Enactivist Interventions

Rethinking the Mind

Shaun Gallagher

OXFORD
UNIVERSITY PRESS
2017

Contents

1

Introduction

Is cognition *in the head* or *in the world*, or in some mix of brainy and worldly processes? Continuing research on embodied cognition in philosophy of mind and the cognitive sciences has motivated numerous debates about questions such as this. There's a strong tradition in both philosophy of mind and cognitive science that takes cognition to be a fully in-the-head event. In this introductory chapter I'll discuss a few of the more recent versions of this view. I'll then begin to sketch the contrasting view of enactivism, an embodied cognition approach that has roots in phenomenology and pragmatism.

Enactivist approaches to cognition suggest that, at least in basic (perception- and action-related) cases, cognitive processes are not just in the head, but involve bodily and environmental factors. This view clearly poses a challenge to what has been the standard science of cognition, especially to cognitive neuroscience, and to any science that claims to provide full and exclusive explanations in terms of one factor, e.g., neural processing. If cognition is not reducible to brain processes, or to any other single factor, and if indeed it does involve many other aspects of embodiment and environment, then how precisely should a scientific study of the mind proceed? Can there be an enactivist science of mind? In sketching out an answer to this question, a number of issues, involving intentionality, representation, affect, agency, and so on, come into focus and I address these issues more fully in subsequent chapters.

1.1 Cognition-in-the-Head: Some Recent Approaches

Even if we define cognitive processes broadly to include not just beliefs and desires, but also states that refer to bodily action and to interactions with other people, we still find that mainstream cognitive science offers narrow accounts that place all the action required for full explanation in mental states that correspond strictly to brain processes. The term 'narrow' is a technical one in philosophy of mind, referring to internal mental representational processing or content. Standard explanations in cognitive science define cognition as constituted by mental or neural representations. I explore the complex question of what counts as a representation in *Chapter 5* (section 5.2 and following). For introductory purposes, a few examples will provide a good sense of this approach.

The first example concerns action—specifically aspects of action that involve planning and intention formation. The processes involved can be characterized at subpersonal and personal levels of explanation, but all of them remain narrowly within the traditional boundary of the mind-brain. Consider the well-known Libet experiments. Libet and colleagues (1983; also see Soon et al. 2008) asked about neural dynamics involved in the readiness potential (*Bereitschaftspotential*)—a brain signal that begins approximately 800 milliseconds prior to any particular bodily movement—and its relation to our immediate sense of deciding to act. The experiments are not only well known but also controversial. Here I won't go into details about the experiments (see section 7.3 for more details) or about controversies that pertain to methodology, but I will summarize briefly the basic idea and say something about the philosophical controversy.

The question Libet tried to answer is whether consciousness plays a role in the initiation of action, and he interpreted this to be a question about free will. Libet's results indicated that on average, 350 milliseconds before the subject is conscious of deciding (or of having an urge) to move, the subject's brain is already working on the motor processes that will result in the movement. That is, the readiness potential is already underway, and the brain is preparing to move before the subject makes the decision to move. The conclusion is that voluntary acts are 'initiated by unconscious cerebral processes before conscious intention appears' (Libet 1985, 529).

There are different interpretations of what these results mean. Most of them focus on the question of free will. Libet himself finds room for free will in the approximately 150 milliseconds of brain activity remaining after we become conscious of our decision, and before we move. He suggests that we have time to consciously veto the movement. Others, however, think that the brain decides and then enacts its decisions; consciousness is epiphenomenal in this regard. The brain inventively tricks us into thinking that we consciously decide to act and that our actions are controlled at a personal level. On this view, free will is nothing more than a false sense or illusion.

For purposes of this introduction I will simply point to a central assumption made about the kind of cognitive processes that are supposed to be involved in free will. The assumption is nicely expressed by Haggard and Libet (2001), who frame the question, and refer to it as the traditional question of free will: 'how can a mental state (my conscious intention) initiate the neural events in the motor areas of the brain that lead to my body movement?' (47). They are right that this is the traditional way to ask the question: it's precisely the way that Descartes, and many thinkers in the modern philosophical tradition, would frame the question. It's the question of mental causation, which places the cognitive processes of free will in the head where brain and mind meet up.

To assume that this is the right way to ask the question overlooks the possibility that free will is not something that can be explained simply by looking where Libet experiments look. For example, one can argue that these experiments have nothing to do with free will. The latter interpretation challenges the assumption that free will can be characterized in terms of the short timescale of 150 milliseconds. This type of response can go one of two ways, however. The first way simply leads us back into the head, into discussions of intention formation where cognitive deliberations generate prior intentions that have a later effect on intentions-in-action. Because the Libet experiments address only motor intentions or, at best, intentions-in-action, they miss the mark since free will is more about deliberation and prior intention formation. Such explanations are worked out in representational terms of beliefs and desires in processes that are best characterized in terms of a space of reasons, but still very much in the head. The second way leads outwards into the world, and to the idea that free will is not a property of one individual brain, mind, or organism, but is relational, so that social and environmental factors contribute to or detract from our ability to act freely.

To say that something like social relations are involved in free will, however, does not necessarily lead beyond traditional concepts of the mind. This is clear when we turn to look at ongoing debates about social cognition or 'theory of mind' (ToM). These debates are framed in terms of methodological individualism, i.e., the idea that ToM can be explained by a causal mechanism (a ToM module or a mirror system) located within the individual. Today the growing consensus is that there are two networks in the brain responsible for our ability to understand others: (1) a ToM network that includes the temporo-parietal junction, medial parietal cortex, and medial prefrontal cortex (e.g., Saxe et al. 2009), and allows for some form of theoretical-inferential mindreading concerning the other person's mental states; and (2) mirror areas in premotor and parietal cortexes, supporting a mental simulation of the other's actions, intentions and emotions. Taken together, the neuroscientific findings may justify a hybrid style of mindreading, or suggest a two-system approach of online perspective taking and offline social reasoning (Apperly and Butterfill 2009).

Complicating such views, however, mainstream theories of social cognition have started to take note of objections coming from embodied cognition (EC) and action-oriented approaches. This and more general concerns about the claims made by EC theorists have motivated a way of thinking about the role of the body that retains a standard representationalism—so-called 'weak' (Alsmith and Vignemont 2012) or minimal EC (see section 2.1). For example, Alvin Goldman and Frederique de Vignemont (2009) suggest that none of the many things that EC theorists usually count as important contributors to cognitive processes—anatomy and body activity (movements and postures), autonomic and peripheral systems, relations with the environment—really do count. Rather, the only 'bodily' things relevant to an account of cognition in general, or social cognition in particular, are body-formatted (or B-formatted) representations in the brain. As they put it, B-formatted representations offer a 'sanitized' way of talking about the body, and 'the most promising' way to promote EC (2009, 155).

B-formatted representations are not propositional or conceptual in format; their contents may include the body or body parts, but also action goals, represented in terms of how to achieve such goals by means of bodily action. Somatosensory, affective, and interoceptive representations may also be B-formatted, 'associated with the physiological conditions of the body, such as pain, temperature, itch, muscular and visceral sensations, vasomotor activity, hunger and thirst' (156).

Social cognition, on this weak EC view, is embodied only to the extent that B-formatted representations involved in perceptual mirroring are used to represent the actions or mental states of others. Similar strategies aiming to 'sanitize' embodied cognition more generally can be found in accounts of broader aspects of cognition. As one example, several theorists point to body-related simulations (representations) as important for language and concept processing (e.g., Glenberg 2010; Meteyard et al. 2012; Pezzulo et al. 2011; Pulvermüller 2005). Goldman (2012; 2014) argues that one can develop an overall EC approach simply by generalizing the use of B-formatted representations. All of this is consistent with the standard representationalist 'mentalistic enterprise' of reconstructing the world (Jackendoff 2002), of 'pushing the world inside the mind' (Meteyard et al. 2012), and a very narrow-minded conception of embodiment.

1.2 Cognition-in-the-World: Phenomenologically Inspired Enactivist Approaches

Enactivist approaches to cognition are inspired and informed by phenomenological philosophy. Varela, Thompson, and Rosch (1991), who first defined the enactivist approach, found significant resources in the phenomenological tradition for rethinking the mind. For example, Husserl's concept of the 'I can' (the idea that I perceive things in my environment in terms of what *I can* do with them); Heidegger's concept of the pragmatic ready-to-hand (*Zuhanden*) attitude (we experience the world primarily in terms of pre-reflective pragmatic, action-oriented use, rather than in reflective intellectual contemplation or scientific observation); and especially Merleau-Ponty's focus on embodied practice, which so influenced both Gibson's notion of affordances and Dreyfus's critique of classic cognitivism (also see Di Paolo 2005; Gallagher 2005a; Noë 2004; Thompson 2007). Less noted are relevant resources in the American pragmatist tradition; many of the ideas of Peirce, Dewey, and Mead can be considered forerunners of enactivism (see *Chapter 3*).

Enactivist versions of EC emphasize the idea that perception is *for action*, and that action-orientation shapes most cognitive processes. Most enactivists call for a radical change in the way we think about the mind and brain, with implications for methodology and for rethinking how we do cognitive science. Enactivist approaches can be characterized by the following background assumptions, explored in subsequent chapters of this book:[1]

1. Cognition is not simply a brain event. It emerges from processes distributed across brain–body–environment. The mind is embodied (see *Chapter 2*); from a first-person perspective embodiment is equivalent to the phenomenological concept of the lived body. From a third-person perspective the organism–environment is taken as the explanatory unit (*Chapters 3* and *9*).
2. The world (meaning, intentionality) is not pre-given or predefined, but is structured by cognition and action (*Chapter 4*).
3. Cognitive processes acquire meaning in part by their role in the context of action, rather than through a representational mapping or replicated internal model of the world (*Chapter 5*).
4. Enactivist approaches have strong links to dynamical systems theory, emphasizing the relevance of dynamical coupling and coordination across brain–body–environment (*Chapters 5* and *6*).
5. In contrast to classic cognitive science, which is often characterized by methodological individualism with a focus on internal mechanisms, enactivist approaches emphasize the extended, intersubjective, and socially situated nature of cognitive systems (see *Chapter 7* for how this relates to the problem of free will).
6. Enactivism aims to ground higher and more complex cognitive functions not only in sensorimotor coordination, but also in affective and autonomic aspects of the full body (*Chapters 8* and *9*).
7. Higher-order cognitive functions, such as reflective thinking or deliberation, are exercises of skillful know-how and are usually coupled with situated and embodied actions (*Chapter 10*).

Enactivist approaches are similar to the ideas of extended mind and distributed cognition insofar as all of these approaches argue that cognition is not entirely 'in the head', but rather is distributed across brain, body, and environment (e.g., Clark and Chalmers 1998). However, in contrast to the extended mind hypothesis, which embraces functionalism and finds a role for minimal representations, enactivists reject functionalism and claim that

[1] These assumptions are drawn from the following sources: Clark (1999); Di Paolo, Rohde, and De Jaegher (2010); Dominey et al. (2016); Engel (2010); Engel et al. (2013); Thompson and Varela (2001); Varela, Thompson, and Rosch (1991).

the material specifics of bodily processes shape and contribute to the constitution of consciousness and cognition in a way that is irreducible to representations, even B-formatted representations (see section 2.1). In contrast to Clark (2008a), for example, who argues that specific differences in body type or shape can be transduced and neutralized via the right mix of representational processing in order to deliver similar experiences or similar cognitive results, enactivists insist that biological aspects of bodily life, including organismic and emotion regulation of the entire body, have a permeating effect on cognition, as do processes of sensorimotor coupling between organism and environment. In regard to the latter processes, for example, Noë (2004; also see O'Regan and Noë 2001; Hurley 1998) developed a detailed account of enactive perception where sensory–motor contingencies and environmental affordances take over the work that had been attributed to neural computations and mental representations.

1.3 Causality, Constitution, and Diachronicity

One clear objection to both enactivist and extended mind proposals involves the relation between causality and constitution. Both theories are said to confuse causality with constitution (Adams and Aizawa 2008). It's an important objection because it points to a clear difference between these theories and the standard cognitivist approaches.

Adams and Aizawa (2008; Aizawa 2010) argue that the extended mind hypothesis, as well as enactivist approaches (see Aizawa 2014), make an unjustifiable inference from *causal* dependence (where bodily and environmental factors play a causal role in support of cognitive processes) to *constitutive* dependence (where the claim is that such factors actually are part of the cognitive processes). This is the causal-constitution (C-C) fallacy. For example, the use of a notebook or a smartphone to support memory should be understood as causally supporting or enabling a cognitive process, but not as being a cognitive process itself, as the extended mind hypothesis claims. The strict distinction between causality and constitution is closely tied to the idea that there is a 'mark of the mental' (a way to determine what processes count as cognitive and what processes do not). Adams and Aizawa, among others, argue that non-derived internal (brain-based) representational content is what constitutes the mark of the mental. Outside of that nothing counts as cognitive.

The standard cognitivist approach adopts a classic metaphysical view that causation and constitution are independent relations—facts about causal relations do not tell us anything about constitution (Bennett 2011; 2004). On this view, for example, it is possible that emotions are causally influenced by bodily or environmental factors, but what constitutes an emotion is just the mental event that is instantiated in the brain. Michael Kirchhoff (2015) argues that this view understands constitution to mean material or compositional constitution: a *synchronic* one–one, or many–one (where one thing is constituted by an aggregate of things) relation between spatially and materially coincident objects of different kinds. The classic example is that the statue of David is constituted by (but not caused by) the piece of marble that it is. The relation between the statue and the marble is what it is at any moment, and does not change (Gibbard 1975). If x constitutes (or composes) y, then x and y exist at the same place at the same time and they share the same material parts. This classic notion, however, does not account for processes and a theory of the mental has to account for processes.

To account for processes, enactivists appeal to the ideas of a dynamical system and diachronic constitution (Kirchhoff 2015). Brain, body, and environment are said to be dynamically coupled in a way that forms a system, and the coupling is not equivalent to identity of material parts; rather it involves physical relational processes. Significant changes in one part of the system will cause changes or adjustments in the other parts. For the enactivist just these dynamical causal relations constitute the system.

Because these processes occur on several timescales, it will be helpful to introduce a threefold distinction in temporal and dynamical registers. The following differentiation, based on neurobiology and phenomenology (see Varela 1999), can clarify the C-C issue, but will also have relevance to a number of the analyses developed in later chapters. Varela argued that cognition involves processes on the following three timescales, which I'll call the elementary, the integrative, and the narrative timescales.

1. The *elementary* scale (varying between 10 and 100 milliseconds)
2. The *integrative* scale (varying from 0.5 to 3 seconds)
3. The *narrative* scale involving memory (above 3 seconds)

The elementary scale is the basic timescale of neurophysiology. It corresponds to the intrinsic cellular rhythms of neuronal discharges roughly within the range of 10 milliseconds (the rhythms of bursting interneurons) to 100 milliseconds (the duration of an excitatory/inhibitory sequence of postsynaptic potential in a cortical pyramidal neuron). Neuronal processes on this scale are integrated in the second scale, which, at the neurophysiological level, involves the integration of cell assemblies. Phenomenologically, the integrative scale corresponds to the experienced living present, the level of a fully constituted cognitive operation; motorically, it corresponds to a basic action, e.g., reaching, grasping. On a dynamical systems interpretation, neuronal-level events on the elementary scale synchronize (by phase-locking) and form aggregates that manifest themselves as incompressible but complete acts on the integrative scale.[2] The narrative scale is meant to capture longer time periods that scale to complex actions and cognitive processes that may involve recollection, planning, intention formation, and so on. Further distinctions could be made (one could think of developmental and evolutionary timescales, for example), and other more rhythmic time patterns could be explicated, but for our purposes the threefold distinction should be sufficient.

On the standard notion of synchronic constitution, subpersonal, elementary-scale neuronal processes constitute contentful, representational mental processes that in some way scale up to conscious mental states. One might think of this as a form of identity theory. Identity theories usually posit mental state = brain state identities—a central-state materialism. All other factors—bodily, environmental, social, etc.—are causal but not constitutive. On some accounts they are, at best, derived or epiphenomenal relative to non-derived mentality. On the enactivist view, however, one requires a more nuanced distributed-state materialism. In their dynamical relations, neural and non-neural, including embodied, environmental, social, etc., may be causal in a way that they are also constitutive.

In contrast to standard synchronic views of constitution, then, enactivists propose a notion of diachronic constitution, where causality and constitution are not independent. Embodied mental processes (i.e., processes of the embodied-enactive mind), distributed across different factors/levels (neural, behavioral, environmental), and across different

[2] This currently has the status of a working hypothesis in neuroscience. See Thompson (2007, 332).

timescales, are constituted in a temporally integrated dynamical system. The constituent elements may very well be in complex, reciprocal causal relations with each other, but just these reciprocal causal relations make the mental process what it is. Thus, an intervention that changes the causal relations in a dynamical system will also change the system as a whole. In a gestalt (what Maurice Merleau-Ponty [1964] called a 'form' or 'structure') the whole is said to add up to more than the sum of its parts. In a dynamical gestalt composed of processes that unfold over time, and characterized by recursive reciprocal causality relations, changes in any processual part (above a certain threshold) will lead to changes in the whole, and changes in the whole will imply changes in the processual parts. In contrast to a synchronic, compositional notion of constitution, these kinds of causal relations are diachronically constitutive of the phenomenon. As Kirchhoff (2015) argues, the notion of a C-C fallacy, where constitution is defined synchronically, does not apply to the type of diachronic processes described in dynamical patterns.

This notion of constitution might be taken in either a strong existential sense or a weaker sense. In the strong sense the claim would be that if one significantly changes, destroys, or removes the causal coupling, the system ceases to exist as such. Coma or death may be good examples of this. If there is a living organism, however, there is always an environment and some kind of causal coupling. In the weaker sense of constitution the claim is simply that cognition is *what* it is because of the nature of the coupling. A change in the way the brain, body, and environment are related will change cognition. The nature of cognition depends on the instantiation of certain dynamical couplings such that a specific kind of cognition would not arise were it not for causal interactions that define the system. The claim here is not, for example, that the environment determines representational contents. One can be in a particular environment and be dreaming or hallucinating—that will depend on the state of the body (e.g., if one is sleeping) or the brain (e.g., if there is an imbalance of neurotransmitters), and that will change one's relation to the environment. One can change the causal interactions with the environment, for example, by putting the organism in a sensory deprivation chamber; that clearly will result in cognition that is different.[3]

[3] If one is in a sensory deprivation chamber, one may be hallucinating (and not perceiving something in a physical environment) precisely because the specifics of the

To be clear, enactivists don't deny the importance of the brain, but they understand the brain to be an integrated part of a larger dynamical system that includes body and (physical, social, and cultural) environments. The explanatory unit of cognition (perception, action, etc.) is not just the brain, or even two (or more) brains in the case of social cognition, but dynamic relations between organism and environment, or between two or more organisms, which include brains, but also include their own structural features that enable specific perception–action loops, which in turn effect statistical regularities that shape the structure and function of the nervous system (Gallagher 2005a; Thompson 2007).

If I reach out to grasp something (or someone), my hand is involved, as is my arm, my shoulder and back muscles, my peripheral nervous system as well as my vestibular system, no less than my brain, which in all of its complexity is making its own dynamical adjustments on the elementary timescale as part of this process of reaching out to grasp. A full account of the kinematics of this movement doesn't add up to an explanation of the action; nor does a full account of the neural activity involved. Likewise, if I reach a decision about how to act, the neural components of this activity are a necessary part, but also my location, and who I'm with, and my past practices, current physical skills, and health status, not to mention my mood, will to some degree play contributory roles in the decision formation. Some of these elements enter into the process on a narrative timescale and are not under my current control. In this respect, my body is not just a sensory–motor mechanism. Affect plays an important role—things like hunger, fatigue, physical discomfort or pain, as well as emotion and mood (see *Chapter 8*). Such things are not well behaved in terms of timescale—they involve all three scales. With respect to discussions of agency and free will, for example, whatever agentive action is, it is both constrained and enabled by all of these different factors. As Clark and Chalmers (1998, 9) suggest, if one of the extra-neural components is taken away, 'the system's behavioural competence will drop, just as it would if we removed part of its brain'. At the very least a removal (or an

organism–environment coupling are different. This should not be an argument for the irrelevance of the environment or of the specifics of dynamical coupling (cf. Prinz 2009).

addition) of any component will entail compensatory adjustments across the system.

Evan Thompson (2014) provides a nice analogy. Saying that cognition is just in the brain is like saying that flight is inside the wings of a bird. Just as flight doesn't exist if there is only a wing, without the rest of the bird, and without an atmosphere to support the process, and without the precise mode of organism–environment coupling to make it possible (indeed, who would disagree with this?), so cognition doesn't exist if there is just a brain without bodily and worldly factors. 'The mind is relational. It's a way of being in relation to the world' (Thompson 2014, 1). For some, these claims may seem obvious or even trivial, and yet we often find ourselves doing science as if the only things that counted as explanatory were neural representations.

Processes of social interaction are also not reducible to neuronal processes (or B-formatted representations) within the individual, since they include physical engagement with another person, who is not just a representation in my mind, but someone who can push back in a way that a mere representation cannot. Social interactions also include physical engagement in a socially defined environment, and processes of 'primary intersubjectivity', including affective processes where distinct forms of sensory–motor couplings are generated by one's perception and response to facial expression, posture, movement, gestures, etc. in rich pragmatic and social contexts (Gallagher 2005a). This is demonstrated in the kind of rich analyses one finds in conversational analysis (e.g., Goodwin 2000). Again, this is not to say that all the essential processes of social cognition are extra-neural. Mirror neurons may indeed make a contribution, not by simulating actions of others, repeating a small version of them inside one's head, but by being part of larger sensory–motor processes that respond to different interaction affordances. On the enactivist view, social cognition is an attunement process that allows me to perceive the other as someone to whom I can respond or with whom I can interact. In the intersubjective context, perception is often *for interaction* with others. In some cases, a relational understanding is accomplished in the social interaction between two people where some novel shared meaning (or some decision or even some misunderstanding) is instituted in a way that could not be instituted within the single brain of either one of them alone (De Jaegher, Di Paolo, and Gallagher 2010).

1.4 How to be an Embodied Theorist without Losing your Head

Take any example of cognition and one can run two different explanations—the standard representationalist one *versus* the enactivist one. Sometimes it seems to be simply a vocabulary substitution; sometimes the enactivist description seems to work better, especially if we think of examples that involve problem solving rather than belief; and other times the representationalist description seems to have the upper hand. Even when the representations involved are action-oriented, minimal, or B-formatted, there are clear differences in explanation.

Consider the well-known example of fielding (trying to catch) a ball (McBeath, Shaffer, and Kaiser 1995). We can run the account in both ways, where running it in one case means representing various aspects of speed and trajectory, and in the other case it means literally running rather than representing.

In the classic representational account the problem is first solved in the fielder's head. Speed and trajectory of the ball are calculated and reconstructed by the brain, which solves the problem offline and then simply sends instructive signals to the limbs to move in the most efficient way to catch the ball. As Michael Anderson (2014, 164) points out, this representation-rich view treats cognition as 'post-perceptual'—something added to perception to make sense of it. It's not likely that anyone still believes this story, and there's evidence against it since outfielders who are standing still are unable to reliably predict where the ball will land. Moreover, the account doesn't predict the actual pattern of movement that the fielder makes to catch the ball.

In a weak EC, *action-oriented representation* (AOR) account, calculations are made online as we move, but part of the process involves quick (on the elemental timescale) offline AORs formed in forward models that contribute to motor control. Sensory feedback is too slow to update the system in a timely fashion; the forward model generates a simulation or representation that anticipates sensory feedback from intended body positions on the run and allows for a fine-tuning of motor control. The AOR stands in, briefly, for a future state of some extra-neural aspect of the movement—a body position (or proprioceptive feedback connected with a body position), which is just about to be accomplished in the action of catching the ball. Since the model represents a state of the system that does not yet exist—a predicted motor state—it is said to be offline, or decoupled from the ongoing action (Clark and Grush 1999), and to occur in the self-contained brain. After catching the ball, such representations can then be simulated and taken further offline, reused, e.g., in memory systems, scaling up to enable additional cognitive states. The brain can run such offline models to accompany states in which no running and catching is involved at all—when, for example, I imagine or remember catching a ball. No need for the body itself or for 'a constant physical linkage' (Clark and Grush 1999, 7; Clark 1999).

On the enactivist account, in contrast, the fielder solves the problem without representations, by vision and movement. She runs on a curved line so as to keep the ball's trajectory through the visual field at a constant speed, i.e., visually stationary on the retina. This reliably gets the fielder to the catching spot (McBeath, Shaffer, and Kaiser 1995; Fink, Foo, and Warren 2009). There is no need to compute in-the-head mental representations—of the ball, its speed, its trajectory, and so on. Rather, the cognitive component of this action just is seeing the ball that is 'out there' in the world, and directly acting in the world. The processes involved are dynamical sensory–motor processes that are fully online. These processes do involve ongoing anticipation, but it's not clear in what sense such anticipatory processes, which Clark and Grush equate with AORs, can be described as 'off-line' or decoupled. The forward anticipatory aspect of neural processing is a constitutive part of the action itself, understood in diachronic, dynamical terms, rather than something decoupled from it. The anticipation of a future state or position (of the ball, or of the body grabbing the ball in the next second) requires ongoing reference or 'constant physical linkage' to one's current bodily state or position. To think of such processes as representational is to think that such anticipations are in some way detached or detachable from perceptual and proprioceptive input, which they clearly are not. Such processes may be one step ahead of real-world proprioceptive feedback—but they are also at the same time one step behind the previous moment of feedback, integrated with ongoing movement and perception (see section 5.3). Moreover, they necessarily and quickly dissipate as the agent continues to move. They are not stored as representations for later reuse, although if you catch enough balls your system becomes more proficiently attuned for further performance as well as for re-enacting the process *via* memory or imagination.

On some views, higher-order cognition is 'representation hungry' (Clark and Toribio 1994). On the enactivist account, however, to scale up to cognitive states such as imagining or remembering, the brain doesn't decouple or recreate a process that was representational to begin with; rather, the system (using the same motor control or forward control mechanism) enacts (or re-enacts) a process that is now coupled to a new cognitive action. In remembering, for example, there may be reactivation of perceptual neural processes that had been activated during the original experience. It has also been shown, using electromyography (EMG) that other non-neural bodily processes, e.g., subliminal tensing of muscles and facial expressions, may be (re)activated in cases of remembering, imagining, reflecting, etc. (e.g., Bakker, Boschker, and Chung 1996; Livesay and Samras 1998; Schmidt and Lee 1999).

Here, however, the line between accounts of AORs and the idea of enactive cognition gets blurred, and some may suspect that the difference is merely one of preferred vocabulary (see *Chapter 5*). Thus, defenders of AORs, like Michael Wheeler, give up the criterion of decoupleability as part of the concept of an AOR (2005, 219); and both Wheeler and Mark Rowlands suggest that AORs involve aspects of a system that includes brain, body, and environment. 'The vehicles of representation do not stop at the skin; they extend all the way out into the world' (Rowlands 2006, 224). When the concept of representation is weakened to this extent, however, one might suspect that what proponents of weak embodiment call AORs can be replaced with what enactivists call affordances. Affordances, however, in whatever way they are conceived, are not meant to be representation substitutes, as if they are standing in and doing the job that representations are said to do in more standard accounts. The notion of affordance is dynamically relational in a way that representations cannot be.

1.5 Beyond Predictive Coding

Since enactivist accounts reject standard computational and representationalist explanations they need to provide a different understanding of how the brain works. In this respect it will be productive to contrast enactivist conceptions to the recently advanced, and comprehensive alternative theory of brain function cast in terms of predictive coding (PC) or predictive processing (Clark 2016a; Hohwy 2013). PC has been an important trend in neuroscience that explains brain function in terms of Baysian inference (Friston 2005). According to Andy Clark (2013a, 181) it 'offers the best clue yet to the shape of a unified science of mind and action'. One might think that PC has already settled on the representationalist side since much of the PC literature assumes or adopts the representationalist vocabulary, along with the terminology of 'inference' and 'hypothesis' formation (e.g., Hohwy 2013). An alternative interpretation, however, emphasizing PC's recent focus on 'active inference', pushes some of the basic concepts of PC more towards the enactivist account (see e.g., Bruineberg, Kiverstein, and Rietveld 2016; Gallagher and Allen 2016).

On one reading of the PC approach the brain is pictured as having no direct access to the outside world; accordingly, it needs to represent that world by some internal model that it constructs by decoding sensory input (Hohwy 2013). The brain attempts to make sense out of sensory data 'within a cascade of cortical processing events in which higher-level systems attempt to predict the inputs to lower level ones on the basis of their own emerging models of the causal structure of the world (i.e. the signal source)' (Clark 2013a, 181). On this view, the brain makes probabilistic inferences (forms 'hypotheses') about the world and corrects those inferences to minimize prediction errors. This involves synaptic inhibitory processes based on empirical priors: that is, based on prior experience or on prior states of the system. Predictions are matched against ongoing sensory input. Mismatches generate prediction errors that are sent back up the line and the system adjusts dynamically back and forth until there is a relatively good fit. This is an efficient process since the only data that need to be sent up the line are the discrepancies (the surprises) from the predicted signal, and in the process of revising the prediction, the brain updates its model of the world and revises its priors (Rao and Ballard 1999).

Within this scheme an agent has two means by which to maintain its structural and functional integrity; either through the accurate internal prediction of hidden (external) causes, or by acting on the environment in ways that minimize sensory surprise. The latter, with the ensuing changes in action and perception, is known as active inference. On the one hand, emphasizing only the first type of operation, prediction-error minimization, PC remains strictly internalistic, and active inference only serves the central processes that do the real work. Accordingly, Hohwy

(2016) argues that PC understood in this narrow way is not consistent with EC approaches.

> PEM [prediction-error minimization] should make us resist conceptions of [a mind-world] relation on which the mind is in some fundamental way open or porous to the world, or on which it is in some strong sense embodied, extended or enactive. Instead, PEM reveals the mind to be inferentially secluded from the world, it seems to be more neurocentrically skull-bound than embodied or extended, and action itself is more an inferential process on sensory input than an enactive coupling with the environment. (Hohwy 2016, 259)

On the other hand, an emphasis on active inference leads to a recognition of the importance of embodiment and interaction, reflected in recent dynamical variants of predictive coding (Friston, Mattout, and Kilner 2011; Friston and Frith 2015; also see Kilner, Friston, and Frith 2007; Wolpert, Doya, and Kawato 2003). Thus, Clark, who sees PEM processes as closely tied to movement and action, argues that PC offers support for a more embodied and enactive theory of cognition (Clark 2013a; 2015; 2016b).

This suggests that we do not have to think that the outcome of the PC process is the creation of a representation in the brain—'a kind of internal model of the source of the signals: the world hidden behind the veil of perception' (Clark 2013a, 184).

> This means that 'inference', as it functions in the [PC] story, is not compelled to deliver internal states that bear richly reconstructive contents. It is not there to construct an inner realm able to stand in for the full richness of the external world. Instead, it may deliver efficient, low-cost strategies whose unfolding and success depend delicately and continuously upon the structure and ongoing contributions of the external realm itself as exploited by various forms of action and intervention. (Clark 2016a, 191)

Clark takes the problem of fielding a ball (mentioned in the previous section) as an example of active inference (2016a, 190, 247, 256ff.). On a PC account, he argues, it may be possible to move away from the vocabulary of representations, even AORs, in the same way that he wants to move away from the idea that PC depends on forming a 'hypothesis'. We can rather think of the brain as engaged in finding the distributed neural states 'that best *accommodate* (as I will now put it) the current sensory barrage' (192). In active inference, the brain does this, not by sitting back and formulating hypotheses, but *via* 'world-engaging action' (192), like running to catch a fly ball. Clark makes room for extra-neural, bodily factors (basic morphology, biomechanical dynamics, kinematics, environmental regularities), citing Gibson's ecological psychology and rejecting the classic 're-constructive' view (246–7). Even on Clark's PC account, however, there remains an ambiguity: on the one hand, active inference (as PC accounts have it) is in the service of generating information that is sent back to the brain for central processing. He cites Lungarella and Sporns (2005, 25) to indicate that the world-engaging action acts as a 'complement to neural information-processing'. This is also how Hohwy (2013) conceives it. On the other hand, Clark pushes towards a more enactive story: the problem solving is distributed across brain–body–environment, and this 'allows the productively lazy brain to do as little as possible while still solving (or rather, while the whole embodied, environmentally located system) solves the problem' (2016a, 248). The enactivist story is in parentheses.

Removing those parentheses, why should we not rather think of this process as a kind of ongoing dynamical adjustment in which the brain, *as part of and along with the larger organism*, settles into the right kind of attunement with the environment—an environment that is physical but also social and cultural (Gallagher et al. 2013)? Neural accommodation occurs in this larger system. Notions of adjustment and attunement can be cashed out in terms of physical states, or more precisely, physical dynamical processes that involve brain and body, autonomic and peripheral nervous systems, as well as affective and motoric changes.

This notion of enactive attunement is seemingly reflected in PC terms that emphasize two 'directions of fit'. The first involves updating predictions or adjusting priors on the basis of ongoing perceptual experience—the world-to-brain direction. The second involves acting on the world to directly shape or resample it in such a way as to test our prior expectations (active inference). In this respect, for PC models, perception may be conceived as an active process whereby I engage in the types of behaviors that are likely to produce sensory experiences that confirm or test my expectations; where, for example, active ballistic saccades do not merely passively orient to features but actively sample the bits of the world that fit my expectations or resolve uncertainty (Friston et al. 2012). Pushing PC towards more embodied, enactive, ecological accounts, Clark suggests that active inference takes the lead in the metaphorical 'circular causal dance'; one might even suggest, more radically, that in contrast to the 'subtly misleading' explanations of PC that makes the world-to-brain direction primary, all inference is really active inference (see, e.g., Clark 2016a, 250–1).[4] In contrast, however, on the enactivist model the dynamic adjustment/attunement process that encompasses the whole of the system is not a *testing* that serves better neural prediction; active inference is not 'inference' at all, it's a *doing*, an enactive adjustment, a worldly engagement (Bruineberg, Kiverstein, and Rietveld 2016; Gallagher and Allen 2016). The fielder is trying to catch the baseball; she is not performing tests or sampling the environment. The brain is not located in the center, conducting tests along the radiuses; it's on the circumference, one station amongst other stations involved in the loop that also navigates through the body and environment and forms the whole.

For example, we know that one's beliefs and values, as well as one's affective states and cultural perspectives (phenomena defined for the most part on the narrative scale), operating as priors, can shape the way that one quite literally sees the world (see section 6.4). How such cognitive and affective states and perspectives enter into (elementary scale) subpersonal processes can be explained in terms of PC models. With respect to affect, for example, Barrett and Bar's *affective prediction hypothesis* 'implies that responses signaling an object's salience, relevance or value do not occur as a separate step after the object is identified. Instead, affective responses support vision from the very moment that visual stimulation begins' (Barrett and Bar 2009, 1325). Along with the earliest visual processing, the medial orbital frontal cortex is activated, initiating a train of muscular and hormonal changes throughout the body, and generating 'interoceptive sensations' from organs, muscles, and joints associated with prior experience, which integrates with current exteroceptive sensory information. This is the organism's response which contributes to shaping subsequent actions. Accordingly, as part of the perception of the environment, we undergo certain bodily affective changes that accompany this integrated processing. In other words, before we fully recognize an object or other person, for what it or he or

[4] As we see in Clark (2016a, 251), however, this simply leads back to the notion of pragmatic representations, serving 'epistemic functions, sampling the world in ways designed to test our hypotheses and to yield better information for the control of action itself'.

she is, our bodies are already configured into overall peripheral and autonomic patterns shaped by prior associations. In terms of the PC model used by Barrett and Bar, priors that include affect are not just in the brain but involve whole body adjustments—what Freund et al. (2016, 1860) call 'anatomically informed priors'.

On the enactivist view, brains play an important part in the ongoing dynamical attunement of organism to environment. Social interaction, for example, involves the integration of brain processes into a complex mix of transactions that involve moving, gesturing, and engaging with the expressive bodies of others—bodies that incorporate artifacts, tools, and technologies that are situated in various physical environments, and defined by diverse social roles and institutional practices. Brains participate in a system, along with all these other factors, and it would work differently, because the priors and surprises in the system would be different, if these other factors were different. If, as Clark (2013a, 189) suggests, 'humans act as rational Bayesian estimators, in perception and in action, across a wide variety of domains', which means that they take into account the uncertainty in their own sensory and motor signals, this is due to the fact that brains evolve to function the way they do because they evolve with the body they are part of, and in environments that are coupled in specific ways to those bodies.

Perception, on the enactivist view, involves transactions in the complete (neural plus extra-neural) system. Enactivists emphasize sensory–motor contingencies (e.g., Noë 2004), bodily affect (see section 8.1), as well as the role that intersubjective interaction plays in shaping perception (section 8.2). Perception thus, rather than the result of narrow inferential or simulative processes, involves complex, dynamical processes at a subpersonal, sensory–motor level (in the elementary timescale)—but these processes are part of an enactive, dynamical engagement or response of the whole organism (in the integrative and narrative timescales), living in and materially engaging with structured environments. As Clark suggests, taking into consideration more embodied and embedded practices, we can use 'a variety of tricks, tools, notations, practices, and media [to] structure our physical and social worlds so as to make them friendlier for brains like ours'—thereby stacking the deck (designing our surrounding environments) to minimize prediction errors (2013a, 195). Such redesigns, however, reflect a *metaplasticity* that goes both ways—changing not only the brain, but also physical, social, and cultural environments (Malafouris 2013). We can intervene at any point on the self-organizing circle of brain–body–environment; that intervention will incur (sometimes friendly, sometimes not so friendly) adjustments to the whole.

Changes or adjustments to neural processing will accompany any changes in these other worldly factors, not because the brain represents such changes and responds to them in central command mode, but because the brain is part of the larger embodied system that is coping with its changing environment. Just as the hand adjusts to the shape of the object to be grasped, so the brain adjusts to the circumstances of organism–environment. And just as it is not clear that we gain anything by saying that the shape of the grasp represents the object to be grasped (cf. Rowlands 2006), it's not clear that we gain anything in saying that brain activations represent the world.

With respect to PC models, enactivist views that emphasize a more holistic system of brain–body–environment would clearly favor a move away from internalist and intellectualist vocabularies (and conceptions) of 'hypothesis', 'inference', and 'representation' in favor of more embodied terms like 'adjustment', 'attunement', and 'affordance'. Such terms are not simply substitutes for the PC terms; they change the way that we think of the brain's engagement.

1.6 Enactivism as a Philosophy of Nature

Enactivist EC approaches present a challenge for science. Enactivists, by focusing on not just the brain, not just the environment, not just behavior, but on the rich dynamics of brain–body–environment, offer a holistic conception of cognition. To put it succinctly, however, it is difficult to operationalize holism. Neither experimental control nor the division of labor in science allows for all factors to be taken into consideration at once. Nor is it clear that there could be one single critical experiment that might decide the issue between the representationalist and the enactivist. On the one hand, enactivism makes empirical claims, for example, about the work of sensorimotor contingencies, and in this respect it resembles a research program that can suggest new experiments and new ways of interpreting data. On the other hand, its emphasis on holism presents problems for empirical investigations. To be clear, nothing prevents science from doing its experiments, controlling for variables, and building up explanations one experiment at a time. Yet, each science tends to develop its theories based on its own particular assumptions, in its own vocabulary, and often in isolation from the insights of other sciences. Triangulation doesn't always work just because of conflicting assumptions, vocabularies, and interpretations. One can encounter what I've elsewhere called the 'clunky robot' problem (Gallagher et al. 2015, 74). That is, just as one can design a robot by assigning teams to construct different modules, which turn out to work well as individual modules, it may happen that when the modules are brought together, they don't play well together. No one has considered the relational aspects of how one module will dynamically connect with another in a complex system, and the result is a clunky machine-like behavior. The same problem can be found in theory construction. Scientific experiments, designed within the framework of their own particular paradigm, often study the pieces of a system but don't always consider how the dynamical relations among those pieces work, and don't always have the vocabulary to address those relations. Even working in an interdisciplinary way we often find ourselves building a clunky theory where insights from different disciplines don't integrate well.

This motivates serious consideration of the idea, first suggested by Cecilia Heyes[5] (drawing on a distinction proposed by Godfrey-Smith [2001]), that enactivism may be better thought as a philosophy of nature than a scientific research agenda. Godfrey-Smith, discussing developmental systems theory, distinguishes between a 'scientific research programme' and a 'philosophy of nature'. As he makes clear, a philosophy of nature is a different kind of intellectual project from science, and although science may be its critical object, the two enterprises do not have to share the same vocabulary. A philosophy of nature 'can use its own categories and concepts, concepts developed for the task of describing the world as accurately as possible when a range of scientific descriptions are to be taken into account, and when a philosophical concern with the underlying structure of theories is appropriate' (Godfrey-Smith 2001, 284). A philosophy of nature takes seriously the results of science, and its claims remain consistent with them, but it can reframe those results to integrate them with results from many sciences. An exclusive

[5] Here I'm pursuing a suggestion made by Cecilia Heyes, commenting on a paper I presented at the Ernst Strüngmann Forum, Frankfurt, in October 2014.

focus on cognitive neuroscience as *the* science of cognition, for example, would be entirely unjustified on this view. Moreover, the requirements of such a reframing may indeed call for a vocabulary that is different from one that serves the needs of any particular science. Although to work out a philosophy of nature is not to do science, it can still offer clarifications relevant to doing science, and it can inform empirical investigations. In this sense, a philosophy of nature is neither natural philosophy (in the traditional sense) nor the kind of naturalistic philosophy that is necessarily continuous with science. It offers critical distance and practical suggestions at the same time. In some cases it may make doing science more difficult.

That enactivism is a philosophy of nature can be seen in the fact that from the very start enactivism involved not only a rethinking of the nature of mind and brain, but also a rethinking of the concept of nature itself (see Di Paolo 2005; Thompson 2007, 78ff.). If enactivism is a form of naturalism, it does not endorse the mechanistic definition of nature often presupposed by science, but contends that nature cannot be understood apart from the cognitive capacity that we have to investigate it. As Cecilia Heyes suggested in her comments, in the context of a philosophy of nature meant to offer an encompassing view, holism is a strength rather than a practical complication.

Enactivism, as 'a non-reductionist yet scientifically engaged philosophy of nature' (Di Paolo, Buhrmann, and Barandiaran 2017, 253), may still motivate experimental science in very specific ways. Even if in some cases it is difficult to apply a holistic view to a given question, in many cases there may not be any special complication in designing experiments that can test enactivist ideas. For example, one can set all factors to work and then test the system to see what happens when we intervene to knock out one of those factors. Moreover, one need not include absolutely everything in every case when dealing with a particular concrete question, although in the end it may be easier to include than to ignore a factor that is crucial. For example, including embodied interactions in explanations of social cognition might actually be more parsimonious if keeping them out of the picture requires the elaboration of more convoluted explanations in terms of theory or simulation mechanisms (De Jaegher, Di Paolo, and Gallagher 2010). Although in this, and other cases, much will depend on circumstances like the availability of the right lab technology, the whole may sometimes lead to simpler explanations. In short, even if enactivism were to be considered a philosophy of nature, it wouldn't be right to conclude that it cannot offer concrete hypotheses or raise novel scientific questions.

The following chapters are meant to be contributions towards formulating an enactivist philosophy of mind, as part of a larger philosophy of nature. *Chapters* 2 and *3* provide a broad background and situate enactivism in contemporary and historical contexts. *Chapter 2* reviews a number of contemporary approaches to embodied cognition in order to clearly distinguish the enactivist version. *Chapter 3* explores a largely ignored background to enactivism in the American pragmatist tradition.

Chapter 4 outlines a theory of enactive intentionality that capitalizes on both the phenomenological and the pragmatist roots of enactivism. Looking at debates about intentionality in neo-behaviorist and neo-pragmatist approaches, I argue that if we frame the notion of intentionality correctly, then the opposition between enactivism and extended mind approaches can be resolved. Specfically, the enactivist approach can borrow from neo-pragmatism and develop an embodied and extended mind account of non-derived intentionality that is immune to objections from the standard internalist theories. On this view, however, intentionality is not equivalent to representation. This leads, then, to an enactivist critique of representationalism in *Chapter 5*. I argue that on standard definitions of representation in philosophy of mind, perception–action processes are non-representational—in contrast to various versions of action-oriented representations defended by Clark, Wheeler, Rowlands, and others.

This motivates the question explored in *Chapters* 5 and *6*: How does the brain work if it is not forming internal mental representations of the world? Here I come back to some of the issues that I touched on above. In contrast to the standard conception of the brain making inferences (as found in classic computational accounts and the more internalist predictive processing accounts) the enactivist view is that the brain, as part of the body–environment system (not only regulating body, but regulated by the body and its affective processes) is, as Jesse Prinz puts it, 'set up to be set off' (2004, 55) by prior experience and plastic changes. The brain works as an integral part of the organism which, as a whole, responds dynamically to environmental changes. It's not clear that this is equivalent to the notion of 'active inference' in predictive processing accounts, but from the enactivist perspective, it may be the best way to think of how the brain works.

Chapter 7 argues that the attempt to locate free will within elementary timescales or in neurophysiological processes just prior to action is wrong-headed. I outline an enactivist response to recent debates around the notion that free will is an illusion. I argue that Libet experiments, which show how neural activations in the elementary timescale of milliseconds prior to action anticipate a voluntary movement, even before the agent decides to move, are about motor control processes (where motor intentions are formed)—not about free will, which is best understood on a narrative timescale and which involves larger processes of distal intentions and action in social contexts.

Chapter 8 argues that an enactivist conception of embodied cognition involves more than sensory–motor contingencies, and more than a critique of representationalism. Here the importance of both affect and intersubjectivity is emphasized. *Chapter 9* continues on this theme by returning to an essay by Erwin Straus (1966), written within the tradition of phenomenological anthropology and focused on the human upright posture. I extend and update this analysis as a way to flesh out a fuller conception of embodiment and its relation to rationality. Finally, in Chapter 10, I address what is variously known as the problem of 'scaling up' or the question of 'higher-order' cognition (although I reject these labels). Can an enactivist approach explain cognitive processes involved in reflective thought, deliberation, memory, imagination, and so forth? To gain some traction on this issue I frame my answer in terms of a recent debate between Dreyfus and McDowell concerning the nature of the mind. The enactivist approach can split the difference and establish a space somewhere between the positions of Dreyfus (who emphasizes embodied skills and coping) and McDowell (who focuses on concepts and the space of reasons). I argue that reflection and conceptually rich cognitive operations (e.g., imagining and doing mathematics) involve a specialized affordance space but are continuous with embodied coping, and share the same structure.