E. Souberman (eds). Harvard University Press.

——— (1965). Psychology and localization of functions. *Neuropsychologica, 3*: 381-386. Translated by A. R. Luri.

Wadham, J. (2016). Common-Sense Functionalism and the Extended Mind. *Philosophical Quarterly, 66*(262): 136-151.

Walker, A. E. (1998). *The Genesis of Neuroscience*. The American Association of Neurological Surgeons and A.M. Walker.

Watson, J. (1930). *Behaviorism*, Norton: New York.

Weiner, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine*. Hermann and Cie; MIT Press.

Weiskopf, D. (2008). Patrolling the mind's boundaries. *Erkenntnis, 68*(2): 265-76.

——— (2010). The Goldilocks problem and extended cognition. *Cognitive Systems Research, 11*(4): 313-323.

Weiskopf, D. and F. Adams. (2015). *An Introduction to the Philosophy of Psychology*. Cambridge University Press.

Wernicke, C. (1874). *Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis*. Breslau, M. Crohn und Weigert. (Title in English: The aphasic symptom complex: a psychological study from an anatomical basis.)

Wheeler, M. (2005). *Reconstructing the cognitive world*. Cambridge, MA: MIT Press.

——— (2010a). In Defense of Extended Functionalism. In Menary, R. (ed.). *The extended mind*. Cambridge: MIT Press: 245-70.

——— (2010b). Minds, Things, and Materiality. In L. Malafouris and C. Renfrew. (eds.), *The Cognitive Life of Things: Recasting the Boundaries of the Mind*. Cambridge: McDonald Institute Monographs. (Reprinted in J. Schulkin (ed.), Action, Perception and the Brain: Adaptation and Cephalic Expression. Basingstoke: Palgrave Macmillan.)

——— (2011). In Search of Clarity about Parity. *Philosophical Studies*, symposium on Andy Clark's *Supersizing the Mind*, *152*(3): 417-25.

——— (2016). Commentary on "The what and the where of mental types" by Karina Vold. MindsOnline Conference. Brains Blog. Available online: <http://mindsonline.philosophyofbrains.com/2016/2016-2/the-what-and-where-of-mental-states-on-what-is-distinctive-about-the-extended-mind-thesis/>

Wheeler, M. and Clark, A. (1999). Genic representation: reconciling content and causal complexity. *The British Journal for the Philosophy of Science, 50*(1): 103-135.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review, 9*(4): 625-636.

Wilson, R. A. (1994). Wide Computationalism. *Mind, 103*(411): 351-72.

——— (2004). *Boundaries of the mind: The individual in the fragile sciences–cognition*. Cambridge, UK: Cambridge University Press.

——— (2014). Ten Questions Concerning Extended Cognition. *Philosophical Psychology, 27*(1): 19-33.

Wilson, R. A. and A. Clark. (2009). How to situate cognition: Letting nature take its course. In Murat Aydede and P. Robbins (eds.), *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press: 55-77.

Wittgenstein, L. J. J. (2001/1953). *Philosophical Investigations*, G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell Publishing.