Chapter 1

1 Merleau-Ponty for Cognitive Science

Preamble: A Remarkable Convergence

My attention was first drawn to the relevance of Merleau-Ponty for cognitive science through reading Andy Clark’s *Being There* (1997). Clark indicates that Merleau-Ponty’s depiction of everyday intelligent activities in terms of “the playing out of whole organism-body-world synergies” is close in spirit and execution to Clark’s own project. Moreover, Clark is particularly fond of Merleau-Ponty’s stress on the activity of the organism in perception, and he coins the term ‘continuous reciprocal causation’ to describe this interaction between perceiver and perceived. Clark thus sees Merleau-Ponty as presenting a kind of ‘free-form interactive dance’ view of perception, which he endorses on the grounds of its compatibility with “recent work in the computational foundations of animate vision.” (Clark 1997, p172) I want to take a moment to look at the compatibility that Clark refers to – with the findings of researchers in animate vision – which I think is surprisingly compelling. I offer this as a kind of intuition pump for appreciating the affinity between Merleau-Ponty’s writing and certain developments within cognitive science. In the main part of this chapter, I’ll go through the ways in which this affinity has been cashed out, supported, and formalised.

Merleau-Ponty and animate vision

Animate vision is a new engineering paradigm of robot vision which seeks to replace the traditional account of a system which constructs and updates an accurate 3-D model of the viewer’s surroundings. This traditional approach is often unfavourably referred to as passive or pure vision and is generally criticised on the grounds that it makes infeasible computational demands on the observer robot, especially when it comes to tasks involving a moving object, a moving observer, or, in the worst case, both. The advantage of passive vision is supposed to be that it maintains a general representation of the visual field that can be employed in any behavioural task. However, this generality is precisely what has come under attack on computability grounds, because the generality of the information received must be compensated for by the specificity and complexity of the algorithm that translates that information into the form of a behavioural response. The idea behind animate vision, then, is to provide only
that visual information required by a particular task, and to employ other so-called "visual behaviours" that modify the visual system and its relation to the object, in order to further simplify the computational task required. I am going to use Dana Ballard’s presentation of animate vision for my purposes here. His account is generally endorsed, and is that employed by Andy Clark in *Being There*.

Animate vision differentiates itself from passive vision, as I mentioned, by its decomposition of vision into visual behaviours. These are behaviours that exploit the observer robot's motilities, i.e. its capacities to alter its perspective through movement of its cameras. Ballard defends this suggestion by pointing to an experiment that suggests that there is no routine algorithm for creating a complete representation of people and objects in a scene. In this experiment saccadic eye movements, i.e. the paths traced by a person's fovea or region of high optical resolution, were studied in the context of three different tasks. Subjects were shown an image and asked to (a) give the ages of the people in the scene, (b) surmise what the family had been doing before the arrival of the "unexpected visitor"; and (c) remember the position of the people and the objects in the room. The results showed that highly specific saccades were employed for each task.

Passive vision theories predict a general-purpose saccade that, to a first approximation, locates and identifies objects and people in a scene. This saccade may be supplemented by additional detailed searches, but should nevertheless be included in every visual task. To the contrary, Ballard concludes that "since the eye movement traces show a specialized signature for this task [i.e. the third task of localisation and identification] as well, it is not done routinely." Rather, "the visual system is used to subserve problem-solving behaviors and such behaviors often do not require an accurate model of the world in the traditional sense of remembering positions of people and objects in a room." (Ballard 1991, 59)

Animate vision is 'animate' in two senses of the word, both of which stem from its decomposition of vision into visual behaviours. Visual behaviours are unlike traditional accounts of vision in that they aim to provide an ongoing representation of the world that is not generally useful, nor complete, but simply suffices for the particular task at hand. That is, we see in order to move. Animate vision is about providing visual content which is timely and relevant to a robot that operates in real time, with several forms of motility. It is therefore animate in the sense that it is geared to the motilities of its subject, which is assumed to be embodied.
The assumption of embodiment straight away brings animate vision close to Merleau-Ponty's phenomenology of the living body. But the link with motility is also pertinent, since Merleau-Ponty encourages a conception of spatiality that is intimately bound up with motility.

Even if subsequently, thought and the perception of space are freed from motility and spatial being, for us to be able to conceive space, it is in the first place necessary that we should have been thrust into it by our body, and that it should have provided us with the first model of those transposition, equivalents and identifications which make space into an objective system and allow our experience to be one of objects. (PhP, 142)

Spatiality and motility are therefore correlative. If they were not, certain behaviours would be impossible. For example, the immediate substitution I perform for limbs that are injured or otherwise unavailable would be impossible if the object of my action did not stand in some functional relation to my body as a whole, and in such a way that it solicits the appropriate movements from the substituting limb.

Admittedly, the animate vision paradigm is less radical in its account. Spatiality and motility are not strictly correlative, but the greater the degree of correlation, the more efficient the vision dependent performance.

Depth Perception, Habitation and Non-Synthetic Unity

The second way in which animate vision is animate is in its recognition that, as much as we need to see in order to move, we also need to move in order to see. The animate vision paradigm responds to difficult problems such as depth perception by searching for behaviours which will minimise that computational load. It turns out that a small, expected shift in perspective, a purposeful shake of the head, allows that load to be significantly lightened, but with the proviso that the robot's gaze must remain fixed. (The gaze is, to my ears, a highly suggestive term for the focal point and plane of vision in real space.) There are two advantages to the employment of this visual behaviour: depth is discerned in relative terms within a coordinate frame fixed at the focal point, and the robot can roughly discern object boundaries via discontinuities in relative depth.

There are two ways in which Merleau-Ponty's phenomenology seems to anticipate animate depth perception. He specifies that perception in general involves centring one's perceptual field on the object of perception - "to look at an
object is to inhabit it, and from this habitation to grasp all things in terms of the aspect which they present to it. ” (PhP, 68) - and this seems to reflect the fixation of the coordinate frame at the focal point in animate vision, and the perception of depth in relative terms. In a later, more extensive discussion of depth, he points out that capacity to see depth and the capacity to retain subjective unity through a change of perspective (recall the shake of the head) are one and the same thing. ”[The] contraction into one perceptual act of a whole possible process constitute[s] the originality of depth.” (PhP, 264)

Ballard’s approach to depth perception requires embodiment to exploit perspective. For Merleau-Ponty, depth is regarded as ”the most ‘existential’ of all dimensions, because...it is not impressed upon the object itself, it quite clearly belongs to the perspective and not to things. Therefore, it cannot either be extracted from, or even put into that perspective by consciousness.” (PhP, 256) "One cannot, therefore, speak of a synthesis of depth...since depth does not posit the multiplicity of perspective appearances to be made explicit by analysis, but sees that multiplicity only against the background of the stable thing." (PhP, 265; my emphasis)

Merleau-Ponty’s point has more than a technological relevance: we learn that depth implies at least potential movement, and a unity through that movement which is not an inferred, synthesised unity produced after the fact, but a unity which is cotemporal with perception. In technological terms however, it encourages us to think of animate vision’s approach to depth perception as taking the burden off thought (symbolic processing), and placing it onto sight. The fixation of the gaze provides an exocentric stable point, which, in combination with movement through a continuity of perspectives, allows the robot to ’see’ a vectorised map of depth.

Spatial Memory and the Thing-Seen-From-Everywhere

A further point can be made regarding the relationship between the gaze and its objects. Merleau-Ponty is predominantly concerned with how we come to have an experience of an object in its objectivity, that is, as something that transcends our perspective on it. He explains the transcendence or permanence of an object is ’guaranteed’ by its situation among other objects as spectators bearing witness to its character and existence.
I can therefore see an object in so far as objects form a system or a world, and in so far as each one treats the others round it as spectators of its hidden aspects and as guarantee of the permanence of those aspects. (PhP, 68)

But in so far as I see those things too [objects in the periphery], they remain abodes open to my gaze [i.e. my gaze may return to them], and, being potentially lodged in them, I already perceive from various angles the central object of my present vision. Thus every object is the mirror of all others. When I look at the lamp on my table, I attribute to it not only the qualities visible from where I am, but also those which the chimney, the walls, the table can 'see' … (PhP, 68)

For Ballard, the establishment of an observer-independent memory involves a similar process of situation, though here that process is restricted to the establishment of mathematical relationships between coordinate frames. Ballard builds his robots to use a another kind of geometry in the construction of spatial memory, which he calls object-centred coordinate frames. This is a bad terminological choice, since fixation-point frames are also object-centred. However, unlike the fixation-point frame, the object-centred frames that Ballard refers to are orientated independently of the gaze of the observer. Ballard does not elaborate on how these object-centred frames of reference are orientated, but one can imagine exploiting natural symmetries within the area of interest for this purpose. Really, it's not important how they're orientated, what is important is that they do not need to be updated to compensate for the movements of the observer robot. What is stored in memory are matrices which translate between pairs of object-centred coordinates. Since the frames are observer-independent so are the matrices of translation. The information that these matrices carry is effectively the displacement of one object from another and the difference in orientation between them. The translation matrix amounts to an extrapolated record of an earlier object's image of the present one. We can therefore interpret Ballard's paradigm of spatial memory as storing the location and orientation of the present object from the point of view of its neighbours.

Recall also that Ballard's robots do not seek a complete representation of the world. The advantage of the creation of this matrix is that it only stores the information needed to return one's gaze to the earlier object, without needing to continually correct that information in the light of future movements. The form of spatial memory they employ is expressly for the purpose of reestablishing one's gaze on a former object of interest by retracing one's steps; 'steps' being represented mathematically by the transformation matrices. It is true that Merleau-Ponty distinguishes perception and memory per se, while Ballard only
distinguishes them by their geometries, but like Ballard, Merleau-Ponty does so because he reserves memory expressly for the purpose of reliving a past experience.

To remember is not to bring into the focus of consciousness a self-subsistent picture of the past; it is to thrust deeply into the horizon of the past and take apart step by step the interlocked perspectives until the experiences which it epitomizes are as if relived in their temporal setting. To perceive is not to remember. (PhP, 22)

I wanted to start with this comparison, because I think it offers a compelling prima facie case for asking after the significance of this apparent convergence of thought, a convergence that has been taken up in a number of different ways.

Overview of the chapter

I've divided the subsequent literature review into three sections, to reflect what I regard as the three distinct dimensions of that reception. Firstly, what I call the psychological aspect continues along the lines of this preamble, illustrating those points of descriptive convergence that Merleau-Ponty's readers point to in support of his relevance to contemporary cognitive science. I refer to this as the psychological aspect, though it might just as well have been called the "phenomenological" aspect (in the minimal, analytic sense of the term) or simply the "descriptive" aspect, because I mean to examine the ways in which Merleau-Ponty's texts are taken to offer precursors to contemporary psychological accounts of intelligence and mentality. This first sub-section covers the way connectionists and enactivists embrace Merleau-Ponty's descriptions. As the reader will see, connectionists tend to take these descriptions more literally, while enactivists tend to treat them as representative of a general approach.

The second, philosophical, dimension of the reception might be considered to be only indirectly related to cognitive science, especially by those inspired by Rodney Brooks' professed philosophical indifference in "Intelligence without Representation" (Brooks 1991). Nonetheless, it is crucially important. The surge in interest in what phenomenology has to offer cognitive science in the last dozen years has its origins further back, in Hubert Dreyfus' philosophical critique of Good Old-Fashioned AI (GOFAI) (Dreyfus 1979; Dreyfus 1992). Dreyfus drew mostly on the work of Heidegger and Wittgenstein to bring into question the

---

3 See the brief note in §7.5 "It isn't German philosophy"
representationalism inherent in this research program. His criticism focused on questioning the ability of that research program to encode, in the form of explicit representations and rules for their manipulation, the kind of background or commonsense know-how that human behaviour depends upon, and to coordinate the retrieval of only the salient know-how at the appropriate time. I’m not going to recapitulate Dreyfus’ well-known arguments here, which relied very little on Merleau-Ponty. Instead, I’m going to discuss two recent attempts – by Charles Taylor and Taylor Carman – to critique contemporary epistemology and philosophy of mind from a Merleau-Pontyan perspective.

When Dreyfus first presented his critique of GOFAI, there was no real alternative to computationalism in either AI or cognitive psychology. As a result, his critiques were construed as a call for the development of such alternatives. Today, with the emergence of connectionism and enactivism, such a call is no longer required. Consequently, the critiques presented on behalf of Merleau-Ponty by Taylor and Carman play a different role. They implicitly support the convergence of descriptions discussed in the preface, by suggesting that Merleau-Ponty’s descriptions and those of researchers in non-cognitivist cognitive science share the same philosophical motivations: the rejection of objectivism, understood as a mix of representationalism and physical reductionism.

This leaves the third, and least mature, dimension of the reception; namely, the attempt to formalise relations between cognitive science and phenomenology, thereby securing the de facto convergence. Such an attempt was begun in earnest by the late Francisco Varela, as part of the Naturalizing Phenomenology project. Varela’s early conception of the relationship in the Embodied Mind focused on Merleau-Ponty, but his recent work focused almost entirely on Husserl, though, as I suggest, Merleau-Ponty is described as a source of support for the later project. Dreyfus doesn’t offer a systematic conception. I attempt to articulate his working hypothesis and criticise the capacity of this conception to stand up to scrutiny. I also consider some developments of Varela’s and Dreyfus’ views in the work of Shaun Gallagher and Ron McClamrock, respectively.
1.1 Psychological Aspects

Connectionist approaches

A) Neurophysiological basis for expert coping: Dreyfus

Dreyfus' prime target is representationalism – the view that intelligence consists in the acquisition and manipulation of internal symbols that stand for salient features of the environment. Merleau-Ponty's relevance is owed to the fact that he "holds that the most basic form of intelligent behavior, skillful coping, can and must be understood without recourse to any type of representation." (Dreyfus 2005,129)

The grand-daddy of reading Merleau-Ponty for cognitive science, Dreyfus grants Merleau-Ponty both negative (i.e. critical) and positive (i.e. descriptive) significance. I'm going to set aside the way Dreyfus appropriates Merleau-Ponty's critiques of intellectualism and empiricism, for the sake of focusing on the convergence Dreyfus presents between Merleau-Ponty's descriptions and concepts within connectionist cognitive science.

Intentional arcs and feedback loops

In the *Phenomenology of Perception*, Merleau-Ponty remarks that:

> the life of consciousness – cognitive life, the life of desire or perceptual life – is subtended by an 'intentional arc,' which projects round about us our past, our future, our human setting. (PhP, 136)

Dreyfus takes this in a few slightly different ways, which he doesn't distinguish. On the one hand, he treats the intentional arc as a relationship between past and present, in which "past experience is projected back into the perceptual world of the learner and shows up as affordances or sollicitations to further action." (Dreyfus 2005,132) This notion of affordances has quite strong Gibsonian overtones, and lends itself quite naturally to the notion of a dialectical relationship between subject and world. Dreyfus finds further support for this in the *Phenomenology*:

> As Merleau-Ponty puts it, a "person's projects polarize the world, bringing magically to view a host of signs which guide
action, as notices in a museum guide the visitor” [PhP, 112]
(Dreyfus 2005, 132)4

On its own, however, this interpretation does not readily lend itself to translation into the language of cognitive science. The key move is made when Dreyfus describes the intentional arc as a kind of “feedback structure,” a “feedback loop between the learner and the perceptual world” (Dreyfus 2005, 132; cf. Dreyfus 2002, 373).

The notion of a feedback loop comes from the engineering of control systems, and is immediately intelligible within cognitive science, particularly dynamical (i.e. connectionist or enactivist) cognitive science. Dreyfus supports his translation of the concept into the language of control systems through an appeal to the Structure of Behavior:

Merleau-Ponty refers to this feedback structure as a dialectical or circular relation of milieu and action: "the relations between the organism and its milieu are not relations of linear causality but of circular causality." [StrB, 15]
(Dreyfus 2005, 132)

The translation from "circular causality" to "feedback loop" is certainly a lot less problematic than it would be from "intentional arc," which is a very opaque phrase. However, it isn’t clear that Merleau-Ponty would himself endorse the equation of the meaning of "intentional arc" and "circular causality" given that, prior to introducing the former phrase in the Phenomenology, he writes that "reciprocal action’ is as yet only a compromise with causal thought, and a contradictory principle". (PhP, 127) Does Merleau-Ponty mean to avoid any implicit appeal to causality, circular or otherwise, or does he merely mean to indicate that, at the time, there was no way to make sense of such a reciprocal action between organism and milieu?

Dreyfus is happy to suppose that the second is the correct interpretation, for he argues that the absence of any satisfactory model on which to understand the reciprocal action characteristic of skilled behaviour forced Merleau-Ponty, unconvincingly, to claim that the perception-action loop was "magical".

---

4 I use Dreyfus own citation, because he modifies the translation in order to avoid reference to the distinction between normal and pathological subjects in Colin Smith’s version and in the original text. As will become apparent later on, I consider this distinction very important to understanding the phenomenological context of Merleau-Ponty’s descriptions.
Dreyfus makes the convergence explicit by introducing the reader to neural networks, which are capable of learning new skills purely through feedback loops.

Fortunately, … there are now models of what might be going on in the brain of an active perceiver forming an intentional arc that do not introduce brain representations. Such models are called simulated neural networks. (Dreyfus 2005,132-3)

Simulated neural networks consist of a layer of input "nodes" (i.e. simulated neurons), a layer of output nodes, and any number of "hidden" nodes between them. The simulated synaptic connections between nodes are weighted, and for any given input, the entire network (usually⁵) comes to an equilibrium in which the activation or output level of each node is determined. This includes, most importantly, the output nodes.

Once the weightings on each connection are allowed to vary freely, the net can be trained. Consider the case of a net being trained to correctly identify handwritten characters. This can be done, roughly speaking, by holding steady the output nodes to reflect the value of the particular character being trained, and exposing the input to various handwritten examples of that character. The combination of input and output states forces the weightings on the hidden nodes to adjust. The net has been successfully trained when, of its own accord, it comes to the appropriate output state given any instance of a handwritten character.

Dreyfus points out that such a network can "produce outputs based on past experience without the net needing to represent its past experience as cases or rules for determining further actions" (Dreyfus 2005,133), and he concludes that:

The neural-net model thus suggests a nonrepresentational, and yet non-magical, brain basis of the intentional arc.
(Dreyfus 2005,134)

The body as a generalisation constraint

Dreyfus presents Merleau-Ponty’s turn to the body as a solution to a problem of generalization. That is, the problem of understanding "how … human beings learn to generalize like other human beings" (Dreyfus 2005, 135). He quotes the Phenomenology:

⁵ Taking the simplest case, and ignoring complex periodic or chaotic effects.
Although our body does not impose definite instincts upon us from birth, as it does upon animals, it does at least give to our life the form of generality. (PhP, 146)

The quote can be understood in two ways, depending on how one reads "the form of generality". It can be read as saying that our body imposes a certain general form upon our lives, although not the specific instincts we associate with animals. This is the reading Dreyfus prefers. But alternatively, it can be read as saying that our body simply ensures that our lives have a certain generality. This reading seems better to capture the strangeness of the phrase, "form of generality," but like intentional arc, has a sense that is difficult to clarify.

Dreyfus takes Merleau-Ponty to be suggesting that having a body provides a constraint on the ways in which we generalize. This differentiates us from an abstract, simulated neural network, which is unlikely to generalize the way we do.

When a neural network has a large number of hidden nodes, it can be set up to respond to not only the association of input and output states, but also the order and frequency of those associations. This kind of neural net could:

be set up to learn by itself, that is, if its connection strengths were arranged so as to adjust themselves on the basis of the input-output pairs that the net encountered in the world, then the order and frequency of the inputs would depend on the interaction of the structure of the embodied network and the structure of the world. (Dreyfus 2005, 135)

Maximum grip, attractor theory and reinforcement learning

As experts getting around in the world, we are all constantly drawn to what Merleau-Ponty thinks of as a maximal grip on our situation. (Dreyfus 2005, 137)

This notion of a maximal or maximum grip is derived from a remark Merleau-Ponty makes about finding an optimal distance from which to view a painting (PhP, 302). Merleau-Ponty suggests that I "tend towards a maximum of visibility" (PhP, 302), and I recognize having achieved such a state when:

my perception presents me with a spectacle as varied and as clearly articulated as possible, and when my motor intentions, as they unfold, receive the responses they expect from the world. (PhP, 250)

The important feature of this notion for Dreyfus is that it suggests that "skillful coping does not require any representation of a goal. It can be purposive without
the agent entertaining a purpose." (Dreyfus 2005, 138) Really, the italics in this sentence are misleading. The emphasis should really be on the lack of any representation of a goal, and of any need for the agent to entertain a purpose. To talk about an optimal situation, and the system's tending toward it as "a normative sensitivity", as Dreyfus does (Dreyfus 2005, 141), is to talk about one's behaviour as having a goal or a purpose. But what is clear is that Dreyfus does not want this to imply that this goal need ever be the content of a representation, or more accurately, that we ever need to explain this tendency toward it as the result of comparing a representation of the system's current state to a representation of the optimal state.

In an earlier paper (Dreyfus 2002), Dreyfus turns to Walter Freeman's attractor theory to explain how there can be a satisfactory result (maximal grip) that the learner tends towards, which nonetheless does not involve any sort of representation. On this theory, neural activity in the brain can be regarded as a dynamical system, and taken together, its possible states can be thought of as marking out a landscape of peaks and troughs of neural activity. It is then suggested that the brain tends toward the troughs, or "basins of attraction", so that if the current brain state is within range of such a basin, it will evolve in such a way that it approaches this minimum energy state; a little like a ball rolling down a hill.

The notion of maximal grip is then explained in terms of the tendency of the system to evolve toward "basins of attraction".

The system thus directs the player to make those movements that result in the brain approaching the lowest accessible point in its current energy landscape, without the player needing to represent where that lowest point is or how to get there, any more than a river needs to represent as its goal the lowest point in the landscape in order to find the optimal path down the hill. (2002, 383)

In the more recent version, Dreyfus implicitly recognizes that this only accounts for expert behaviour, and does not address the question of how it is "possible to learn to make the right decision at an early stage without immediate [and here he means explicit] feedback as to whether that decision increases or decreases the

---

*There's an implicit tension in Dreyfus' account between his 'phenomenological' characterisation of the 'intentional arc' as a tight connection between agent and world, and the connectionist, neurophysiological model he appeals to as providing an appropriate explanation of the phenomenon. The restriction to the neurophysiological threatens to undermine the embeddedness that Dreyfus seems to want to emphasize. This corresponds roughly to the enactivist criticism of connectionism, as we will see below.*
Dreyfus explains this in terms of the development of a second feedback loop, which "represents the prospect of future reward – all without needing to represent its goal." (Dreyfus 2005, 140) The second loop is still "myopic," it corresponds not to any objective measure of the system/organism's distance from the optimal, but rather its sense of that distance. Dreyfus argues that this model reflects Merleau-Ponty's description of learning a skill in *Structure*:

> [The organism] builds up aptitudes, that is, the general power of responding to situations of a certain type by means of varied reactions which have nothing in common but the meaning... Situation and reaction... are two moments of a circular process. (StrB, 130)

As we can see, in a similar move, Merleau-Ponty's talk of a "circular process" has been translated into the control system concept of a feedback loop.

**B) Neural network modelling: Borrett, Kelly and Kwan**

Borrett, Kelly and Kwan set themselves the task of showing how "phenomenological principles can be introduced into [the] physical models" upon which neurophysiological accounts of human motility are based. (Borrett, Kelly, and Kwan 2000, 226) On their view, phenomenology offers direct descriptions of the experience of perceiving and acting in the world, "without metaphysical presuppositions inherited from psychological, scientific, historical, sociological, or other theoretical frameworks." (Borrett, Kelly, and Kwan 2000, 213) These descriptions are thus taken as data to be explained by the brain sciences. While their focus is on the kinds of models that fit Merleau-Ponty's descriptions, the authors claim to subscribe to a dual-constraint model, in which neurophysiological plausibility is as much of a constraint as phenomenological accuracy. However, these both seem to be taken as constraints on the models – there's little sense that neurophysiological plausibility is to be understood as a constraint on phenomenological description.

The approach outlined in Borrett, Kelly and Kwan (2000) is distinctly connectionist in flavour. The master idea is a variety of anti-representationalism, i.e. that perception and action do not need to be mediated by a governing representation, combined with the belief that neural networks demonstrate the neurological plausibility of non-representationalist approaches. The authors’ present their neural network models as alternatives to those accounts of behaviour that presuppose the need for an error function which would compare
Chapter 1

the current state of the system to a general representation of the objective of the performance, e.g. the canonical view of an object, or the ideal hard/arm position for gripping a cup.

The feature of neural networks that the authors emphasize is that the changes in the network as it moves toward the target state cannot be described as continual, improving approximations to the latter. Its trajectory through its state space is not a function of its difference from the target state, but is determined entirely by its initial conditions. The target state is a kind of "attractor," a nonequilibrium stable point, toward which the system simply evolves. This evolution of the system toward an attractor is called "network relaxation."

Action: grasping vs pointing

This dependence only on initial conditions becomes the first point of comparison with Merleau-Ponty. The authors take up Merleau-Ponty's distinction between pointing and grasping behaviours, and suggest that neural networks provide better models of grasping than representationalist/cognitivist alternatives. In the authors' opinion the latter tend to treat grasping behaviours as elaborations of pointing. They argue that their neural network model, which involves "little or no sensory feedback" (Borrett, Kelly, and Kwan 2000, 217), is a better fit to Merleau-Ponty's claim that "From the outset the grasping movement is magically at its completion." (PhP, 103-4)

In the authors' dynamical network model, distance to the object – which would be the argument of the relevant error function in a cognitivist account – is not a quantifiable variable. They claim that this respects Merleau-Ponty's description of the spatiality of grasping. Merleau-Ponty writes that "The distance from me to the object is not a size which increases or decreases, but a tension which fluctuates round a norm." (PhP, 302) The authors pick up on the opposition suggested by Merleau-Ponty between space as an objective, measurable quantity – the "objective space" of pointing – and the "lived space" of grasping. Error functions presuppose that the distance to the target is measurable, i.e. that the target (e.g. the coffee cup) can be located within objective space. By contrast, modelling grasping behaviour on neural network relaxation does not involve positing any explicit measurement of distance. They suggest that, on their model, the only sense of distance to the target is to be understood entirely in terms of the pending evolution of the system toward an attractor state, which they suggest is
a good fit for distance as a kind of tension characteristic of lived space. Network relaxation is therefore the relief of this tension.

The authors also stress the need for an account of the relation (in fact, the reciprocity) between the spatiality of pointing and that of grasping, and provide a first pass at an account of the constitution of objective space out of lived space as they’ve construed it, in which objective distance is the result of a reflective mode that somehow tracks the number of iterations involved in network relaxation. However, this suggestion is not very clear –the notion of ‘iterations’ seems out of place in the context of a continuous trajectory through state-space – and I won’t dwell on it here.

**Perception**

The other case that Borrett, Kelly and Kwan consider is the ability to deal with perspectival distortion. For example, a plate viewed from an angle offers an ovular appearance. Cognitivists take the task of object recognition to lie in the construction of viewpoint-invariant 3-D representations out of viewpoint-dependent 2-D representations. Examples of this approach are the linear combination view of Ullman and Basri (1991) and interpolation by basis functions presented in Poggio and Edelman (1990). Even more phenomenologically sophisticated models such as that offered in Edelman and Wienshall (1991) assume that objects are 3-D representations reconstructed out of constancy relations among perspectival appearances. The authors suggest that Merleau-Ponty would be highly critical of such an approach, demanding that the self-evidence of the thing have precedence over, and ground any constancy of form across, perspectival views.

Their interpretation of the task is quite different from the cognitivists’. They treat the problem as one of transposing from a contingent, distorted appearance to the “canonical” or optimal view of the object. Following earlier work in Yoon, Borrett and Kwan (1995), they report that a neural network can be trained to evolve toward this canonical view from any other perspective – even perspectives which were not involved in the training itself. Moreover, this is achieved without

---

7 Every discussion of the non-representational character of neural networks involves the proviso “after training”. This proviso should be read as a concession to the fact that during training, classical representational constraints are applied. Dreyfus’ model of skill development is supposed to suggest that a representational characterisation which may be appropriate to the early stages of learning does not continue to be appropriate when applied to expert performance. If this were not true, then the need for representations or rules in the training of neural networks would undermine the claim that these networks offer excellent analogies to anti-representationalist descriptions of perception and action.
incorporating any error function providing comparative feedback with respect to an explicit representation of the canonical view.

Having experienced how an object appears as it undergoes translation and rotation, the network could anticipate the next image along the relaxation trajectory without the need for an explicit representation of the object other than the image as it appears in immediate experience. (Borrett, Kelly, and Kwan 2000, 222)

The authors focus on the lack of any explicit representation which would capture invariant relationships between all possible views, and argue that their model best fits Merleau-Ponty's view, because:

Far from its being the case that the thing is reducible to constant relationships, it is in the self-evidence of the thing that this constancy of relationships has its basis. (Borrett, Kelly, and Kwan 2000, 221)

The idea seems to be that what Merleau-Ponty means by the self-evidence of the thing is something like the immediate availability of anticipations about how the appearance of the thing will change under changing displacement and orientation without appeal to some form that remains invariant under these transformations.

The unity of the perceptual experience is guaranteed not because perceptual experience conforms to certain laws and relationships but because pre-scientific experience reveals an object in the world, the size and shape of which is evident in my usual daily interaction. In this immediate experience, the object is never finally constituted. It appears in perception as an image that anticipates further views and this anticipation is a property of the organism-environment system and not my reflection on the object. (Borrett, Kelly, and Kwan 2000, 221)

That is, the dynamical network model they propose lacks any explicit, general representation of the objective.

**A criticism: organism and/or environment?**

Connectionist references to organism-environment interaction, like Dreyfus' interpretation of the 'intentional arc', are somewhat controversial. The aim is to provide an account of the "appropriate [neural] substrate from which representationality itself arises" (Borrett, Kelly, and Kwan 2000, 226) – if you like, to offer an account of neurological conditions of motor intentionality. As such, their account presumes that "we can model a controller to act and perceive as if it has a human body based on current understanding of brain function without demanding that it physically act and perceive through a body." (ibid.)
Engineering practicalities will only become relevant when an attempt is made to "fully reflect the human experience," by incorporating these networks into bodies like our own. (ibid.) The authors don’t deny the need to complete the project through embodying the network and thereby providing it with a human environment. They quote the following remark by Dreyfus:

> Nothing is more alien to our life-form than a network with no up/down, front/back, no interior/exterior, no preferred way of moving, such as moving forward more easily than backwards, and no tendency towards acquiring a maximum grip on its world. The moral is that the way brains acquire skills from input-output pairings can be simulated by neural networks, but such nets will not be able to acquire our skills until they have been put into robots with a body structure like ours. (Dreyfus 1996)

But can this approach of providing a neural network model of experience first, and then implementing it in a body second, really do justice to Merleau-Ponty’s view? It may even be incompatible with Dreyfus’ own criticisms of the traditional, internalist view. Alan Costall, for one, claims to be unconvinced that these models "fully capture the radical implications of Merleau-Ponty’s philosophy" (Costall 2000, 230). In contrast to the authors’ deferral of engineering practicalities, Costall proposes that:

> Merleau-Ponty and Gibson’s point about embodiment was not that it is a convenient if incidental "input-output peripheral", if one happens to want to get about in the world. ... embodiment and situatedness are essential to any proper understanding of coordinated action... (Costall 2000, 230)

This concern is echoed by Fred Keijzer, who goes as far as to suggest that such a deferral amounts to implicitly endorsing "the traditional view to which authors like Merleau-Ponty are opposed." (Keijzer 2000, 244)

> The embedded view is definitely not a combination of inner experience and some bodily extras that can be subtracted at will. This latter view simply is the traditional one. (Keijzer 2000, 245)

It is with reference to these sorts of considerations that enactivists make the case that their models are more in tune with the spirit of Merleau-Ponty’s *Phenomenology*, since they not only reflect the letter of his descriptions, but have also (implicitly or explicitly) incorporated aspects of his philosophical standpoint into their methodological approach.
Enactive approaches

A) The Embodied Mind: Varela, Thompson and Rosch

The publication of the *Embodied Mind* by Varela, Thomson and Rosch in 1991 announced the arrival of an 'enactive approach' to cognition, an approach to cognitive science which aims to "situate cognition as embodied action within the context of evolution as natural drift" (Varela, Thompson, and Rosch 1991, 205). Enactivists contend that perception is essentially action-directed, which is to say they would agree with Ballard on the need for a shift away from pure (representationalist) vision to animate vision. However, enactivists would also contend that the visual behaviours that Ballard describes need to be understood in terms of the evolutionary coupling of organism and environment.

At one point, the authors helpfully describe their approach as mid-way between the dualism of representationalist accounts, and the monism of an ecological approach like that of J. J. Gibson. On the one hand, the authors agree with Gibson against representationalism that perception is action-directed, and that therefore the world needs to be described in terms that are meaningful for the perceiver, i.e. in terms of affordances. On the other hand, however, they disagree with Gibson when he claims that this involves picking up invariances in the ambient light. That is, affordances are not simply there to be perceived. They are rather the result of a history of coupling which codetermines the structure of both the organism and its environment. For example, the authors point to the fact that "the colors of flowers appear to have coevolved with the ultraviolet sensitive, trichromatic vision of bees." (Varela, Thompson, and Rosch 1991, 201)

Thus, the enactivist approach stresses a consummate reciprocity between organism and environment, ruling out any inquiry into perception that would take either pole of the relationship as self-sufficient. Merleau-Ponty is regarded as a forerunner to this conception on the basis of passages like the following:

> When my hand follows each effort of a struggling animal while holding an instrument for capturing it, it is clear that each of my movements responds to an external stimulation;

---

8 See pp202-204.

9 Varela et al come to speak of "perceptually-guided action", but since their reading of Merleau-Ponty is concerned with perception, I have chosen to use the locution "action-directed perception" to convey the same tight connection between action and perception.

10 Presumably enactivism could thereby avoid the criticisms levelled against Gibson's ecological approach by Jerry Fodor and Zenon Pylyshyn (Fodor and Pylyshyn 1981).
but it is also clear that these stimulations could not be received without the movements by which I expose my receptors to their influence … Thus the form of the excitant is created by the organism itself, by its proper manner of offering itself to actions from the outside. (StrB, 13)

The authors claim that, in this quote, "Merleau-Ponty clearly recognized, then, that we must see the organism and environment as bound together in reciprocal specification and selection." (Varela, Thompson, and Rosch 1991, 174) It needs to be said, however, that this quote can only be taken to support the claim about a reciprocity between action and perception, and the inadequacy of a representationalist or pure vision model of perception. It does not necessarily support the evolutionary account of the origin of that reciprocity in the "coevolution" of organism and environment.

The *Embodied Mind* claims to do more than to contribute to contemporary cognitive science, though. It introduces a new conception of the relationship between science and experience. The authors hope to "open a space of possibilities in which the circulation between cognitive science and human experience can be fully appreciated and to foster the transformative possibilities of human experience in a scientific culture." (Varela, Thompson, and Rosch 1991, xviii-xix)

The authors consider their project to be "a modern continuation" of Merleau-Ponty’s work, although they are quick to point out that theirs is not an exercise in scholarship.

By continuation, we do not mean a scholarly consideration of Merleau-Ponty’s thought in the context of contemporary cognitive science. We mean, rather, that Merleau-Ponty’s writings have both inspired and guided our orientation here. (Varela, Thompson, and Rosch 1991, xv)

Nonetheless, the authors are prepared to claim that their employment of Merleau-Ponty for enactive cognitive science constitutes "a radical new approach to the implementation of Merleau-Ponty’s vision". (Varela, Thompson, and Rosch 1991, xvii)

What is this vision? It is difficult to clarify what is attributed to Merleau-Ponty here, because of the reluctance to engage in a scholarly discussion – a decision that is, for all that, understandable given the intended audience of the book. But it seems to consist in the claim that any attempt to achieve a proper rapport between science and experience needs to take embodiment as its locus of inquiry.
The authors suggest that cognitive science is "janus-faced", and can be polarized into two extreme views:

we suppose either that our human self-understanding is simply false and hence will be eventually replaced by a mature cognitive science, or we suppose that there can be no science of the human life-world because science must always presuppose it. (Varela, Thompson, and Rosch 1991, 13)\(^{11}\)

Merleau-Ponty's insight is his recognition that embodiment itself is also janus-faced:

For Merleau-Ponty, as for us, *embodiment* has this double sense: it encompasses both the body as a lived, experiential structure and the body as the context or milieu of cognitive mechanisms. (Varela, Thompson, and Rosch 1991, xvi)\(^{12}\)

What Varela et al draw from this is that the scientific inquiry into the role of embodiment in cognition is particularly apposite for opening up the "circulation" that would resist the polarization of views just mentioned. It is through the study of embodied cognition that the sciences will learn to engage directly with human experience. However, the authors don't discuss Merleau-Ponty's conception of embodiment in any detail. Moreover, the authors announce that they are concerned with "the sensorimotor structure of the perceiver (the way in which the nervous system links sensory and motor surfaces)." (Varela, Thompson, and Rosch 1991, 173) It is this structure that the authors mean to refer to when they speak of *embodiment*.

Ultimately, however, it seems that Merleau-Ponty's work is a convenient stepping stone to a direct engagement between cognitive science and the Buddhist tradition of mindfulness meditation, which is the authors preferred basis for inquiring into human experience.\(^{13}\) It is embodiment as a theme for cognitive science, rather than embodiment per se, that seems to be the vision – attributed to Merleau-Ponty – that guides the *Embodied Mind*.

---

\(^{11}\) Stephen Stich (1983), Patricia Churchland (1986) and Paul Churchland (1979) are taken to represent the former, eliminativist extreme, while Hubert Dreyfus (1979) and Charles Taylor (1983) are taken to represent the latter, anti-naturalist alternative; though the authors note Dreyfus' subsequent rapprochement with connectionism in "Making a Mind Versus Modelling the Brain" (Dreyfus and Dreyfus 1986).

\(^{12}\) Of course, as I mentioned in the introduction, it is precisely this interpretation of Merleau-Ponty’s conception of embodiment as both experiential and scientific that Dreyfus finds so objectionable in his review.

\(^{13}\) I’ll consider the reasons for this when discussing Varela's conception of the relationship between phenomenology and cognitive science in the third part of this chapter, Securing the Convergence.
B) Consciousness makes a difference: Ellis

Ralph Ellis (2006) speaks of a "Kuhnian revolution now occurring in neuropsychology," that has lead to a newfound affinity between cognitive science and phenomenology. Ellis goes as far as to suggest that the "enactive" approach to cognition "is consistent which Husserl and Merleau-Ponty on virtually every point." (Ellis 2006, 33)

The first part of this account of their convergence focuses on the claim that consciousness should not be regarded as "an extra layer superimposed over physiological information processing, enabling us to be aware of what is going on in a computer-like subconscious mechanism." (Ellis 2006, 34) Consciousness is not the equivalent of "the effluent waste given off by a chemical manufacturing plant," (Ellis, 37) but is an important component in explaining neuropsychological functioning. That is, "consciousness drives and organizes the process rather than being a mere causal by product or spinoff" (Ellis 2006, 34)

The enactive approach grants causal efficacy to consciousness without compromising causal closure by regarding consciousness as a self-organizing system which can maintain its structure while replacing its material constituents. The explicit comparison with Merleau-Ponty comes with identifying self-organizing phenomena with the concept of a psychophysical form, to be found in the *Structure of Behavior*; 14 Ellis stresses Merleau-Ponty's assertion, against the Gestaltists from whom the concept is borrowed, that organic forms are irreducible to physical forms:

"It is not a question of risking one hypothesis among others, but of introducing a new category, the category of 'form,' ... defined as total processes whose properties are not the sum of those which the isolated parts would possess. ... We will say that there is a form whenever the properties of a system are modified by every change brought about in a single one of its parts and, on the contrary, are conserved when they all change while maintaining the same relationship among themselves." (Merleau-Ponty 1942/63, 47)

The complaint that the computational theory of mind tends to treat consciousness as epiphenomenal is taken up and radicalised as a failure to distinguish the living and the non-living, and what is specific to the living is its
purposiveness, it's reserve capacity to initiate movement, i.e. not merely to react, but to act.

This capacity for initiative has become a prominent part of the neuropsychological understanding of not only action, but perception. That is, perception means something altogether different, and requires a different explanatory framework in a system that is capable of action.\(^{15}\)

The convergence is made explicit in Ellis (2006). In the *Phenomenology*, Merleau-Ponty favourably quotes R. Déjean's remark that 'It is necessary to "look" in order to see.' (PhP, 232\(^{16}\)). Ellis suggests that this remark is to be compared with the claim that "perceptual consciousness requires self-initiated efferent activity." (Ellis 2006, 41) Ellis is referring to the conclusions of a series of neuropsychological studies (...) which show that activation of the occipital lobe by a visual stimulus does not result in consciousness of the object unless or until the parietal and frontal lobes are also activated. Other studies (...) have ruled out treating the activation of the parietal and frontal lobes as a causal consequence of the activation of the occipital lobe, the former are related to emotion-motivation activity, which is largely self-generated and self-energizing (cf. Panskepp and Damasio). (Ellis 2006, 40)

So the model of the mind as a passive receiver of causal work done by stimulus inputs and other mechanical computations places the cart before the horse. The organism must first purposely act, and only then can consciousness of the environment result. It is this fundamental shift in the direction of causation which is now sometimes referred to as the "enactive" view of the mind. Rather than a stimulus causing a response, it is the response which must occur first, and then act on the incoming afferent signals to produce a stimulus. (Ellis 2006, 41)

Ellis also finds a convergence in the description of the way this anticipatory activity and the resulting stimulation interact. Merleau-Ponty describes sense experience by saying that: "I give ear, or look, in the expectation of a sensation, and suddenly the sensible takes possession of my ear or my gaze, and I surrender a part of my body, even my whole body, to this particular manner of vibrating and filling space known as blue or red." (PhP, 212) Ellis reads this as comparable to the claim that: "Conscious experience (by contrast to unconscious information

\(^{15}\) Thinkers like Alva Noë have embraced this enactive view of perception, and recognize a debt to Merleau-Ponty in the process. See e.g. Noë (2004), 17.

\(^{16}\) The quote is taken from R. Déjean, *Étude psychologique de la distance dans la vision*, p111.
processing) entails an emotionally interested anticipation of possible sensory and proprioceptive input such that the pattern of the subject's interest determines the modality, patterns, and emotional significance of the anticipated input." (Ellis & Newton 1998, 432; quoted in Ellis, 45-46)

1.2 Philosophical Aspects

Countering mediational epistemology: Taylor

For contemporary readers within the perspective I am considering, Merleau-Ponty's philosophical relevance lies in his work's ability to break out of a particular picture about ourselves and our relations with the world in which we live. Charles Taylor in his contribution to the Cambridge Companion to Merleau-Ponty, argues that "Merleau-Ponty ... helped to break the thrall of the mediational picture." According to this mediational picture, our relations with the world are indirect and mediated (usually by representations). (Taylor 2005, 30) Taylor picks up a Wittgensteinian remark to suggest that this picture has "held us captive (heilt uns gefang)" to such an extent that it continues to distort our entire approach to epistemological questions, provoking perennial sceptical worries. Taylor argues that Merleau-Ponty’s stress on the importance of the ways in which we cope with our environment corrects this distortion and relieves those sceptical concerns.

Taylor uses this notion of the mediational view to locate Merleau-Ponty with respect to some key figures in analytic philosophy of mind. In doing so, Taylor implicitly moves Merleau-Ponty beyond the Dreyfusian portrayal of him as an anti-representationalist. Merleau-Ponty's significance does not lie solely in rejecting the mediational picture in its dualist, Cartesian form. While he shares with much contemporary analytic philosophy the recognition that it is illegitimate to posit "little objects in the mind that could be seen as a copies of external reality ... and [as] claims that something is the case," Merleau-Ponty goes much further than our contemporaries in avoiding any presupposed "distinction between inside and outside." (Taylor 2005, 27, 26)

Taylor argues that the mediational view persists even among anti-representationalist thinkers like Donald Davidson and Richard Rorty. Davidson

\[17\] Wittgenstein (1953), §115
and Rorty agree that the only thing that can justify a belief is another belief.
Moreover, they accept it as a truism. To this Taylor asks:

Why is this so obvious? Because, they insist, the only way you could find an alternative would be to "get outside our beliefs and language," in Rorty's formulation. (Taylor 2005, 29)

But in asserting that we can't get outside, "we find the picture invoked within an argument that is meant to repudiate [it]," since this idea of a separation between inside and outside is the most basic theme of the picture. "This," Taylor informs us, "is what it means to be held captive." (Taylor 2005, 29)

It helps to think of the problem in terms of the challenge of scepticism. In effect, the attempt to present a coherentist account of knowledge is one which seeks to eliminate scepticism, though not by refuting it with some sort of foundationalist claim about some indubitable knowledge we have of the world. Instead, it denies that our inability to check for a correspondence between our beliefs and the world says anything about the epistemic validity of those beliefs.

From Taylor’s perspective, and presumably from Merleau-Ponty's and Heidegger's as well, this response insulates the conception of belief, a fundamentally representationalist conception, from critique. The church of traditional epistemology that Davidson and Rorty will never step into must be a representationalist one.

The discussion comes to a head on the topic of belief-formation. Merleau-Ponty and the anti-representationalists would agree that the problem with representationalism is that it ultimately must find a way of combining a causal theory of belief-formation with a rationalist theory of epistemic validity. The way representations connect with the world (non-normative relations of causal origin) and the way they connect with each other (their normative relations of justification) seem to be incompatible, yet they need to be combined to make sense of the normative significance of reference. The advantage of Merleau-Ponty's approach is that it allows us to "think in terms of the kind of embedded knowing." (Taylor 2005, 29) Embedding knowledge does not allow us to reconcile the causal and rational connections that define a representation empirically (genetically) and epistemically in the traditional model. It allows us to reconfigure the whole problem, by subordinating the formation of belief to the establishment of a perceptual sense within a background of "abilities to get around and deal intelligently with things". (Taylor 2005, 31)
As Taylor explains:

The idea I'm invoking here is not the Quine-Davidson holism. That is a holism of verification, first of all; it reflects the fact that propositions or claims can't be verified singly. It is only derivatively a holism about meaning... The holism I'm invoking is more radical. It undercuts completely the atomism of the input because the nature of any given element is determined by its "meaning" (Sinn, sens), which can only be defined by placing it in a larger whole... (Taylor 2005, 31)

The sense of a percept, like the content of a belief, only exists by virtue of its place within a whole, but the whole – and thus the sense – is nothing like that invoked when talking about a belief or a theory: "the whole in which particular things are understood, bits of information taken in, is a sense of my world." (Taylor 2005, 32) I understand this whole not by reconciling its components with one another or testing their correspondence with something outside. I understand it in coping with (or really, within) it. Once we acknowledge the possibility and ubiquity of this kind of understanding, the mediational picture comes to seem quite far-fetched, because it just doesn't seem natural to try to analyse this kind of understanding in the absence of the worldly things to which it refers. "My ability to throw baseballs can't be exercised in the absence of baseballs. My ability to get around this city, this house, comes out only in getting around this city and house." (Taylor 2005, 33)

The very idea of an inner zone with an external boundary can't get started here, because our living things in a certain relevance can't be situated "within" the agent; it's in the interaction itself. (Taylor 2005, 38)

Although Taylor does not explicitly recognize this, exposing a kind of corporeal understanding that does not involve a solipsistic separation of subject and object is, on its own, insufficient for overturning the mediational or epistemological picture. The key feature of that picture is that it posits an interface between mind and world. Showing that we do not need to posit an interface between body and world only shifts the problem onto relations between mind and body, albeit conceived as two modes of understanding, one conceptual and one non-conceptual. We'll take this problem up in a moment, when we turn to the contemporary relevance of Merleau-Ponty's critiques of empiricism and intellectualism.

The last, but possibly the most important, aspect of Merleau-Ponty's relevance to contemporary analytic philosophy, in Taylor's opinion, is the way in which his
contribution "dissolves the temptations to antirealism." (Taylor 2005, 46) Antirealism is only a viable option so long as we remain within the mediational picture. It is only once the commonsense distinction between the way things really are and the way they appear has been reified into a dichotomy by a representationalist account that the antirealist move of denying any independent sense to the way things really are becomes appealing. It is incoherent once one adopts a standpoint, like Merleau-Ponty's, which locates all our doubts about the veracity of our beliefs or the impartiality of our opinions within the context of "the world as the all-englobing locus of my involvements." (Taylor 2005, 38)

If we see that our grasp of things is primordially one of bodily engagement with them, then we can see that we are in contact with the reality that surrounds us at a deeper level than any description of significance-attribution we might make of it. These descriptions and attributions may be wrong, but what must remain is the world within which the questions arose to which they were the wrong answers, the world from which I cannot escape because I need it ... even to know who I am and what I'm about ... My first understanding of reality is not a picture I am forming of it, but the sense given to a continuing transaction with it. (Taylor 2005, 47)

Taylor describes this as an "unproblematic realism," distinguished by the fact that it is not offered as 'a daring philosophical "thesis"'. (Taylor 2005, 39) To justify his reading of Merleau-Ponty, Taylor has to account for remarks like the following:

There is, however, no question of justifying realism, and there is an element of final truth in the Cartesian return of things or ideas to the self. (PhP, 369)

In response, Taylor asserts that what Merleau-Ponty means by "realism" is "the view according to which everything, including human thought and perception, can be explained in terms of objective, third-person processes." (Taylor 2005, 47) This is compared to Heidegger's refusal to treat Dasein as a kind of entity, because Dasein's understanding of its world "can never be simply equated with any objective processes of exchange between agent and surroundings." (Taylor 2005, 47-8)

Taylor's presentation doesn't sit well with the relevance Merleau-Ponty is taken to have for psychology. In opposition to the neural network modelling we saw in Borrett, Kelly and Kwan, and even Dreyfus' own presentation of Freeman's attractor theory as an explanation of phenomenological structures, Taylor denies that there can be a neurophysiological account of our
Chapter 1

preconceptual engagement with the world. Taylor regards the suspicion that a proper neurophysiological account would place our abilities within the head or body as the source of "that weird, post-Cartesian philosophical dream, the brain in a vat." (Taylor 2005, 34)

One might be inclined to believe that this would put Taylor on a path similar to that of the enactivist strand of cognitive science. Such a view would find support in Taylor's remark, which I mentioned earlier, that the sense of our engagement with the world is "in the interaction itself". However, we've just heard that Taylor denies the possibility of equating our preconceptual understanding with any objective exchange between agent and surroundings; a claim which seems to at least put pressure on the plausibility of an enactivist account, which does still regard co-origination as an objective process. Taylor goes on briefly to suggest that these are "modes of understanding whose bases we can never fully render transparent to ourselves." (Taylor 2005, 48) The question this raises is whether we can affirm an in-principle opacity of human understanding to itself and continue to regard the project of cognitive science as a meaningful one.

Primacy of perception: Carman

Merleau-Ponty's dual critique of empiricism and intellectualism is central to the Phenomenology of Perception. It is arguable that Merleau-Ponty's descriptions are inseparable from this dialectic rhetoric. It is therefore no wonder that commentator's, like Taylor Carman, are motivated to find contemporary analogues for these positions. Carman attempts to update our sense of the originality of Merleau-Ponty's description of the "phenomenal field" by reviewing Merleau-Ponty's critiques in contemporary terms. But Carman's review is also important because he raises the key theme of the normative character of our bodily embeddedness.

Sensation and empiricism

While the concept of sensation is firmly entrenched in the language we use to speak about perception, Merleau-Ponty argues, following the Gestalt psychologists, that the idea of a pure sensation, of an "undifferentiated, instantaneous dot-like impact" which purportedly constitutes the most basic way in which I am affected by things, in fact "corresponds to nothing in our experience." (PhP, 3) This only demonstrates that sensations should not be part of any rational explanation of behaviour. However, it still leaves room for
analyzing experience into sensations for the purpose of explaining, though not understanding or evaluating, perception.\textsuperscript{18}

Merleau-Ponty denies even this explanatory role for sensations. He writes:

> When Gestalt theory informs us that a figure on a background is the simplest sense-given available to us, we reply that this is not a contingent characteristic of factual perception, which leaves us free, in an ideal analysis, to bring in the notion of impressions. It is the very definition of the phenomenon of perception, that without which a phenomenon cannot be said to be perception at all. (PhP, 4)

The crucial point, however, is not merely the indiscreteness of these elementary perceptibles, but the fact that when we explore their contextual character, we find that they are what they are only with reference to other features of the object to which they belong, and even the scene within which we find that object. Once we recognize that "this red would literally not be the same if it were not the 'wooly red' of a carpet", we must conclude that "quality is not an element of consciousness, but a property of the object... indeed the whole spectacle."

Merleau-Ponty therefore warns that it is a mistake to "make [quality] into an element of consciousness, when in fact it is an object for consciousness." (PhP, 4-5) That is, against the empiricist view that it is composed of atomic elements, and can be analyzed like a collection of facts, Merleau-Ponty argues that perception is essentially \textit{intentional}.

How do we come to the empiricist view, though, if there is no basis for it in our experience? Merleau-Ponty suggests that two errors are involved. The first of these is in committing "what psychologists call the 'experience error', which means that what we know to be in things themselves we immediately take as being in our consciousness of them. We make perception out of things perceived." (PhP, 5)

The second, which is distinguishable but nonetheless complementary to the first, is to suppose that the objects of perception are completely discrete from one another and fully determinate.\textsuperscript{19} That is, that the objects perceived are just as well-defined as the objects of scientific inquiry.

\textsuperscript{18} In this distinction between explaining and understanding, we can already see the seeds of the opposition between empiricist and intellectualist conceptions of consciousness. The very condition of their opposition lies, according to Merleau-Ponty, in the postulation of sensations.

\textsuperscript{19} This is close to what Dreyfus called the "ontological assumption" in his critique of Good Old-Fashioned AI (GOFAI), i.e. the assumption 'that the world can be exhaustively analyzed in terms of context-free data or atomic facts.' (Dreyfus 1992, 205)
Chapter 1

The determinate quality by which empiricism tried to define sensation is an object, not an element, of consciousness, indeed it is the very lately developed object of scientific consciousness. (PhP, 6)

Carman (2005) summarizes Merleau-Ponty’s argument by saying that the problem with the concept of sensation is that it attempts to fulfill two incompatible functions. The concept is supposed to provide both a description and a causal explanation of perceptual experience. The concept fails, however, to describe experience just in so far as it attempts to explain it in terms of a relationship between determinate objects – i.e. a relationship between the object perceived, construed as the determinate cause, and our perceiving of it, construed as a determinate effect, i.e. the stimulus of our sensory surfaces. (Carman 2005, 58)

As Carman points out, "the ordinary notion of sensation is meant to capture how things look." (Carman 2005, 55-56) As soon as it is laboured with the additional task of capturing the way I am causally affected by the world, it is unable to fulfill its more basic function. Take the example of the Müller-Lyer illusion. If sensation were to fulfill its task of capturing how things look, then the sensation should reflect the (phenomenological) fact that the lines look unequal. However, given its other task of expressing the causal impingement of a scene on the perceiver, the sensation must also reflect the fact that the lines are measurably the same length.

Judgement and intellectualism

Merleau-Ponty treats intellectualism as defined by its refutation of empiricism. For example, he writes that judgement, the key concept for the intellectualist, "is often introduced as what sensation lacks to make perception possible." (PhP, 32) Phenomena like the Müller-Lyer illusion, which thwart the empiricist idea that appearance should be isomorphic with objective stimulus, are grist for the intellectualist’s mill, which interposes between the objective stimulus and the appearance the concept of attention, interpretation or judgement.

Descartes and Kant20 count among the intellectualists that Merleau-Ponty criticizes. Carman points out, for example, that Descartes argues that:

20 Husserl is often – erroneously – included in this list. Taylor Carman demonstrates an uncommon respect for Husserl when he resists the temptation to mention him in the context of an exposition of Merleau-Ponty’s critique. By contrast, in his article for the same collection, Sean Kelly frames his understanding of Merleau-Ponty by criticising Husserl. (Kelly 2005)
imperfect or confused perception is not a matter of having defective or obscure material available for mental scrutiny, but of scrutinizing it imperfectly or confusedly. What is given by God cannot be imperfect; error or illusion flow from our own willful misconstructions. (Carman 2005, 59)

Illusions like that described by Müller and Lyer are perfect examples. Because we fail to pay a careful and diligent attention to the main, parallel lines, but are somehow distracted by their auxiliary rays, we misjudge their length and take them to be unequal. The intellectualist will explain the moon’s appearing larger on the horizon as the result of the fact that the presence of objects on the horizon distorts our interpretation of its size, while at a higher elevation, no such distortion occurs.

The intellectualist may also point to the fact that such illusions can be overcome by modifying the visual field in some way. For example, if we remove or conceal the auxiliary rays, the main lines of the Müller-Lyer illusion will be recognized as equal. Similarly,

if we look at [the moon] attentively, for example, through a cardboard tube or a telescope, we see that its apparent diameter remains constant. (PhP, 27)

Carman suggests that Dennett can plausibly be regarded as a contemporary target of Merleau-Ponty’s critique of intellectualism.

Dennett is an even more extreme intellectualist than Descartes, for whereas Descartes’s characterization of all mental phenomena as modes of “thought” is largely a terminological idiosyncrasy, Dennett maintains that every conscious experience, even the most visceral and concrete, is literally a kind of judgment or supposition that something is the case. (Carman 2005, 60)

Take the case of apparent movement, popular with Gestalt psychologists; for example, the apparent movement of the text on a scrolling LED display. The pertinent question is whether this movement is seen or merely inferred. In particular, do we see continuous motion, or do we infer that the motion is continuous on the basis of a sequence of successive perceptual images? To be sure, you seem to see a continuous movement, but does that not merely imply that, along with believing that the movement is continuous, you also believe that you are seeing a continuous movement. For Dennett, “quite literally, seeing is

---

21 This sort of phenomenon is commonly and mistakenly referred to as a phi phenomenon. In fact, it is better described as a beta phenomenon. Phi phenomena are actually much more interesting. See e.g. Steinman, Pizlo and Pizlo (2000).
believing," i.e. it is believing that you see X; "to lack a belief about a perceptual experience is to lack the experience altogether." (Carman 2005, 62)

Dennett's position focuses on situations in which it is difficult to distinguish between what is present in one's visual field from what one takes (i.e. judges) to be present in that field. He uses these situations to argue for:

\[ \text{the radical thesis that, although we ordinarily suppose things are given in perception, and that we then form judgments about them, there is in fact no difference in principle between a perceptual experience and a judgment about a perceptual experience. (Carman 2005, 61)} \]

Carman reads Merleau-Ponty as disputing the intellectualists' erosion of the distinction between perception and judgement. The following passage, Carman argues, might almost have been written with Dennett in mind:

Ordinary experience draws a clear distinction between sense experience and judgement. It sees judgement as the taking of a stand, as an effort to know something which shall be valid for myself every moment of my life, and equally for other actual or potential minds; sense experience, on the contrary, is taking appearance at its face value, without trying to possess it and learn its truth. (PhP, 34)

So Merleau-Ponty presents himself as a defender of a distinctive role for perception in contrast to judgement, even if there are situations in which the distinction between them is blurred. He denies the intellectualist the right to treat these cases as typical or as having the significance the intellectualist takes them to have.

Merleau-Ponty also challenges whether the intellectualist can actually account for those fringe phenomena on which her radical thesis rests. Merleau-Ponty observes that the intellectualist must treat illusions as erroneous judgements. "Intellectualism simply reduces the phenomenon to a mistake, bringing in the auxiliary lines and their relation to the main ones, instead of comparing their main lines themselves." (PhP, 35) The question that needs to be asked though is why I bring in these auxiliary lines at all. Have I simply mistaken the task given to me? If so, then why do we all mistake this task in particular? To explain our judgement of the lines, do we really need to appeal to a judgement concerning the task? And why is it so hard to correct the mistake, if it is merely a misunderstanding that I presumably can understand?
For the intellectualist, the only way perception can contribute to judgement is by providing – in Spinoza’s words – "conclusions without premises". So the structure of the percept is analyzed as a collection of beliefs about objective relationships between the various parts of the scene. In this way, the intellectualist continues to preserve the constancy hypothesis, which treats these objective relations as the only relations that make up our cognitive milieu. But if this were the case, why should I be unable to recognize an inverted landscape, when all these objective relationships remain identical to the more natural case?

So, on the one hand, the intellectualist goes too far in seeing the inadequacy of the sensationalist account of perception as legitimizing the subordination of perception to judgement. On the other, the intellectualist does not go far enough in failing to question the constancy hypothesis, with its presumption that only objective relationships are relevant to our cognitive apprehension of a scene.

Carman points out that Merleau-Ponty’s core critique of intellectualism is that (like empiricism) it "helps itself uncritically to the view of the world as described by the physical sciences." (Carman 2005, 67) For Merleau-Ponty, by contrast, "Perception is not a science of the world." (PhP, x)

**Normativity and body schema**

Carman attributes to Merleau-Ponty the achievement of having "effectively reconceived perception itself as neither a mere registration of stimuli nor a radically free initiation of mental acts, but as the way in which the body belongs to its environment, the essential interconnecteness of sensitivity and motor response." (Carman 2005, 68) This constitutes "a fundamental challenge to the conceptual distinction between the mental and the material," in so far as "body and world are conceptually, not just causally, two sides of the same coin." (Carman 2005, 68)

The key feature for Carman is, however, the normativity of the relation between my body and the perceptual milieu that this implies. It means something to perceive a scene **properly**, presumably as opposed to perceiving it inadequately or askew, by virtue of a certain "bodily equilibrium", which, when upset, is experienced as an incapacity, discomfort or a distortion. To talk of "motor intentionality" is to talk of "our bodies ... constantly adjusting themselves to integrate and secure our experience and maintain our effective grip on things." (Carman 2005, 69)
This self-correcting grip on things constitutes a body schema, akin to "Kant's insight that intentional content does not just magically crystalize in the mind but is so to speak sketched out in advance by the dispositions that allow things to appear to us as they do." (Carman 2005, 70) These dispositions had to have the structure of explicit rules for Kant, whereas for Merleau-Ponty they are understood in terms of favoured postures, habits and skills. Together these provide "a normative domain, without which perception could not be intentional."

Indeed, what makes motor intentionality worthy of the name is precisely its normativity ... The rightness and wrongness of perceptual appearances is essentially interwoven with the rightness and wrongness of our bodily attitudes, and we have a feel for the kinds of balance and posture that affords a correct and proper view of the world. (Carman 2005, 70)

What Carman has in mind are the very familiar ways in which we get ourselves in a position to best perceive something – holding something up to the light better to see its colour, circumnavigating an unusual structure to get a grip on its proportions or shape, leaning toward the photograph in order to find a familiar face, etc. The idea is that these adjustments are not simply reactions to causal stimuli, though naturally, a causal relationship is implied. They are normatively governed. This normative character of the perception – i.e. its capacity to refer our bodies to a better posture or position from which to apprehend it – is not only essential to any accurate description of perception. It is also precisely what gets lost when one takes the constancy hypothesis as one's starting point, since the objective world that resides at the origin of the empiricist account, and at the destination of the intellectualists, is what it is entirely irrespective of our seeing it aright. This depersonalisation of the objective world, while making it the cradle of the most lasting truths, is what makes that world incapable of acting as the condition of the intelligibility of human cognition.22

1.3 Securing the Convergence

There is a variety of views on how to appropriate phenomenological insights for cognitive science. In this sub-section, I consider those of Dreyfus and Varela, as well as McClamrock and Gallagher. Dreyfus and McClamrock regard

---

22 We may, however, have taken Carman further than he would wish, by speaking of different worlds. Though Merleau-Ponty differentiates between the perceived and the objective world, Carman refers singularly to "the world".
phenomenology as prior to cognitive science, as providing its explicanda (Dreyfus) or initial taxonomy (McClamrock), while Varela and Gallagher see the two as developing synchronically, either as complementary methods within a single inquiry (Varela), or as complementary inquiries (Gallagher).

Explicanda for psychology: Dreyfus

In his introduction to *Husserl: Intentionality and Cognitive Science* (1982), Dreyfus makes an analogy that characterises the relationship he sees between phenomenology and psychology. There, he suggests a parallel between Husserl's prescription of an *epoché* or bracketing of the existence of things, and Fodor's methodological solipsism.

Fodor argues that any causal account of intentional states would have to wait for a completed account of physics (and possible all other science) before it could begin to type-individuate intentional states. Thus, Fodor suggests that any rational approach to cognitive (i.e. intentional) psychology must adopt a representational theory of mind – i.e. a theory which regards mental activity as the processing of representations which can be individuated without concern for the causal relations among their referents – to avoid having "wait forever" for its taxonomy.

Dreyfus suggests that Husserl's *epoché*, i.e. that phenomenological reduction that enables the development of a phenomenological psychology, amounts to the same methodological prescription. Like Fodor's methodological solipsism, it involves abstracting away from any naturalistic or causal considerations to establish the appropriate taxonomy of psychology – which is what it means to say that phenomenological psychology is an eidetic science of consciousness.

In Hilary Putnam's terminology, taken over by Fodor, this bracketing of the concerns of naturalism, along with the implicit denial of the relevance of the causal component of reference, makes Husserl a methodological solipsist. ... Fodor holds that the cognitive psychologist must make precisely this "epoché," and he contends that the cognitive psychologist must be content with a solipsistic psychology, since he will never have the final story on the physical world and thus on what is causally affecting the psyche, and since even if he could have a final physical theory, there might well be no theory of the complicated social and psychological interactions which determine reference. (Dreyfus 1982, 15-16)
Even though Dreyfus will come to dismiss this approach as motivated by an erroneous commitment to internalism, this claim nonetheless effectively sets the scene for Dreyfus' conception of the relationship between phenomenology and cognitive science, since Dreyfus will continue to regard the sole relevance of phenomenology to lie in its capacity to provide explicanda for psychological inquiry.

At least that is the impression Dreyfus gives. He never spells out his conception of the relationship in any detail, preferring to refer to it as an isomorphism, for example between Merleau-Ponty's descriptions and Freeman's attractor theory. However, if we presume, and I think it is reasonable to presume, that Dreyfus wants to suggest that this isomorphism lends some weight or legitimacy to Freeman's theory, then Dreyfus risks embracing the same circular justification for which he criticised cognitivists. For example, in Dreyfus (1992), Dreyfus points out that the only justification cognitivists have for treating the linguists' digital model of built-in rules as an improvement over the Gestaltists' analogue equilibrium model is that it better suits the cognitivists' own digital hypothesis. "It would indeed be nice to have a programmable model in linguistics, and in psychology in general," Dreyfus admits, but there is no reason to think that nature will be that kind. (Dreyfus 1992, 247)

It seems to me that at least the rules the linguists are looking for would be, if they were found, the sort of rules one could process with a digital computer which we already understand whereas the gestaltist equilibrium principles could only be simulated on a brain-like analogue computer, which no one at present knows how to design.

This is no doubt true, but it reminds me of the story of the drunk who lost a key in the dark but looked for it under a street lamp because the light was better. (Dreyfus 1992, 247)

Now, the crux of this argument is that the mutual compatibility of the working hypotheses may be convenient for researchers, but is not justification for endorsing them. Could we not make a similar criticism of the pairing of Merleau-Ponty's phenomenology and Freeman's neural attractor theory? Dreyfus must argue either (a) that the neural attractor theory confirms the phenomenology, or (b) that the phenomenology justifies the neural attractor theory; but he can't have it both ways.
Taxonomies for intentional psychology: McClamrock

One promising development of Dreyfus' analogy is offered by Ron McClamrock in his *Existential Cognition* (1995). McClamrock takes up Dreyfus interpretation of Husserl's phenomenological reduction as a kind of methodological solipsism. (McClamrock 1995, 183ff) McClamrock also suggests, like Dreyfus, that the existentialist turn is akin to embracing an externalism about mental content. 23

However, on the other hand, and unlike Dreyfus, McClamrock does not really look to Merleau-Ponty for an alternative account of perception, so much as an alternative methodology for approaching the same question that motivates Fodor's methodological solipsism: namely, what is the appropriate taxonomy for intentional psychology?

McClamrock's more general goal in his book is to transform this into a genuinely scientific question; that is:

> to contribute toward transforming the general individualism vs. holism question from a largely metaphysical or even verbal dispute to a real, at least partially empirical question with a solution that avoids the triviality of simplicitic reductionism ("only the little individuals matter") or the vapidity of complete ecumenicalism ("any level – individualistic or holistic – is as good as another"). (McClamrock 1995, 31)

If the appropriate taxonomy of psychology is an empirical issue, then it cannot and should not be finessed by an appeal to some general methodological principle of science, à la Fodor. McClamrock proposes that the question be tied to the systematicities of behaviour that intentional psychology is supposed to capture. He then makes the point (often expressed) that to do this will probably require that we individuate mental states in ways that take into account aspects of the ecological situation of the organism, aspects which may not be explicitly represented by the organism at the time. 24 That is, to capture the systematicity of intentional behaviour, psychology will likely have to take its objects to be broader than the internal states of the organism.

---

23 “The existential turn in the phenomenological tradition involved a shift from taking intentional properties of thought as *immanent* (or directly present to reflection), as in Husserl, to seeing them as essentially determined by their relationship to the world, as in Heidegger and Merleau-Ponty. … By relocating the determinants of the content of thought and action into the shared world – the environment – the structure of intentionality becomes analyzable via the structure of the intentional objects, which are now the objects in the environment with which we all interact.” (McClamrock 1995, 191)

24 The idealization about the organism which underlies an *intentional* characterization of its states may not only idealize on the basis of its current internal state, but also idealize about what will happen on the basis of those internal states, given the organism's ecological situation. So in some pretty interesting and systematic ways, it might seem that the causal power (at least for the purposes of psychology) of physically type-identical thoughts can differ. (McClamrock 1995, 35 - original italics)
However, Fodor has a threat waiting for anyone who toys with methodological individualism: before they can even begin to construct a non-individualistic intentional psychology, they will be condemned to wait forever in the purgatory of unfinished science, that is:

if individuation of contents is done via their "external" semantic features (like their referents), we won't be able to type individuate mental states without finishing up our chemistry (and presumably, the rest of our sciences) first. As Fodor puts it: "No doubt it's all right to have a research strategy that says 'wait awhile.' But who wants to wait forever?" (Fodor 1987, 248; McClamrock 1995, 186)

The idea that behaviour should be explained in terms of the contents of mental states is utterly central to the idea of intentional psychology, so if we can't distinguish mental contents without completing the non-intentional science first, it seems as though we'll never get started.

McClamrock disputes the cogency of Fodor's threat, on the grounds that it implies that the intelligibility of any explanation of a higher-order phenomenon depends on the complete explication of the lower-order phenomena upon which it supervenes (taken in a minimal sense). To see why this is untrue, consider a simple case from a different domain. We strongly suspect that the air-flow control properties of a throttle determine its correlation with engine speed. But we would never think to say that we cannot begin to explain engine speed in terms of throttle position until we have cashed out those air-flow control properties in terms of a completed fluid-dynamics. While we haven't got a completed fluid dynamics, we do indeed know what throttles are and we do (roughly at least) understand the properties that make them causally potent with regard to engine speed. Similarly, though we may not be able to fully type-individuate mental states without a completed science, we certain do have something with which to begin to construct an intentional psychology, namely the rough intentional descriptions that we ourselves ordinarily provide.

Fortunately, we don't have to wait for an ideal science to pick out the objects and kinds in the world such that relations to them are intrinsic to the taxonomizing of psychological states that captures rational systematicities. We can pick them out ourselves. We are, clearly, terrific detectors of the intentional objects of human thought. At a first glance, we simply need to use ourselves as the detectors of the properties in the world relevant for the explanation of human behaviour and intelligence.
This may in fact suggest that the only kind of science that will in any way characterize such facts for us will be in a certain sense an *intrinsically human science.* (McClamrock 1995, 186-7)

That is,

There might well be explanations of the behavior of complex systems (e.g., human beings in context) where the appropriate taxonomy is such that its properties can only be detected by systems suitably like those being explained; in this case, human cognitive systems. (McClamrock 1995, 187)

This is precisely where the significance of Merleau-Ponty’s existentialist modification of Husserl’s reduction becomes critical, for McClamrock reads Merleau-Ponty as interpreting the reduction as a kind of “heuristic for getting us to notice certain aspects of the real, transcendent world,” namely, those properties that make up the basic taxonomy of human cognition.

Merleau-Ponty’s reconstruction has the effect of transforming the idea of phenomenology somewhat – from the idea of a transcendentally purified analysis of the immediately given structures of consciousness, to something more like a reflective examination of the general structure of human directedness, *and the ontology of the world that it presupposes.* (McClamrock 1995, 188 – my emphasis)

Merleau-Ponty’s reconception of “bracketing as heuristic” offers a methodological response to Fodor’s threat. Bracketing allows us to bring out what McClamrock calls the “task content” of behaviour, and the properties of the context that are relevant to its systematicity (i.e. the intentional objects of that behaviour). Phenomenology, as Merleau-Ponty comes to construe it, is that first-personal science of cognition that is necessary for developing a intentional psychology appropriate to complex, contextually embedded beings like ourselves. Phenomenology is thus in the business of developing the appropriate taxonomy of intentional psychology, a description that strongly resembles Husserl’s own description of phenomenological psychology.

McClamrock goes further, however, in suggesting that the information-processing hypothesis characteristic of cognitivism might be understood as just another heuristic. It doesn’t reveal the same task content revealed by the phenomenological reduction, but instead the process structure of cognition. In any event:

Both sides are needed: The conception of the activity as contentful and the conception of the underlying mechanisms
McClamrock's focus on a methodological alternative to Fodorian individualism allows him to present Merleau-Ponty as an advocate for externalism, without equating externalism with a rejection of the information-processing hypothesis.

**Neurophenomenology: Varela**

*Modifying the methodology of neuropsychology*

Throughout his career, Francisco Varela attempted to discern the methodological implications of the newly found relevance of Merleau-Ponty's phenomenology for cognitive science. In his work with Thompson and Rosch, especially in the *Embodied Mind*, this took the unusual form of advocating the development of a variety of mindfulness meditation as a practical component of the investigation of conscious experience. What connection do the authors see between phenomenology and meditation?

The *Embodied Mind* takes seriously Husserl's criticism that the Galilean style of theorisation obscures the relationship between scientific accounts of the world and human experience. However,

> Husserl's turn toward experience and 'the things themselves', was entirely *theoretical*, or, to make the point the other way around, it completely lacked any *pragmatic* dimension. It is hardly surprising, therefore, that it could not overcome the rift between science and experience, for science, like phenomenological reflection, has a life beyond theory... (Varela, Thompson, and Rosch 1991, 19)

In the authors' opinion, Merleau-Ponty takes a step toward overcoming this rift. Merleau-Ponty is praised for his illustration in the *Structure of Behavior* of how a relation of "mutual illumination" might exist between "phenomenology, psychology and neurophysiology". (Varela, Thompson, and Rosch 1991, 15) As I mentioned earlier, that mutual illumination is made possible by a focus on the "double sense of embodiment". Merleau-Ponty is taken to show us how to avoid choosing between seeing our bodies as *either* "outer," "biological" or "physical structures", or as *inner," "phenomenological," "lived, experiential structures". Instead, we learn to recognize that "we continuously circulate between them." (Varela, Thompson, and Rosch 1991, xv)
However, the criticism of Husserl's theoretical bent extends to Merleau-Ponty, who – like Heidegger – "stressed the pragmatic, embodied context of human experience, but in a purely theoretical way". The authors suggest that Merleau-Ponty came to recognize this as an inherent limitation of phenomenology:

In Merleau-Ponty's view, both science and phenomenology explicated our concrete, embodied existence in a manner that was always after the fact... precisely by being a theoretical activity after the fact, it could recapture the richness of experience; it could be only a discourse about the experience... Merleau-Ponty admitted this in his own way by saying that his task was infinite. (Varela, Thompson, and Rosch 1991, 19)

It is on this basis that the authors turn to the Buddhist philosophy of Magarjuna, presented in the Madhyamika ("middle way") for a new method. The hope is that the mindfulness meditation that this tradition endorses can grant an access to human experience which is not 'after the fact', thereby achieving what phenomenology could not.

But the word 'access' is misleading here. It suggests a clear distinction between experience and reflection upon that experience, where the authors stress that "reflection is not just on experience, ... reflection is a form of experience itself." (Varela, Thompson, and Rosch 1991, 28) The idea is that mindfulness meditation can put one in a position in which reflection would not result in the reification of the mental à la Descartes.

We are suggesting that Descartes's conclusion that he was a thinking thing was the product of his question, and that question was a product of specific practices – those of disembodied, unmindful reflection. Husserlian phenomenology, though it embraced experience in a radical way, nonetheless continued the tradition by reflecting only upon the essential structures of thought. And even though it has recently become quite fashionable to criticize or "deconstruct" this standpoint of the cogito, philosophers still do not depart from the basic practice responsible for it. (Varela, Thompson, and Rosch 1991, 28)

It is this theme of praxis, this view of both philosophy and neuropsychology as experimental inquiries into experience that is taken up by Varela in his later work.

Reduction as a methodological response to the Hard Problem
In his "Neurophenomenology" paper (1996), Varela places himself among a "motley" group of theorists, including Ray Jackendoff, George Lakoff and Mark Johnson, John Searle, G. Globus, Owen Flanagan and David Chalmers, who are distinguished by "giving an explicit and central role to first-person accounts and to the irreducible nature of experience." (Varela 1996, 333) The group is motley because of the different methodological consequences they see as a result of this declaration of irreducibility. For example, Jackendoff is almost silent on the issue, suggesting that experience constrain the computational theory of mind without providing any methodological prescription except an atmosphere of mutual trust. To this Varela responds with:

Mutual trust indeed! What is needed is a strict method and that is where both the difficulty and the revolutionary potential of the topic lie. (Varela 1996, 335)

But the most striking difference between Varela and the other members of this entourage is that his aim is not to provide a missing ingredient that would bridge the gap between phenomenal/lived experience and a third-person account of cognition – i.e. solve Chalmer's "hard problem" with some theoretical fix. Instead Varela wants to use what he calls the phenomenological "gesture of reduction" as a methodological response to the problem. As a result:

The nature of 'hard' becomes reframed in two senses: (1) is it hard work to train and stabilize new methods to explore experience, (2) it is hard to change the habits of science in order for it to accept that new tools are needed for the transformation of what it means to conduct research on mind and for the training of succeeding generations. (Varela 1996, 347)

It is a lack of "phenomenological expertise" that lies "at the root of the opacity of consciousness for science today". (Varela 1996, 341) What the 'hard problem' points us to is the need for the development of the practice of phenomenology, i.e. to build a tradition of inquiry that is rigorous, intersubjectively validated and respects the irreducibility of first-person phenomena. However, this inquiry is to be not only a complement to cognitive science, but an effective constraint:

The novelty of my proposal is that disciplined first-person accounts should be an integral element of the validation of a neurobiological proposal, and not merely coincidental or heuristic information. This is why I choose to describe the situation by the hypothesis that both accounts be mutual constraints on each other. (Varela 1996, 344)
Reduction without anti-naturalism

This use of the reduction poses a scholarly problem for Varela, however. His aim of developing a rigorous, phenomenological tradition of inquiry into consciousness that would act as a mutual constraint upon cognitive science stands at odds with the deeply anti-naturalistic interpretation that the founder of phenomenology gave to his method.

It is this internal tension in the project that acts as the impetus behind the publication of the *Naturalizing Phenomenology* (1999) collection. Working with J. Petitot, B. Pachoud and J-M. Roy, Varela attempts to divorce Husserlian phenomenology from Husserl's own antipathy toward any naturalistic science of consciousness. The aim is to develop a phenomenology that is "naturalized or naturalistic … in the minimal sense of not being committed to a dualistic kind of ontology." (Petitot et al. 1999, 19) The editors of the collection recognize in their introduction that this stands at odds with Husserl's claims, for example in *Ideas I*, that consciousness "is not a component of Nature, and is so far from being one that Nature is possible only as an intentional unity motivated in transcendental pure consciousness by immanent connections." (Husserl 1913/82, §51; quoted in Petitot et al. 1999, 38)

Linking Husserlian descriptions of cognitive phenomena and the contemporary sciences of cognition thus seems to require cutting Husserlian phenomenology from its antinaturalist roots, that is to say, naturalizing it. Is this possible? And if so, how? (Petitot et al. 1999, 43)

Varela and his colleagues offer the clarification that, by naturalisation, they mean to "make intelligible how an entity can have both the properties characteristic of matter and those characteristic of mentality in spite of an apparent homogeneity between them." (Petitot et al. 1999, 46) The paradigm of naturalisation is the neutralization of the conflict between "sublunar" and "supralunar" worlds by classical mechanics. The editors presume that it will be mental properties that will have to be transformed to achieve a similar reconciliation in the study of consciousness.

Apart from the parallelism that Husserl sees between phenomenological psychology and transcendental phenomenology, Varela and the other editors regard the development of Husserl's thought by his followers to be a strong indication of the plausibility of their project. Merleau-Ponty plays a particularly prominent role here.
Merleau-Ponty's use of "explicit references to the neurosciences and psychology of his time," his use of "basic and clinical materials," and his strong affinity for Gestalt psychology and Goldstein's account of the organism, all suggest that he has reorientated Husserl's phenomenological perspective in a broadly naturalistic direction. Just as important, however, is his emphasis on "the body as the locus of the experiencing subject." According to the editors, "This represented a daring departure from Husserl's transcendental leaning most in view during Merleau-Ponty's time". (Petitot et al. 1999, 33) Thus Merleau-Ponty is presented by Varela as an explicit forerunner to the naturalisation of phenomenology.

Mutual enlightenment: Gallagher

Varela's proposal that we think of the relationship between cognitive science and phenomenology as one of 'mutual constraint' is thrown into relief by Shaun Gallagher's preference for conceiving it as one of 'mutual enlightenment'. Gallagher's view is in one sense more expansive than Varela's: Gallagher suggests that, in addition to training cognitive scientists in phenomenological method in order to practice neurophenomenology, one can also engage in what Gallagher calls "front-loaded phenomenology", i.e. one can use phenomenological insights to influence experimental design in the cognitive sciences.

Gallagher (2003) describes neurophysiology experiments by Farrer and Frith (2001) and Chaminade and Decety (2002), which are informed by a phenomenological distinction between senses of ownership and agency. These experiments do not involve any phenomenological training of the subjects involved, since the subjects are not even required to give experiential reports. In Chaminade and Decety (2002), for example, the subjects are required to lead or to follow an image on the screen by moving a mouse, while PET scans are taken of their brains. The experiment revealed a distinctive activation of the left inferior parietal cortex over the right when the subject is undertaking (agency-rich) leading movements. Gallagher regards the experiment as a paradigmatic example of front-loaded phenomenology, because "it is clear that the phenomenological description of the sense of agency both informs the experimental design (the experiment is set up to find the neural correlates of precisely this experience) and part of the analytic framework for interpreting the results." (Gallagher 2003, 94) In its essence, then, front-loaded phenomenology is about using phenomenological descriptions to guide the kinds of questions that experimenters attempt to answer, and not just the kind of data that they collect.
In another sense, Gallagher’s view is more restrictive than Varela’s, in that he is sympathetic to objections to the constraint of phenomenological descriptions by empirical results within cognitive science.

In another sense, Gallagher’s view is more restrictive than Varela’s, in that he is sympathetic to objections to the constraint of phenomenological descriptions by empirical results within cognitive science.

The putative constraint phenomenology receives from cognitive science – is harder to argue out. Because these phenomenologists will argue – it doesn’t matter how the brain operates, it’s not going to change experience; phenomenological descriptions of consciousness will remain the same no matter what the neuroscientists or psychologists discover. (Gallagher and Sasma 2003)

This is not to say that cognitive science can’t inform phenomenological research, just that it cannot correct it. Take the well-known study of inattentional-blindness, in which a subject is asked to concentrate on counting the number of basketball passes by a particular team, and in doing so, completely fails to notice the guy in the gorilla suit who walks into the middle of the group and waves at the camera. Gallagher acknowledges that even a trained phenomenologist would leave the gorilla out of their description of the experience. Nonetheless, Gallagher balks at the idea that the cognitive scientist is in any position to criticise the phenomenological description as false. It does not follow from the fact that the gorilla is there, or – more accurately – that the experimenter can see the gorilla, that the subject also really sees the gorilla.

That said, the experimental result is still relevant to phenomenology since "cognitive science is … enlightening the phenomenologist about something that has to do with the structure of attention – and this is surely something the phenomenologist should be interested in." Specifically, it should prompt the phenomenologist to ask: "What is it about my experience, or about the structure of my experience, that prevents me from seeing the gorilla?" (Gallagher and Sasma 2003)

Here we come up against a difficulty, because it is quite tough to even pose such a question in properly phenomenological terms. How can we treat the absence of any experience of the gorilla as a part of the structure of my experience, when phenomenology attempts to examine such structures starting from the experience alone? How is the phenomenologist meant to incorporate this information into their inquiry? As Sasma points out, it isn’t clear that the phenomenologist and the cognitive scientist have the same object of reference. For the phenomenologist, the objects of reference are the "purely formal correlates of my intentional acts", while the cognitive scientist is talking about real objects, object that have
presumably been bracketed by the phenomenologist in performing the *epoche*.
(Gallagher and Sasma 2003)

To this, Gallagher responds by saying that the cognitive scientist and the
phenomenologist do share the same object of inquiry, "both .. are interested in the
experience" (Gallagher and Sasma 2003), but it's not clear though that this
response adequately addresses the issue, since Gallagher does think that
experience is intentional, and therefore unintelligible in abstraction from the kind
of object it is directed toward.

While he doesn't stick to it exclusively, one can detect in Gallagher's writings a
quite distinct, alternative conception of the relationship at issue. For the other
authors we considered, the central question is of finding a way in which results or
methods can be share or coordinated between the two disciplines. This is
achieved more directly, as on Dreyfus' explanandum-explanans view, or more
indirectly, via shared or coordinated methodologies, as in McClamrock and
Varela. These approaches look toward a convergence of phenomenology and
cognitive science, as two practices within the same disciplined study of the mind.
The alternative one can hear in Gallagher's work involves an exchange of
questions, rather than of results or methods. The relationship is interrogative,
rather than complementary, in character.

Gallagher often reverts to a view in which cognitive science verifies
phenomenological descriptions, and phenomenology clarifies the object of
cognitive scientific research. Nonetheless, hints of the alternative are available.
What I find attractive about an interrogative model is that it offers the prospect
of preserving a more radical distinction between phenomenology and cognitive
science, that is, of granting them an autonomy that is more than a difference in
method, as though everyone had the same task in view. I merely want to raise
that prospect at this point, in the hope of taking it up later in the thesis.

1.4 Merleau-Ponty Naturalised?

Merleau-Ponty's work has clearly played an important role in inspiring and
legitimising the development of post-cognitivist varieties of cognitive science. As
we saw – for example, in our discussion of Borrett, Kelly and Kwan's neural
network models – Merleau-Ponty's descriptions have provided starting points for
scientific inquiries into cognition. Moreover, Merleau-Ponty’s criticism of intellectualism and empiricism has helped to legitimise the shift away from representationalist and cognitivist approaches and, more broadly, helped to locate Merleau-Ponty within the analytic philosophical landscape that has traditionally been the first port-of-call for cognitive science.

Evaluated in terms of its stated aims, which are to demonstrate that Merleau-Ponty's descriptions can have significance for cognitive scientists, this reception has definitely been a success. As a reading of Merleau-Ponty, however, one might be concerned about a certain domestication of his philosophy as a result of being put to this purpose.

Consider how the reception of Merleau-Ponty's phenomenology differs from Husserl's. At one level, Husserl is merely another source of phenomenological insight and/or data – grist for the cognitive science mill. At another level, though, there is a concerted effort to respond to Husserl's anti-naturalism. As I mentioned, the editors of the *Naturalizing Phenomenology* anthology explicitly claim that contemporary developments in mathematics, specifically in morphodynamics, mean that Husserl's arguments about the inappropriateness of a scientific treatment of the inexact lifeworld phenomena no longer applies. Contemporary mathematics is capable of modelling such informal essences.

There is no doubt that this is a controversial and disputed claim. Consider, for example, Dan Zahavi's comments in his review of *Naturalizing Phenomenology*. There Zahavi argues that the focus on the inexactitude of essences misses the point of Husserl's anti-naturalism. He claims that Husserl's commitment to anti-naturalism is based on a distinction between philosophy and natural science. That is, in giving an anti-naturalist interpretation of his descriptions, Husserl is simply defending their specifically philosophical significance.

From this perspective, the phenomenological reduction is then not simply a way of acquiring accurate first-person experiential data, as Varela suggests, but rather "the requirement always to abide by the sense of the proper investigation, and not to confuse epistemology with a natural scientific (objectivistic) investigation." (Husserl 1984, 410; quoted in Zahavi 2004, 337) In this way, the reduction is tied to a transcendental inquiry into "the conditions of possibility for meaning, truth, validity, and appearance" (Zahavi 2004, 339). Zahavi points out that proponents of the project of naturalisation may take heart from the distinction Husserl makes between transcendental phenomenology and phenomenological
psychology. The latter is a regional, eidetic inquiry into consciousness from a first-personal perspective that, unlike transcendental phenomenology, remains within the natural attitude.

One way to facilitate the naturalization of phenomenology is to abandon the transcendental dimension of phenomenology and to make do with a phenomenological psychology. If that is done, the likelihood of a success certainly looks more promising. ... The main problem with this way out, however, is that a good part of what makes phenomenology philosophically interesting is abandoned. Phenomenology is basically, I would insist, a transcendental philosophical endeavour, and to dismiss that part of it, is to retain something that only by equivocation can be called phenomenology... it is not phenomenology understood as a philosophical discipline, tradition, or method. (Zahavi 2004, 339-40)

My concern, though, is not really with evaluating this aspect of the reading of Husserl, so much as it is to point out that no similar question arises with respect to the reading of Merleau-Ponty among commentators on cognitive science. That is, no need is felt to naturalize Merleau-Ponty's phenomenology in this sense; the task is merely to appropriate it, and this can only be because it is presumed that Merleau-Ponty's phenomenology takes up no transcendental position comparable to Husserl's. Indeed, from Taylor one gets the impression that the opposite is the case, that Merleau-Ponty abandons any distinctively epistemological concerns when he abandons all forms of anti-realism.

Even in Dreyfus' work, where there is no attempt to naturalize Husserl, we find the transcendental aspect of Husserl's phenomenology compared with the greatest excesses of cognitivist psychology and artificial intelligence. Dreyfus attributes two claims to Husserl: firstly, that representational content is a constitutive condition of intentionality, and secondly, that all human understanding can be analyzed in terms of mental representations. (Dreyfus 1982, 8-9) The first claim, in Dreyfus' opinion, anticipates cognitive psychology, while the second anticipates good old-fashioned artificial intelligence. Dreyfus understands Husserl's epoché and noematic analysis as precursors to Fodor's methodological solipsism and representational theory of mind, which his

25 Questions concerning Merleau-Ponty's relationship to naturalism and to transcendental philosophy are raised by others, most notably Renaud Barbaras (1999; 2003; 2004), Maxine Sheets-Johnstone (1999) and Dan Zahavi (2002). Nonetheless, this has had little impact on the way in which Merleau-Ponty is presented in the literature I am interrogating here. The inclusion of papers by Barbaras into two notable collections emphasizing Merleau-Ponty's significance for cognitive science (Carman and Hansen 2003; Petitot et al. 1999) demonstrate a dim awareness of an alternative approach, but these only serve to reinforce the sense that there is no real understanding of how to incorporate Barbaras' insights into their own naturalistic readings.
transcendental reduction anticipates Minsky's attempt to construct a comprehensive database which would contain, in representational form, the contents and structure of common sense knowledge. (Dreyfus 1982, 19ff) This allows Dreyfus to establish an analogy between the rejection of cognitivism and the rejection of Husserl's transcendental reduction by his existential successors, especially Heidegger and Merleau-Ponty.26

As Heinämaa points out, this analogy is a powerful one, in that it "allows both [critics and enthusiasts] to formulate their views in terms of specific philosophical positions." (Heinämaa 1999, 50) Nonetheless, she argues that it provides a "simplistic" account of Husserl's relation to the existentialists. This prompts her to attempt to "show that Merleau-Ponty does not abandon Husserl's idea of phenomenology but rather reformulates or modifies this idea." (ibid.)

The reading of Merleau-Ponty for cognitive science is interesting, from the perspective of Merleau-Ponty scholarship, because it exposes a presumptive dismissal of any transcendental aspect of Merleau-Ponty's thought. I hope to show in the following chapter that this presumption extends far beyond the cognitive science reading, and is adopted for a variety of different reasons. Still, I would suggest that it is only when this presumption takes the form of an implicit naturalisation of Merleau-Ponty's philosophy that one actually balks at the consequences. In this way, the "mundane" reading I've just outlined has a rhetorical or dialectical significance in that it prompts a reconsideration of Merleau-Ponty's relation to transcendental philosophy, by bringing that aspect of his thought into negative relief.

Ironically, Dreyfus himself has recently begun to acknowledge that Merleau-Ponty might not be comfortable with the naturalisation of his philosophy as cognitive science. Consider the following passage:

> It would be satisfying to think that Merleau-Ponty would happily embrace some such [connectionist] model, but there are passages in *Phenomenology of Perception* in which Merleau-Ponty seems to foreclose the possibility of *any* account of brain function that could in any way be the basis of motor intentionality. He states categorically, "How significance and intensionality could come to dwell in molecular edifices or

---

26 It is interesting to note that, as early as the first paragraph of his introduction to the same article, Dreyfus himself dismisses the possibility of providing "a clear explanation in non-Husserlian terms of what the reduction is, what it reveals, and why according to Husserl one must perform it in order to do philosophy." (Dreyfus 1982, 1)
masses of cells is something which can never be made comprehensible, and here Cartesianism is right” [PhP, 351]

One would think that it is an empirical question whether and how brain activity underlies motor intentionality and that the conviction that a naturalized account must be possible (as Koffka and John Searle, for example, maintain) or that it is inconceivable (as Merleau-Ponty contends in the preceding passage …) both go beyond what we have a right to claim. (Dreyfus 2005, 144)

One could presume that it is an empirical question, only if one had already ruled out a transcendental account of motor intentionality, i.e. if one had presumed that Merleau-Ponty's conception of motor intentionality is not presented from within a transcendental reduction.