Roly Perera

Cartesian Creatures

Watching Ourselves Watching the World

Abstract: *I propose that a scientific account of perceptual consciousness decomposes into two sub-problems: the conceptual problem of reframing our internal, first-personal perspective in external, third-personal terms, and the scientific one of providing a theory that accounts for the phenomenon reframed in this way. In this paper I consider both aspects of the problem, drawing on the ideas of Dennett and Metzinger. For the first part, I use Dennett's method of heterophenomenology to argue that perceptual experience should be understood as a structured space of possible behaviours. On this view, each phenomenal 'detail' that we consciously apprehend is not quintessentially private, but rather a reflexive affordance: an opportunity to behave, perhaps covertly, about how the world looks, sounds, or feels from our perceptual vantage point. I then present a novel argument in favour of a higher-order, self-representationalist account of this aspect of our phenotype, along the lines of Metzinger's phenomenal model of the intentionality relation (PMIR). I conclude with the suggestion that a conscious organism has a design which is distinctly Cartesian in flavour: a self-simulating agent able to observe itself perceiving.*

1. Introduction

Self-representationalist proposals such as Metzinger's *phenomenal model of the intentionality relation* (Metzinger, 2003; 2005) explain perceptual consciousness as a kind of higher-order, reflexive intentionality: the ability of an organism to represent its own intentional

Correspondence:
Roly Perera, The Alan Turing Institute, 96 Euston Rd, Kings Cross, London NW1 2DB. *Email: rperera@turing.ac.uk*

Journal of Consciousness Studies, **26**, No. 3–4, 2019, pp. 131–54

relationship to the world. I present a novel argument in favour of such an approach, borrowing a notion of intentionality from Dennett's (2015) 'ontology of the organism': the idea that every organism, through morphology and behaviour, embodies a theory or model of its environment. This will provide a biologically grounded way of understanding organisms and their activities as being 'about', or representing, certain features of the world.

Organisms and their phenotypes are an ongoing response to other organisms (through inferential mechanisms like evolution, learning, and perception), so biological intentionality is often higher-order. Metzinger's proposal explains subjectivity as a special reflexive case of higher-order intentionality — *self-intentionality* or *self-representation* — where the organism not only represents features of the environment, but also represents itself in the act of representing those features.

Metzinger's proposal is compatible with enactivist and sensorimotor accounts of perception. Whereas those accounts usually assume that active engagement with the environment is sufficient to 'bring forth a world' (Varela, Thompson and Rosch, 1992, p. 206) for the organism, in this paper I make the case that this can only arise through reflexivity: only by treating *yourself and your own active engagement with the environment* as part of the environment with which you are actively engaged can you come to an appreciation of yourself as inhabiting a world. (Metzinger's own position on the necessity of a PMIR for perceptual consciousness is unclear; since my goal here is not exegesis, this need not concern us.) Moreover, I will claim that an organism that is reflexive (and thereby perceptually conscious) in this kind of way is a 'Cartesian creature' that is able to observe itself perceiving, although not in a way that is problematic for the usual reasons that 'inner observation' is considered problematic.

1.1. Overview of paper

I start in §2 by getting clearer on what a scientific theory of perceptual consciousness should actually explain. I identify a class of highly contentful behaviours that have as their ontology or subject matter *what it is like to be me*, perceptually speaking: the way the world looks, sounds, and feels from my perceptual vantage point. It is these 'first-personal' behaviours that I shall then suggest a theory of perceptual consciousness should be trying to explain.

As a methodological move, this has the significant advantage of fully locating the explanandum in the natural world. On the other hand the move is warranted only if we can convince ourselves that an account of these behaviours has the potential to 'save the phenomena', in particular what we consider to be the *subjective* phenomena: if both the fact that there *is* something it is like to be me (perceptually speaking) in the first place, and then what being me is actually *like* can be recast somehow in terms of my ability to *behave about* what it is like to be me. But this seems quite implausible: the content of subjective experience is a vivid first-personal world, and moreover a first-personal world that I seem to 'have' rather than merely behave about.

Two further considerations may help make this more plausible. First, and *contra* Dennett, I argue that first-personal behaviours have as their ontology or subject matter a three-dimensional, egocentric, multimodal *image* of how the world appears from my point of view. Second, I suggest these behaviours also interpret *me* as the owner of that first-personal world, so that the only version of 'me' that I have any expertise in — that I am able to behave about, either overtly or covertly — is an agent whom I take to be having perceptual experiences.

This analysis of the explanandum, then, is potentially controversial both to those sceptical of Dennett's deflationary position, *and* to those, like Dennett, determined to avoid anything that might resemble a Cartesian theatre. Regardless, it is assumed for the rest of the paper, which argues for the essential role of self-modelling or self-representation in accounting for this distinctive aspect of our phenotype. The argument also assumes the (uncontroversial) idea of perception as inference (Helmholtz, 1867; Friston, Kilner and Harrison, 2006).

First I will consider in §3.1 what perception contributes to the ontology of the organism from the perception as inference perspective, and argue that what it contributes is *world* knowledge: for example the egocentric location of prey. While the organism and its perceptual vantage point form the *inferential basis* for this worldly knowledge — the raw data from which the world model is inferred — they are not *themselves* things the organism comes to know about as a consequence merely of being a perceiver. Perception alone therefore cannot account for phenomenology as characterized in §2 simply because it fails to deliver the relevant kind of knowledge.

Then I will suggest in §3.2 that self-representation *can* deliver the relevant kind of knowledge: by tying a recursive knot in the theory of the world embodied by the organism, self-representation allows the

organism to come to know about aspects of itself, namely that it has a perceptual vantage point from which it is inferring a world model. Thus it is only self-representation which has the potential to 'bring forth a world' for the organism in the sense of equipping it with an appreciation of *itself* situated in the world.

I conclude in §4 by considering how self-representation gives rise to an 'audience': an agent which can not only perceive in the non-conscious sense set out in §3.1, but one which is able to 'observe' itself perceiving, a functional organization which is distinctly Cartesian in flavour. I close with the suggestion that perhaps the reason 'the persuasive imagery of the Cartesian Theater keeps coming back to haunt us — laypeople and scientists alike' (Dennett, 1991, p. 107) is that that is roughly how we work.

2. Perceptual Phenomenology as Reflexivity

We are intimately familiar with perceptual experiences: I'm having them right now, as I sit at my desk in front of my computer, typing. But if we want to explain them using science, we have no choice but to somehow locate the things we are talking about in the world *observable* to science. This is not a case of identifying things in the world that experiences might (mysteriously) 'correlate' with (Chalmers, 1998), but rather of finding a way of *reconceptualizing* experiences — convincingly — as something taking place in the observable world, and therefore amenable to scientific explanation.

To be convincing, we must do this in a way that is compatible with our intuitions about what perceptual experience is *like*: a colourful 'movie in the brain' (Damasio, 1999), unobservable from the outside. So in this sense the scientific understanding of perceptual consciousness partitions into two sub-problems: the conceptual problem (which we could call 'naturalizing the explanandum') of reframing our internal, first-personal perspective in external, third-personal terms (Stalnaker, 2008), and the scientific one of providing a theory that accounts for the explanandum.

I will return to the scientific problem in §3; in this section I make a stab at the conceptual problem, using Dennett's (2003) abductive method of *heterophenomenology*. However, I will adjust Dennett's story in one important respect. The Dennettian part will be the claim that perceptual experiences can be understood as the *subject matter* of a particular kind of *behaviour*. These are the highly contentful, mostly covert behaviours that are 'about', or constitute expertise in, *what it is*

like to be me, perceptually speaking: the way the world looks, sounds, and feels from my perceptual vantage point. The adjustment will be to insist, *contra* Dennett, that perceptual experiences (so understood) amount to a vivid, egocentric *image* of the world that we are able to 'observe', and that in fact conscious perception consists precisely in this ability not just to perceive but to apprehend an image of what one is perceiving.

2.1. The ontology of the organism

Let's start by getting clear on how behaviour can be intentional, or have a 'subject matter'. I will appeal here to what Dennett (2015) calls the *ontology of the organism*. This is the familiar biological idea that each organism, through its various adaptions and learned behaviours, embodies a *domain of expertise* or model of the world: the 'set of things in the world that matter to it and which it (therefore) needs to discriminate and anticipate as best it can' (*ibid.*, p. 5). Theoretical accounts of perception and evolution take this kind of view, construing organisms as models of their environments (Friston *et al.*, 2012; Bruineberg *et al.*, 2018).

To identify the domain of expertise of the organism — its ontology — is thus to adopt a 'design stance' (Dennett, 1987) towards the organism, interpreting its phenotype as targeted towards, or *about*, specific features of the world. The ontology of an echolocating bat includes things like the egocentric location of bodies of water and insect prey, because these are the things that we, as theorists interested in what bats do, conclude that bat morphology and behaviour is optimized for and that bats seem mostly concerned with.

Adopting this sort of stance also means interpreting the behaviour of the organism as purposive and driven by rationales. We say things like: the bat swerved *to avoid the cave wall*, the bat was *pursuing its prey*, the bat drank *because it was thirsty*, and suchlike. Sometimes we may be tempted to interpret these attributions of reasons and other 'attitudes' to the bat in a phenomenal way, supposing that *to the bat*, subjectively, it is drinking because it is thirsty. But on Dennett's view this is a mistake. Dennett's key observation is that for an organism to be 'aware' of its own agency — for it to have reasons for acting which it might take to be *its* reasons — is more than just a matter of acting in *accordance* with rationales. Acting for your *own* reasons involves being able to act in complex ways about *yourself as a haver of reasons* (Dennett, 2004): the model of the world that you, as an

organism, embody must be *reflexive* enough for you to become your own intentional object.

So unless the behavioural repertoire of bats enjoys this kind of reflexivity, there is no reason (on Dennett's view) for supposing them able to interpret themselves as agents, acting for reasons. The rationales that *we* might posit for their actions are 'free-floating' (Dennett, 2004; 2009): part of the causal language bat-theorists use to explain bat behaviour, but not themselves things it would appear that bats are equipped to know about.

This distinction — between subjective and 'free-floating' rationales — also highlights the difference between the ontology of the organism and von Uexküll's (1957) related concept of *Umwelt*. For von Uexküll the *Umwelt* is an inherently phenomenal or first-personal notion: the world as it 'appears' to the organism. For Dennett the ontology of the organism picks out what it makes third-personal sense to suppose the organism is concerned with, not what (if anything) might 'matter' to the organism in a way that the organism itself is able to appreciate.

So the ontology of the organism provides a straightforward sense in which behaviour is intentional. Building on ideas of Dennett and Metzinger, I will now develop the claim that perceptual phenomenology can also be understood as a kind of reflexive intentionality: an organism's ability to behave about *how the world appears to it*. What is Dennettian (but will require some non-Dennettian work to make plausible) about the claim is that it proposes that there are no facts about how the world 'really' appears to the organism beyond its ability to exhibit these sorts of behaviours.

2.2. The ontology of perceptual experiences

Perceptual judgments, like rationales, often feature in causal explanations of an organism's behaviour. For example, many species of bat are susceptible to a perceptual illusion whereby they interpret any echoacoustically smooth horizontal surface as a body of water; it is easy to trick such a bat into attempting to drink from a smooth plate (Greif and Siemers, 2010). In such a situation we might say that *it seems* to the bat that the smooth surface is a body of water.

As with rationales, we might be tempted to interpret these perceptual attributions in a phenomenal way, as the supposition that *to the bat*, subjectively, the smooth surface 'appears' how a body of water appears. We might be curious about the 'phenomenal character' of echolocation (Nagel, 1974), given that it resembles audition in

some respects and vision in others. But analogously to Dennett's line of reasoning about rationales, I shall defend the view that such a supposition of phenomenality is warranted only if the bat is able to exhibit a certain kind of reflexive behaviour.

The key claim is that perceptual experiential content always has consequences for behaviour. For things to look (or sound, or feel) a certain way is precisely for you to be able to act in complex but highly structured ways *about* how they look (or sound, or feel). On this view, unless our bat's behavioural repertoire can be construed as extending to behaving about *how the smooth surface appears to it*, the appearance of the surface to the bat remains something that we (as bat-theorizers) may appeal to in order to explain and predict its behaviour, not something we have any reason for thinking the bat itself is equipped to know about.

To briefly illustrate the idea of perceptual phenomenology as a structured space of possible behaviours, consider the detailed ways that I am able to behave right now about perspectival matters like apparent shape and apparent colour. I have fine-grained 'expertise' — a rich set of opportunities to behave — about the details of specular and diffuse reflections from my desk lamp. I can behave, also in a variety of revealing ways, about how the colours of the window blinds are muted when reflected in my desk; how as I move my head, nearer things appear to move in specific ways relative to things in the background; and so forth.

The suggestion, then, is that to characterize perceptual experiences as something taking place in the observable world (and therefore amenable to scientific explanation) we should make the following naturalizing move: we should interpret each phenomenal 'detail' that we ascertain, not as something quintessentially private, but as a *reflexive affordance*: an opportunity to behave about *how things appear to us*.

A radical consequence of this move is that there are no details of perceptual 'experience' that lack behavioural consequences. To come to a *conscious* appreciation of something is to acquire the potential to behave one way rather than another in the future in relation to that conscious content: to be inclined to move a paintbrush a certain way, to press a button now rather than later, to have a slight preference for one word over another from a lexicon or for one colour over another from a palette. To make this more plausible, I will now flesh out the basic idea with two complementary claims: first (§2.2.1), that the notional 'subjective world' that is the subject matter of my first-

personal behaviours is (*contra* Dennett) a three-dimensional, first-personal, multimodal image, and that this is easily verified by 'sampling' my capacity for such behaviours; second (§2.2.2) that, more precisely, it is *me as the owner of such a subjective world* that is the subject matter of these behaviours, and it is this ability to behave about myself *as* a subject that is constitutive of my *being* one. I will try to show how these observations help close the gap between first-personal behaviours and how we tend to think of our own phenomenology.

2.2.1. Imagistic behaviour

The first claim will be that first-personal behaviours are highly 'imagistic', in the sense of having a vivid three-dimensional, egocentric image as their subject matter. The suggestion is not that such an image exists in our heads (whatever one might construe that to mean), but rather that such an image is what these behaviours are *about* — that they can be considered to constitute *expertise in*, or to enact a *theory of*.

This is an empirical claim, and one that we could test by 'sampling' a subject's potential for first-personal behaviour, interpreting self-reports, and any other behaviours that we deem relevant, as *data* best explained by supposing that the subject is having perceptual experiences of some kind: in other words by positing an *ontology of subjective experiences* as a way of making sense of those behaviours. This is roughly the abductive method that Dennett (2003) calls *heterophenomenology*. Here we will just indulge a short heterophenomenological thought experiment, but one that will hopefully lend credibility to the claim that our first-personal behaviours are indeed imagistic.

Suppose a normally sighted person were sitting at my desk, like I am. What would their visual experiences have to be *like* in order to be broadly consistent with their potential for outward behaviour? Well, one structural feature it is easy to infer is *egocentricity*: the way the apparent size and shape of things, as our subject might report verbally or attempt to sketch on paper, is determined by the geometry of first-person perspective (Evans, 1982). It is also apparent, purely on the basis of what they are inclined to say and do, that they are able to deliver more fine-grained information about the appearance of nearby things than further away things. The information they have about my computer screen 'occludes' information about things behind the screen. These are all facts about their behaviour.

Moreover, how they are able to act about perspective and occlusion has a distinctive sensorimotor profile (O'Regan and Noë, 2002): we can experimentally verify that, when they move their head, what they (ostensibly) consider to be the apparent shape of the screen — as disclosed, intentionally or otherwise, by their behaviour — changes in a certain predictable way. Some of their appearance-related knowledge which was previously 'occluded' is 'revealed' and vice versa.

We could also determine that our subject's expertise in apparent colour is similarly perspectival (Noë, 2004, chapter 4): because of incident lighting and shadows, their behavioural repertoire relating to the appearance of the upper surface of my monitor, although it is made of the same material as the front and side-facing surfaces, differs from how they are inclined to behave about the appearance of the other surfaces. As with perspective and occlusion those behavioural profiles are transformed in certain reliable ways as they move their head.

We could easily continue in this vein. The take-away point is this: while these are all public capabilities of our subject, and therefore observable to the third-personal world of science, there is also a clear sense in which (regardless of the 'actual' phenomenology of the subject) they *concern* a first-personal perspective. They constitute rich, dynamic expertise in how the world looks from the subject's perceptual vantage point, expertise which is transformed in systematic ways as that vantage point shifts and as the world itself changes.

So it seems clear enough that *the way the world looks to us*, as a heterophenomenological posit, has the dynamics of an egocentric image. But is this hypothetical first-personal 'world' actually an image in the sense of having the same sort of content as a movie or first-person computer game? I think we can show that it is by sampling the subject's capacity for first-personal behaviour more exhaustively, for example by setting them the task of painting a picture of how the world looks. (Consider Ernst Mach's, 1914, famous *View from the Left Eye* in Figure 1.)

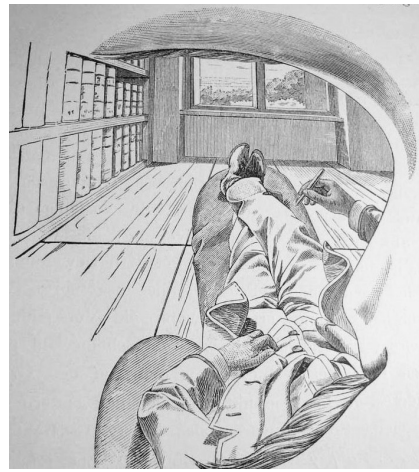

Figure 1. View from the Left Eye by Ernst Mach (1914).

If we interpret behavioural outputs such as these as heterophenomenological data, it seems clear that our competences regarding perspectival matters like foreshortening and apparent luminance do indeed have a *bona fide* image as their collective subject matter. It is not an image which resides in our heads (or indeed anywhere else), but rather an image that we can be considered to 'know about', or have 'access to', in virtue of what we are able to do.

What these thought experiments suggest is that, by considering only someone's capacity for first-personal behaviours, we can conclude — in a third-person verifiable and therefore scientifically intelligible sense — that the intentional object of those behaviours is an egocentric image. Moreover, the more systematically we sample the behaviours, the more closely the inferred first-personal world resembles what we consider our own phenomenology to be like. (If you doubt this, look carefully at John Baeder's photorealistic painting in Figure 2 and then try to interpret it as heterophenomenological data.)

Figure 2. John's Diner with John's Chevelle by John Baeder (2007).

This conclusion moves us towards our 'naturalizing the explanandum' goal by lending credibility to the Dennettian idea that the content that one can *objectively infer* from our capacity for outward behaviour converges, in the limit, on what we *subjectively take* to be the content of perceptual experience (Dennett, 2005, chapter 2). We haven't yet considered how the ability to behave *about* a first-personal world could be the basis for *having* one, but it at least shows that the two things are plausibly aligned in terms of content.

Ironically, perhaps, this conclusion lends credibility to one Dennettian perspective by undermining another. Dennett regards any imagistic talk in relation to perceptual experience as a misguided legacy of Cartesian thinking: not just the uncontroversially problematic idea of an image of the world somehow located in our heads, but anything that plays the functional role of such an image (Dennett, 1991, chapter 11). For example, he argues that there is no recourse to images when it comes to explaining how people see, because perception evolved to produce *behaviour*, and behaviour itself is not 'imagistic':

> The eventual 'products' of vision are such things as guided hand and finger motions, involuntary ducking, exclamations of surprise, triggering of ancient memories, sexual arousal,… and none of these is imagistic in any sense… (Dennett, 2002, p. 189)

But this requires Dennett to take an overly selective view of the behavioural 'products' of vision, restricting them to low-level actions like finger movements and exclamations of surprise, and ignoring the most distinctive things that humans do with vision.

We didn't identify the distinctive ontology of the bat by considering low-level behaviours like wing movements. Instead we considered the activities that bats, at the organismic level, seem to be optimized for and to spend most of their time concerned with: drinking water, pursuing prey, and so forth. Similarly, the question of whether our ontology involves images can only be settled by considering what people spend most of *their* time concerned with.

But when do that, as I have argued here, we find ourselves in the company of behaviours which can only be made sense of by supposing that people have rich, imagistic knowledge of how the world looks. Thus Dennett's own heterophenomenological method suggests that our imagistic intuitions about visual consciousness are at the core of the phenomenon to be explained, and should not be dismissed as a theoretical confusion.

2.2.2. Naturalizing subjectivity

Recall our 'naturalizing the explanandum' goal from §2.2. The aim was to characterize perceptual experiences as a capacity for *first-personal* behaviours, highly structured behaviours which concern *how things appear to me*. What I have argued so far is that 'how things appear to me' — the first-personal world implicit in these behaviours — is (*contra* Dennett) a rich egocentric image. This is progress inasmuch as this behavioural content seems to agree, as we probe it in increasing detail, with what we intuitively take to be the content of perceptual experience.

What is less clear is how we might turn this into a characterization of experience itself. After all I think of myself as *having* a first-personal world, rather than just being able to behave about one. (Metzinger, 2007, calls this crucial 'proprietary' aspect of consciousness *mineness*.) To make more progress, we must tweak our understanding of what we have been calling first-personal behaviours, considering behaviours which are not so much about *my subjective world* but more specifically about *me as the owner of a subjective world*. This is a subtle shift in emphasis that incorporates my tendency to think of myself as having subjective experiences as itself part of my repertoire of first-personal behaviours. In turn this 'naturalizes' subjectivity because we no longer need to explain *experiences*

(whatever they might be) but only my conviction that I am having them.

To unpack this a little from the perspective of the ontology of the organism: behaviour embodies or enacts a theory of the world; first-personal behaviours embody or enact a theory of me-in-the-world; and the claim then is that, *in the ontology of such a theory*, I am a creature with perceptual experiences. That is what I take myself to be — where we must understand 'take myself' not in any cognitive sense but rather as a rich set of dispositions to behave in certain reflexive ways.

There is, then, in a certain sense, a 'me' who *is* having perceptual experiences: the notional subject that is the intentional object of my first-personal behaviours. The perceptual experiences of this notional subject — 'me' — are something I have a huge amount of sub-personally held expertise in, as §2.2.1 showed can easily be demonstrated. Things that I say or do that have *me* as their subject matter — including verbally expressed convictions but also moments of reflection or introspection and other covert behaviours — will therefore be consistent with my *really* having perceptual experiences, because that is precisely what I take myself to be having.

So our attempt at 'naturalizing the explanandum' was initially unsatisfying because it appeared to change the subject from actual experience to merely behaving about experience. Most likely I will retort by insisting that I *really am* having certain experiences; I'm having them *right now* so I can be absolutely certain about that. But you can in turn point out that that kind of conviction is part of the phenomenon to be explained, not a privileged form of first-personal evidence.

This is not to say that I am necessarily *wrong* when I suppose myself to be having experiences; only that the contrast that the word 'really' is reaching for (between having experiences and merely supposing oneself to be having them) might not be quite as substantive as it first seems. It might turn out (although this is a topic for another paper) that one way to genuinely have experiences, perhaps the only way worth wanting (*cf.* Dennett, 1984), is simply to suppose oneself to be having them.

2.2.3. Summary

We often (albeit implicitly) take the explanandum of a theory of perceptual consciousness to simply be the 'first-personal fact' that we are having perceptual experiences. This is taken to be obvious and

undeniable, yet lacks a well-defined third-personal meaning. Trying to explain it using science is therefore a highly problematic endeavour.

What I have motivated here is the idea that we can make these supposed first-personal data amenable to scientific explanation by reconceptualizing them as third-personal data: as our capacity for highly contentful, reflexive behaviours that have as their intentional object a version of 'me' equipped with a subjective world. The proposal has the potential to work as a characterization of experience, counter-intuitive though it may be, because the notional 'me' — the 'me' in the theory of me-in-the-world that I embody, for whom the potted plant on my desk looks a certain way — is what all my first-personal behaviours, even introspective ones, have as their subject matter.

This analysis is potentially controversial *both* to those sceptical of Dennett's deflationary approach (since it denies that there are first-personal data that cannot be reduced to third-personal data), *and* to those, like Dennett, wary of anything that resembles a Cartesian theatre (since it embraces rather than denies an imagistic notion of perceptual experience). It is, however, less open to the charge of 'denying the undeniable' (Levine, 1994) than the usual Dennettian story. It allows that there *are* such things as 'feels', 'smells', and 'looks'; they are facts about *ourselves* that we can robustly and reliably behave about. Dennett himself moves towards such a position in recent work (2013; 2015); see Clark (2016) for discussion.

At any rate, for the rest of the paper I will assume this naturalistic characterization of the phenomenon of perceptual experience. I now argue that it can be explained only by supposing that the organism embodies something like Metzinger's (2003; 2005) *phenomenal model of the intentionality relation* (PMIR), a theory of itself as a locus of perceptual activity.

3. Perceptual Phenomenology through Self-Modelling

Although much of our perception is unconscious (Adams, 1957; Hassin, Uleman and Bargh, 2005), as conscious creatures we find it easy to conflate perception with perceptual experience. *I* know how things look just by looking at them, how things sound just by listening to them, and so forth, so it is natural to suppose that this is somehow (if mysteriously) what perception is all about, or even *for* (Akins, 2002, p. 216). This intuition surfaces even in modern theories of perception, such as the enactivist idea that perceptual engagement

with the environment suffices to 'bring forth a [phenomenal] world' for the organism (Varela, Thompson and Rosch, 1992).

The point of §2 was to argue that we should understand perceptual experience as a highly structured set of reflexive *capabilities*: the ability to behave, both overtly and covertly, about a rich subjective world. In §3.1, I will argue that merely being a perceiver is insufficient to explain these capabilities, at least on the usual view of perception as inference about the world from the dynamics of sensory data (Helmholtz, 1867; Friston, Kilner and Harrison, 2006; O'Regan and Noë, 2002). Instead, as I shall argue in §3.2, these capabilities only come about if the organism, as a locus of inferential activity, becomes the object of its own inferential processes, and thus able to interact internally with its own mechanisms for interacting with the world. Recent proposals along these lines include Metzinger (2003; 2005), Cleeremans (2011; this issue), Dennett (2013; 2015), and Hohwy and Michael (2017).

On this kind of view, it is possible to be a perceiver without being a phenomenal perceiver simply because it is possible to have all kinds of perceptual competences without having any 'metacompetences' whatsoever in *yourself* as a locus of perceptual competences. If you lack the latter, then you lack a phenomenal world, however rich your interactions with the environment. I conclude §3.2 by identifying this view as a particular (reflexive) species of *higher-order representationalism* (Armstrong, 1981; Rosenthal, 1986; Lycan, 1996) which offers a novel account of how (reflexive) metarepresentation gives rise to subjectivity.

3.1. Perception without phenomenology

To see why perception alone is insufficient for perceptual experience, we need to consider the domain of expertise that perception imparts to the organism: the kind of subject matter, or ontology, of the behaviours that it enables. I will assume the broadly Helmholzian idea of perception as the inferential process by which the organism comes to embody a model of the contingent features of the world which 'matter' to it (Helmholtz, 1867; Friston, Kilner and Harrison, 2006). According to this picture, the inputs to perception are perspectival, observer-relative sensory information, or *sensorimotor contingencies* (O'Regan and Noë, 2002); the outputs are changes to the internal organization of the organism enabling behaviour which (from the perspective of a third party) can be interpreted as being 'about'

various features of the world. Perception and action are thus dual mechanisms by which the organism maintains an intentional relationship with its environment: avoiding predators and pursuing prey, moving away from unfavourable conditions and towards favourable ones, and so on.

To make the point about perceptual experience I will use the example of binocular vision, a mode of inference about the world based on various sensorimotor contingencies particular to light and (pairs of) eyes: motion parallax (Rogers and Graham, 1979), binocular disparity, foreshortening, reflectance profiles of surfaces, and suchlike. Although vision relies heavily on perspectival data, the *ontology* or subject matter of the inferred model — the domain of expertise that binocular vision imparts to the organism — is rather different: information about the actual size and shape of things, their egocentric location, and so forth. This worldly expertise consists in the organism's ability to navigate through a forest without bumping into trees, successfully grasp food, recognize predators, and so forth.

So while vision is a complex process of inference from perspectival data particular to the vantage point of the animal, what the organism acquires expertise in as a consequence are various properties of the world. Crucially, no opportunities to behave 'reflexively' about itself as the owner of a perceptual vantage point, or about the inferential processes involved in vision, are afforded to the organism *by vision itself*: that would require information about the *visual process* to be available to the organism. But there is no mechanism by which an organism has automatic 'knowledge' of — in the sense of being able to behave in reliable ways about — the goings on of arbitrary processes in its brain.

A mere visual perceiver, therefore, however rich its world model, lacks the skills to have any appreciation of itself as an agent engaged in vision. That sort, or indeed any sort, of self-knowledge is simply not part of the domain of expertise imparted to the organism by vision *simpliciter*, at least on the perception as inference view. Therefore there is no phenomenology that comes along for the ride with vision.

What of our echolocating bat? To many people, a phenomenology of echolocation is intuitively plausible, driven presumably by analogy with our own vision. But, as with vision, it no longer looks possible that any phenomenology automatically accompanies echolocation once we consider the domain of expertise that it imparts to the organism. Echolocation is an inference *from* the dynamics of motion parallax, Doppler shift, and other sensory data (Pollak and Casseday,

1989) *to* features of the world relevant to the bat: things like the size and egocentric location of bodies of water and flying insects. So although it makes *extensive use* of perspectival data particular to the bat's vantage point, echolocation in itself provides no mechanism for the bat to come to 'know' about such things.

The point common to both of these examples is that there is no automatic means by which information about an inferential process going on in the brain becomes expertise available to the organism: no brain process can itself be the mechanism by which information *about that process* becomes available for other purposes. Thus your bat brain may make inferences from apparent shape to the flight path of insect prey, exploiting patterns in that data to fine-tune your pursuit; but in the absence of further mechanisms by which *those inferential processes themselves* can be incorporated into your domain of expertise, in addition to your 'business as usual' activities of tracking prey, there are no means by which you can come to know anything about apparent shape. You simply have no opinions about how things look.

None of this is to say that bats categorically lack a phenomenology of echolocation; only that merely being an echolocator is not enough to equip you with that capability.

3.2. Perceptual self-modelling

Perception *simpliciter*, then, fails to equip the organism with perceptual experiences simply because perception alone can only account for 'worldly' expertise, not the rich appreciation that a conscious creature has of itself as the owner of a perceptual vantage point. O'Regan and Noë gesture to something else being required beyond mere perceptual competence when they suggest that to perceive consciously is to deploy one's mastery of sensorimotor contingencies in the service of 'planning, reasoning and speech behaviour' (2002, p. 944). But no reasons are offered for why using one's 'mastery' in one way rather than another should be special.

Metzinger's *phenomenal model of the intentionality relation* (PMIR) does offer a basis for distinguishing two senses of 'mastery', and I will now argue that these are central to understanding the extra faculties that a conscious perceiver enjoys over a mere perceiver, namely the *first-personal* behaviours that we considered in some detail in §2.2.1. I focused there on two core aspects of perceptual phenomenology. The first of these was our 'knowledge' of a first-

personal world, the perspectival information that we can acquire just by looking around; the second was the property that Metzinger (2007) calls *mineness*, our tendency to behave as though this egocentric image were *our* subjective world. So these are our explananda: how we come to have 'access' to a vivid, egocentric image that transforms according to the geometry of first-person perspective as we shift our vantage point, and then how we come to attribute such first-personal viewpoints to *ourselves*.

Metzinger's proposal is that the organism can (over evolutionary time) come to acquire these capabilities by directing its inferential mechanisms at itself, and, in a higher-order way, at its own perceptual mechanisms. This allows the organism's own perceptual situation — its 'intentionality relation' — to become something the organism has rich expertise in. Rather than merely representing (or modelling) the world, the organism also 'co-represents the *representation relation* itself' (Metzinger, 2005, p. 28).

What modelling one's own 'intentionality relation' might actually involve is left mostly unspecified, so here is how I propose we understand it in terms of perception as inference: as a (higher-order) model of the inferential processes that constitute the organism's mastery of the relevant sensorimotor contingencies, a perceptual metamodel or metarepresentation of the vantage point of the organism. Such a metamodel would be an example of what Karmiloff-Smith (1990) calls 'representational redescription': a *procedure* of the brain — in this case our mastery of the sensorimotor contingencies — being treated as *data* by some other (higher-order) brain process that operates on it, allowing information about the first process to be made available for other purposes (Cleeremans, this issue).

This is a crucial — necessary although insufficient — building block of an account of perceptual phenomenology because it explains a key aspect of the shift in *subject matter* that we need to go from mere perception as characterized in §3.1 to conscious perception as characterized in §2.2: our knowledge of an egocentric world-image. What I argued in §3.1 is that an organism by default has what we could call *tacit* mastery of the sensorimotor contingencies: it can make competent use of the appropriate perspectival information in the service of behaving about worldly things, but is unable to behave *about* that perspectival information itself. All organisms enjoy this kind of (non-conscious) perceptual competence. On the other hand something like a PMIR equips the organism with what we could call *explicit* mastery of the sensorimotor contingencies: rich higher-order

competences regarding its own first-order mastery, allowing it to behave about the inferential basis of its own perceptual judgments. Then perspectival properties like apparent shape and apparent size, which are part of the inferential basis of vision but not what vision is intrinsically *about*, can be incorporated into the domain of expertise of the organism, enriching its behavioural repertoire with expertise in a first-personal world.

To see that this kind of metarepresentation is necessary for perceptual experience, consider *what it would be like* if you were to lack perceptual meta-expertise of this kind. I suggest that you would simply lack the associated phenomenology, since there are no prospects — in the sense of a plausible mechanism — for you to even introspectively make the 'observation', or mental note, that the screen in front of you has a certain (apparent) shape or size, if those facts are never the subject matter of any subpersonal inferences you make. You may quite competently be able to *make use* of foreshortening and motion parallax in inferring the location and orientation of the screen, but it would be a form of blindsight: you would lack the further faculty of being able to *appreciate* (in the sense of behaving reliably and directly about) the foreshortening or parallax itself.

The PMIR proposal, in involving meta-inference or metarepresentation in an essential way, is a variety of *higher-order representationalism* in philosophy of mind (Armstrong, 1981; Rosenthal, 1986; Lycan, 1996). However, it is important to point out that the suggestion is *not* that higher-order representation is sufficient for consciousness, i.e. that merely in virtue of being the object of other inferences certain perceptual inferences 'become conscious'. The *ad hoc* nature of such a supposition is captured by Goldman's famous objection to higher-order theories, which Stubenberg (1998) dubbed 'the problem of the rock':

> How could possession of a meta-state confer subjectivity or feeling on a lower-order state that did not otherwise possess it? Why would being an intentional object or referent of a meta-state confer consciousness on a first-order state? A rock does not become conscious when someone has a belief about it. Why should a first-order psychological state become conscious simply by having a belief about it? (Goldman, 1993, p. 366)

Rather, the suggestion here is that only a particular *reflexive* kind of metarepresentation, namely *self*-representation, is sufficient to explain perceptual consciousness. Perceptual metarepresentation only accounts for the first of our explananda: how we come to have rich knowledge of a first-personal world. This perspectival, imagistic

knowledge is explained by the supposition that the organism internally models its own perceptual vantage point, equipping it with a notion of *how things look from where it is*. But to account for the second explanandum — why we unquestioningly take that egocentric world-image to be *our* subjective world — we have to refine this notion slightly, and suppose (following Metzinger) that the organism models *itself* as the owner of such a vantage point. Only self-representation, not metarepresentation more generally, is able to explain how we attribute a subjective world to *ourselves*, because only self-representation is able to deliver behaviours that have the *organism* as their subject matter, and thus provide a way to understand consciousness as 'a property I have by virtue of my brain's attributing it to me' (McDermott, 1992, p. 217). And so our friend the echolocating bat could in principle have some kind of perceptual experiences *if* it were to model *itself* as the owner of a first-personal perspective. Some of the perspectival sensorimotor contingencies involved in echolocation, such as occlusion and apparent shape, resemble those of vision; and if the bat, rather than merely putting that perspectival data to use in the service of chasing insects, were to internally model *its use of that information*, then it would in principle be able to acquire expertise in the sorts of things that we have expertise in through visual phenomenology: how the apparent shape of an insect's wing changes as its orientation relative to the bat changes, how nearby prey appear larger and to move faster than more remote targets, and so on. Unless it can model its own perceptual vantage point in that kind of way, and moreover model *itself* as the owner of that vantage point, there are no prospects for a *phenomenology* of echolocation, since the bat is simply bereft of the relevant reflexive meta-competences.

4. Cartesian Creatures

Materialists since Ryle (1949) have objected to the idea of perception involving the observation of a private image: the so-called 'Cartesian theatre' (Dennett, 1991). And it is not hard to see why explaining vision as a kind of 'inner vision' is likely to be problematic. Yet I want to conclude that a perceptually conscious organism is a 'Cartesian creature', one able to see the world through an image.

A more nuanced objection than the usual worry about infinite regress, due to Dennett (1998; 2015), is that a second 'transduction' of perceptual discriminations that have *already been made once* by the brain, into some internal depictive medium, for the benefit of a

notional audience, would be functionally redundant. When we consider an apple to have a certain subjective appearance, Dennett suggests, we need not suppose that there is a further thing — *the appearance of the apple* — rendered in some kind of physical or virtual space, and that judging the apple to appear a certain way consists in somehow 'inspecting' that appearance; instead we can suppose that we simply *judge the apple to have a certain appearance*.

The intuition Dennett wants us to discard is conveyed by an imagined interlocutor reflecting on the experience of apparent motion:

> I know there wasn't *actually* a moving spot in the world — it's just apparent motion, after all — but I also know the spot *seemed* to move, so in addition to my judgment that the spot seemed to move, there is the event which my judgment is *about*: the seeming-to-move of the spot. There wasn't any real moving, so there has to have been a real seeming-to-move for my judgment to be about. (Dennett, 1991, p. 134)

It is this 'real' seeming-to-move that Dennett wants us to regard as a legacy of Cartesian thinking and to recognize as already subsumed by a judgment whose *content* is a seeming-to-move.

To relate the view I have presented here to Dennett's stance on perceptual experiences, we need to separate his aspirations into two strands. One part of the claim is reframing experiences as subpersonal suppositions about ourselves: his suggestion that there is 'no such phenomenon as real seeming — over and above the phenomenon of judging in one way or another that something is the case' (*ibid.*, p. 364). This we can wholeheartedly embrace; indeed this inspired the methodological observation in §2 that we cannot distinguish — *I* cannot distinguish, even from the 'inside' — my merely *taking* the apple to look a certain way from it 'actually' looking like that. So we can share Dennett's view that the problem of explaining perceptual consciousness can indeed be reduced to the problem of explaining how we come to form various reflexive judgments about appearances.

The second of Dennett's goals is to purge any role for an image from a potential account of perceptual consciousness. But if we buy the argument from Denett's own heterophenomenological method in §2.2.1, we are forced to disagree with Dennett's claim that a first-personal image of the world, 'observed' by a notional audience, would be functionally redundant. What heterophenomenology reveals instead is a first-personal image *at the heart* of any functional characterization of perceptual consciousness. Our ability to behave about such an image — and in particular to behave about such an image as though it

were 'ours' — is, functionally speaking, exactly what perceptual consciousness is.

Dennett is right to insist that there is no rendering of an image somewhere that underwrites your judgments about how things seem to you, but wrong to doubt — and often to outright deny — that the *content of those judgments is an image*. Modelling yourself modelling the world gives rise to an 'audience': not an inner homunculus whose power of vision is the source of our power of vision, but a self-simulating agent who can not only *see* but can *appreciate that it sees*, and is thus able to become its *own* audience. To borrow a metaphor from Douglas Hofstadter, we 'watch ourselves watching the world' (interviewed in Hoenderdos, 1988). There is no threat of regress here because phenomenal perception is not explained in terms of phenomenal perception, but bootstrapped out of mere perceptual competence.

Acknowledgments

Thanks to Mike Anderson, Axel Cleeremans, Max Jones, François Kammerer, Tom McClelland, Takuya Niikawa, Stuart Presnell, and Keith Wilson for helpful feedback on the several earlier drafts. I also thank Jakob Hohwy for a conversation about the role of conceptual analysis in relation to the natural sciences, which motivated my description of 'naturalizing the explanandum' in §2.

References

Adams, J.K. (1957) Laboratory studies of behavior without awareness, *Psychological Bulletin*, **54**, pp. 383–405.

Akins, K. (2002) A question of content, in Brook, A. & Ross, D. (eds.) *Daniel Dennett*, Cambridge: Cambridge University Press.

Armstrong, D.M. (1981) *What is Consciousness?*, Ithaca, NY: Cornell University Press.

Bruineberg, J., Rietveld, E., Parr, T., van Maanen, L. & Friston, K.J. (2018) Free-energy minimization in joint agent–environment systems: A niche construction perspective, *Journal of Theoretical Biology*, **455**, pp. 161–178.

Chalmers, D.J. (1998) On the search for the neural correlate of consciousness, in Hameroff, S., Kaszniak, A. & Scott, A. (eds.) *Towards a Science of Consciousness II: The Second Tucson Discussions and Debates*, Cambridge, MA: MIT Press.

Clark, A. (2016) Strange inversions: Prediction and the explanation of conscious experience, in Huebner, B. (ed.) *Engaging Daniel Dennett*, Oxford: Oxford University Press.

Cleeremans, A. (2011) The radical plasticity thesis: How the brain learns to be conscious, *Frontiers in Psychology*, **2**, art. 86.

Cleeremans, A. (this issue) Consciousness (unconsciously) designs itself, *Journal of Consciousness Studies*, **26** (3–4).

Damasio, A. (1999) *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, San Diego, CA: Harvest Book/Harcourt Brace.
Dennett, D. (1984) *Elbow Room: The Varieties of Free Will Worth Wanting*, Cambridge, MA: MIT Press.
Dennett, D.C. (1987) *The Intentional Stance*, Cambridge, MA: MIT Press.
Dennett, D.C. (1991) *Consciousness Explained*, Boston, MA: Little, Brown.
Dennett, D.C. (1998) The myth of double transduction, in Hameroff, S., Kaszniak, A. & Scott, A. (eds.) *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, pp. 97–107, Cambridge, MA: MIT Press.
Dennett, D.C. (2002) Does your brain use the images in it, and if so, how?, *Behavioral and Brain Sciences*, **25** (2), pp. 189–190.
Dennett, D.C. (2003) Who's on first? Heterophenomenology explained, *Journal of Consciousness Studies*, **10** (9), pp. 19–30.
Dennett, D.C. (2004) *Freedom Evolves*, New York: Penguin.
Dennett, D.C. (2005) *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*, Cambridge, MA: MIT Press.
Dennett, D.C. (2009) Darwin's 'strange inversion of reasoning', *Proceedings of the National Academy of Sciences*, **106** (suppl. 1), pp. 10061–10065.
Dennett, D.C. (2013) Expecting ourselves to expect: The Bayesian brain as a projector, *Behavioral and Brain Sciences*, **36** (3), pp. 209–210.
Dennett, D.C. (2015) Why and how does consciousness seem the way it seems?, in Metzinger, T.K. & Windt, J.M. (eds.) *Open MIND*, chapter 10(T), Frankfurt am Main: MIND Group.
Evans, G. (1982) *The Varieties of Reference*, Oxford: Oxford University Press.
Friston, K., Kilner, J. & Harrison, L. (2006) A free energy principle for the brain, *Journal of Physiology-Paris*, **100** (1), pp. 70–87.
Friston, K., Adams, R., Perrinet, L. & Breakspear, M. (2012) Perceptions as hypotheses: Saccades as experiments, *Frontiers in Psychology*, **3**, art. 151.
Goldman, A. (1993) Consciousness, folk psychology, and cognitive science, *Consciousness and Cognition*, **2** (4), pp. 364–382.
Greif, S. & Siemers, B.M. (2010) Innate recognition of water bodies in echolocating bats, *Nature Communications*, **1**, art. 107.
Hassin, R., Uleman, J. & Bargh, J. (2005) *The New Unconscious*, Oxford: Oxford University Press.
Helmholtz, H. von (1867) *Handbuchderphysiologischen Optik*, Allgemeinen Encyclopädie der Physik, Leipzip: Leopold Voss.
Hoenderdos, P. (1988) *Victim of the Brain*, documentary.
Hohwy, J. & Michael, J. (2017) Why should any body have a self?, in de Vignemont, F. & Alsmith, A. (eds.) *The Subject's Matter*, pp. 363–391, Cambridge, MA: MIT Press.
Karmiloff-Smith, A. (1990) Constraints on representational change: Evidence from children's drawing, *Cognition*, **34** (1), pp. 57–83.
Levine, J. (1994) Out of the closet: A qualophile confronts qualophobia, *Philosophical Topics*, **22** (1–2), pp. 107–126.
Lycan, W. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.
Mach, E. (1914) *The Analysis of Sensations, and the relation of the physical to the psychical*, Chicago, IL: Open Court.
McDermott, D. (1992) Little 'me', *Behavioral and Brain Sciences*, **15** (2), pp. 217–218.

Metzinger, T. (2003) *Being No One: The Self-Model Theory of Subjectivity*, Cambridge, MA: MIT Press.
Metzinger, T. (2005) Précis: *Being No-One*, *PSYCHE: An Interdisciplinary Journal of Research On Consciousness*, **11**, pp. 1–30.
Metzinger, T. (2007) Empirical perspectives from the self-model theory of subjectivity: A brief summary with examples, in Banerjee, R. & Chakrabarti, B.K. (eds.) *Models of Brain and Mind*, vol. 168 of *Progress in Brain Research*, pp. 215–278, Amsterdam: Elsevier.
Nagel, T. (1974) What is it like to be a bat?, *Philosophical Review*, **83** (October), pp. 435–450.
Noë, A. (2004) *Action in Perception*, Cambrdige, MA: MIT Press.
O'Regan, J. & Noë, A. (2002) A sensorimotor account of vision and visual consciousness, *Behavioral and Brain Sciences*, **24** (5), pp. 939–973.
Pollak, G.D. & Casseday, J.H. (1989) *The Neural Basis of Echolocation in Bats*, Zoophysiology and Ecology Series, Berlin: Springer-Verlag.
Rogers, B. & Graham, M. (1979) Motion parallax as an independent cue for depth perception, *Perception*, **8** (2), pp. 125–134.
Rosenthal, D.M. (1986) Two concepts of consciousness, *Philosophical Studies*, **49** (May), pp. 329–359.
Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson & Co.
Stalnaker, R.C. (2008) *Our Knowledge of the Internal World*, Oxford: Oxford University Press.
Stubenberg, L. (1998) *Consciousness and Qualia*, Amsterdam: John Benjamins.
Varela, F., Thompson, E. & Rosch, E. (1992) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.
von Uexküll, J. (1957) A stroll through the worlds of animals and men, in Schiller, C.H. (ed.) *Instinctive Behavior: The Development of a Modern Concept*, pp. 5–80, New York: International Universities Press