This is an unedited version of a paper written in 2012 and accepted for publication in a forthcoming *Festschrift* for Mark Platts.

In it I argue that the Helmholtz/Bayes tradition of free energy neuroscience begun by Geoffrey Hinton and his colleagues, and now being carried forward by Karl Friston and his, can be seen as a fulfilment of the Quine/Davidson program of radical interpretation, and also of Quine’s conception of a naturalized epistemology. This program, in turn, is rooted in Helmholtz’s scientific reconception of Kant’s notion of a concept-led synthesis of affections by an extra-sensory reality that creates human self-consciousness, a topic already discussed in my (2012) *Psychoanalysis, Representation, and Neuroscience*.

I suggest how far I think 20th century analytical philosophy went astray when Frege, Russell, and Wittgenstein turned away from what should have been their legacy from Helmholtz to espouse of a conception of perception and epistemology that Helmholtz had already shown to be false.

**Kantian Neuroscience and Radical Interpretation:**

***Ways of Meaning* in the Bayesian Brain**

Mark’s first book, *Ways of Meaning,* played an important (if non-evangelical) role in disseminating Davidsonian ideas in the philosophy of language. In his preface to the latest edition, he comments on ‘talk of the death of the kind of philosophy of language presented in the first edition of this book’. As he says, philosophical desires ‘to initiate funeral rites’ are commonly exaggerated and premature. Still this raises the question as to how – fluctuations in philosophical fashion apart -- we should now evaluate the philosophical program to which the book contributed.

1. **Radical interpretation.**

In answering this we should bear in mind the full ambition of that program: to explicate the human abilities of understanding and meaning.Members of our species are capable of co-ordinating their activities, for good as for ill, to an extent that is unique among vertebrate animals. This is evident in the fact that almost everything we do, or make, or use – the food we eat, the clothes we wear, the books we read, the tools we use, devices by which we communicate and share experience, the science and technology that plays an ever-increasing role in our lives – involves tacit or explicit co-ordination with very many others, in intersecting networks of co-operation or competition that span the globe.

All this pivots on the way each of us can produce behaviour that is meaningful both to ourselves and to others. This enables us to pursue goals by moving in ways that are perceived as actions informed by desires, beliefs, and intentions, and to communicate these and other states of mind by making sounds (or producing other interpretable signs) perceived as meaningful utterances. In thus knowing and communicating our thoughts we establish a socially shared consciousness which that gives us power over life and death never before concentrated in the a single species. All these accomplishments – including the most precise and abstract reaches of science and mathematics -- are sustained by the interpretive abilities that underpin meaning and understanding, and would wither in the instant should these fail.

Davidson rightly regarded explicating these abilities as an important philosophical task. He sought to cast light on them by constructing a fully explicit theory that would enable an imaginary scientific investigator – a ‘radical interpreter’ – to explain partly similar data (the production of observable behaviour) by construing it in a partly similar interpretive framework (explaining the behaviour as action, including speech, caused by the beliefs and desires of its agents). In this, as is familiar, Davidson was extending Quine’s project of explicating meaning via an imaginary scientific radical translator. And Quine can be seen as following the later Wittgenstein, who had also sought to investigate our ability to understand meaning and motive, and particularly the way it sustained a social extension of individual intentionality. His final investigations also were conducted via the figure of an imaginary explorer – the first radical interpreter – whose task, like that set for Quine’s translator, was to understand peoples whose languages and forms of life he had never encountered before, but including, for Wittgenstein’s particular purposes, an group who spoke only in monologue.[[1]](#endnote-1)

**2. Kant, Helmholtz, and unconscious prediction**.

In seeking to understand perceptual and cognitive abilities via the construction of a scientific theory Quine and Davidson perforce located themselves in another, and Kantian, tradition. This stemmed from the great 19th century scientist and mathematician Hermann von Helmholtz (see Patton 2012)[[2]](#endnote-2), and had previously influenced both Frege and Wittgenstein. Among other things Helmholtz was one of the first formulators of the principle of the conservation of energy. His research in sensory physiology – supplemented by a range of personal inventions such as the opthalmascope – expanded the science of optics and revolutionized the treatment of malfunctions of the eye. More recently his work has been formative both for the empirical psychology of perception and current developments in neuroscience. His relevance analytical philosophy, however, remains to be fully acknowledged.

**Helmholtz and Kant**

Tim Lenoir describes Helmholtz’s relation to philosophy, and to Kant in particular, as follows:

Although he considered himself a physicist, Hermann Helmholtz devoted more pages of his published work to physiological psychology and philosophical problems related to spatial perception than to any other subject…Helmholtz pursued fairly consistently over many years a line of investigation centered around the questions, *How do our mental representations of external objects get constructed? And how do those representations relate to the world of external objects…?*

And as regards Kant

Helmholtz considered his work in epistemology…as updating Kant**’**s views in light of new developments in experimental physiology…Helmholtz thus considered himself to be more consistently Kantian than Kant had been himself.

In his *Critique of Pure Reason* Kant had advanced the first detailed model of the human mind that did not assume, but rather sought to explain, basic aspects of consciousness and perceptual experience. According to Kant we have a manifest image of ourselves as self-conscious subjects of experience and agents of action. We take our experience to be internal to our minds, but to disclose a world of physical objects and events external to our minds, and in spatio-temporal and causal relations with one another, including our own bodies and organs of perception. Our actions in moving among these objects and events are guided by our experience in perceiving them, and this arises from their effects on our sensory organs as we do so.

Philosophers had long had reason to know that experience was not simply produced by effects on the sensory organs. Descartes had stressed that full conscious and apparently perceptual experience regularly arose in dreams, and hence without the participation of the sensory organs themselves. Accordingly Kant recognized that we must understand experience – and the whole conscious image of ourselves of which it is part – as somehow created by internal activity of our minds, and in a way that was prior to the sensory impingements that we understood as conscious experiences of the objects we perceived*.* Since this creative activity apparently preceded the conscious experience it produced, it had to be unconscious. Kant called the consciousness-creating activity *synthesis* and argued that it was governed by the basic concepts – e.g. of *self*, *experience*, *time*, *space*, *substance*, *object*, *event*, *cause* – in whose terms we experience the world.

Kant’s notion of unconscious processes that created perceptual consciousness remained in place as (most) later investigators came to acknowledge that the organ of synthesis must in fact be the brain: so that, as Freud was to put it, ‘consciousness is the subjective side of one part of the physical process in the nervous system.’[[3]](#endnote-3) Helmholtz himself had definite ideas on this topic. As Lenoir says, he criticized the Kantianism of his time for

setting up two worlds, an objective physical world and a subjective world of intuition, somehow causally related to one another but existing as independent, parallel worlds relying on some unexplained pre-established harmony between perceptions and the real world as the basis for the objective reference of knowledge claims. Helmholtz, believing it was necessary to escape the subjectivity of idealist positions asked: But how is it that we escape from the world of the sensations of our own nervous system into the world of real things?’ [[4]](#endnote-4)

Helmholtz’s question -- how do we escape from the world of the sensations of our own nervous system into the world of real things? – arises in both scientific and philosophical forms. As noted above, we are familiar with dreams as creating a kind of virtual reality: an emotionally charged and presence-embodying image of ourselves as (seemingly) engaged in interaction with the world. In waking life, by contrast, this same internally generated image is so constrained by sensory impingements as to yield the largely veridical perceptual experiences by which we govern our movements and actions. But given the internality of its generation, how is the accuracy of this image secured?

We can take this as a scientific question about the brain and nervous system, enclosed as they are within the skull and skin, and so physically remote from the objects and events we take them to represent. But a scientific answer to this question – an account of how the brain in fact achieves representational accuracy in perception -- should also help to address the philosophical question as to how a conscious subject can know that she is not a brain housed in a vat, or embedded in a Matrix. So it is clear that Helmholtz’s enquiries, although scientific in their aim, also relate to questions that would be regarded as philosophical even in the narrowest confines of the analytic tradition.

Helmholtz was deeply concerned with the relation between science and epistemology. Indeed, like Quine after him, he took epistemology to be the domain in which the differing disciplines could most fruitfully inform one another. Thus he says

…the problems considered fundamental to all science were those of the theory of knowledge: *What is true in our sense-perceptions and thought?* And ‘*In what way do our ideas correspond to reality?* Philosophy and the natural sciences attack these questions from the opposite directions, but they are the common problems of both.[[5]](#endnote-5)

And he notes that Kant’s purpose was

to examine the source of our knowledge and its justification, something that will always remain the task of philosophy and that no generation can shirk with impunity….[[6]](#endnote-6)

Helmholtz did not shirk this philosophical task. Rather he brought science to bear on it, in what he hoped would be a continuing collaboration between the disciplines. And his account of mental representations and their relation to the physical world combined scientific and justificatory thought in an extraordinarily prescient way.

**Synthesis and prediction**

In explaining his notion of synthesis, Kant had stressed that the concepts that enable us to form our conscious image of the world are used both normatively and predictively in relation to it. As he says

By *synthesis,* in its most general sense, I understand the act of putting different representations together, and of grasping what is manifold in them in one knowledge…If we enquire what new character *relation to an object* confers upon our representations, what new dignity they thereby acquire, we find that it results only in subjecting the representations to a rule, and in necessitating us to connect them in one specific manner…[[7]](#endnote-7)

The concepts by which we understand an object of experience require – so that our use of these concepts tacitly predicts – that experiences unfold ‘in one specific manner’. So, as Kant sought to bring out in the *Analogies of Experience[[8]](#endnote-8)*, in construing experiences via concepts of substances we expect, or tacitly predict, that such experiences will represent substances as enduring in space and time. Likewise (and as Hume had previously stressed) in construing experiences as of causes, we tacitly predict experiences of their effects. Again, in construing experience as of a material object like a house, or yet again as of our own bodies, we expect (predict) one perceptual sequence if we walk around a house one way, another if another. Perception guides movement, and so varies with it.

**Spatial determination, perception, and action.**

Helmholtz saw the relation of movement to perception in a further and deeper way. Scientific observation indicated that the nervous system actively controls the collection of sensory data via bodily movements and changes that it initiates. Hence, as he stressed, all ‘sensations of the external senses’ were ‘preceded by some sort of innervation’ so that they were all ‘spatially determined’. This determination, moreover, was not solely governed adjustments of the eyes or other organs of perception. Rather it was also brought about by intentional actions and other purposive movements, whose sensory feedback served to regulate and synchronize them.

**Unconscious predictive models**

Helmholtz therefore reconceived perception and action as co-ordinate parts of a single underlying process – one that related movement and perception in a predictive way, so as to achieve the active control of sensory data that was required for us to thrive. The ‘spatial determination’ of sensory data for this purpose entailed the existence of powerful internal models of the objects and events we perceive, manipulate, and navigate; and these models, as Helmholtz saw, would have to be at least partly constructed via the same active collection of sensory data as they were to facilitate.

Helmholtz thought that such a construction of internal models should be regarded as a psychological process, and one that began early in life. So he hypothesized ‘a field of mental operations which has seldom been entered by scientific explorers’, the drawing of *induktionsschluesses[[9]](#endnote-9)*, or inductive conclusions (see Hawthorne 2012).[[10]](#endnote-10) He repeatedly stressed that these were ‘not conscious activities, but unconscious ones’, and according to his account they were more or less continuously engaged with the data collected by the nervous system. Since these processes involved the use of models or hypotheses, they might now be described as a form of tacit model-based abduction (see Douven 2011)[[11]](#endnote-11). As philosophers of post-Chomskian cognitive science were later to do, Helmholtz argued that such inferential processes should be seen as required for many kinds of knowing-how that could not be understood as knowing-that.[[12]](#endnote-12)

These unconscious abductive inferences were to be described in psychological or mental terms because their conscious conclusions -- judgments of perception, belief, or knowledge -- were already described in this way. Judgments of perception, such as ‘I am seeing and touching a table’ were based on data arising in (internal and external) sensory systems, and made use of concepts in synthesizing this data. The objects and events so perceived and conceived (the table, the positions and movements of the arm and hand in touching it, etc.) were causes of the data on which they were based. So the unknown concept-applying models (encompassing the seeing eyes and the extended arm, hand, and fingers, as well as the table, its surface, the space between eye and table, eye and hand, etc.) could be compared to those by which scientists drew conclusions about causes of perceptual data that were beyond their pretheoretical perceptual ken. As Helmholtz argued

An astronomer, for example, comes to real conscious conclusions of this sort, when he computes the positions of the stars in space, their distances, etc., from the perspective images he has had of them at various times and as they are seen from different parts of the orbit of the earth. His conclusions are based on a conscious knowledge of the laws of optics. Still it may be permissible to speak of the psychic acts of ordinary perception as unconscious conclusions, thereby making a distinction of some sort between them and the common so-called conscious conclusions. And while it is true that there has been, and probably always well be, a measure of doubt as to the similarity of the psychic activity in the two cases, there can be no doubt as to the similarity between the results…[[13]](#endnote-13)

In the tacit instances Helmholtz was considering the inferential process yielded particular predictive conclusions about the causes of sensory data. In his technical writings, as Lenoir explains, Helmholtz explicated such inferences as answerable to complex physiological analogues of theoretical simplicity and statistical accuracy.[[14]](#endnote-14) In his popular expositions, by contrast, he described them in statements that were lucid and compact, but far-reaching in their implications. Thus

Each movement we make by which we alter the appearance of objects should be thought of as an experiment designed to test whether we have understood correctly the invariant relations of the phenomena before us, that is, their existence in definite spatial relations.[[15]](#endnote-15)

This was a radical claim, and Helmholtz intended it to hold not only for conscious perceptual experience but also at multiple levels of motor and sensory processing in the nervous system. As we shall see below, it can be also be said, with appropriate qualification, to hold for the guidance of animal movement generally, and over the whole domain of life.

**3. Hypothetical synthesis and computational neuroscience.**

Perceptual synthesis is accomplished by the use of concepts, and these are realized in the brain. How can we understand the brain as using concepts to model, predict, and explain experience – ‘the appearance of objects’ – in the way Helmholtz envisaged? To see more about this let us use his example of astronomical thinking, and take the models he envisages for unconscious abductive inference as analogous to scientific models of the solar system, such as those propounded by Ptolemaic and Copernican astronomy.[[16]](#endnote-16)

Such models can be more or less abstract and general than the particular hypotheses they may be used to generate, where the latter might explain and predict, say, data consisting in actual observations of the relative positions of celestial bodies (Helmoltz’s ‘perspective images’) in the night sky. Also such models may themselves be generated by models that are yet more general and abstract, as in the case of Copernican models and Newton’s or Einstein’s theory of gravitation. In such cases the confirmation or disconfirmation of a particular data-explaining hypothesis may be transferred to a series of more general models and hypotheses, with increasing distance, and increasing indirectness of evidential impact, from the data originally involved.

Taken this way the view we are considering would be that we tacitly use our concepts to create models of ourselves and our situation in the world, and these in turn to generate particular hypotheses that explain the appearances of the objects and events we perceive, both externally and internally to ourselves. For these we can use ‘H’, ‘d’, and ‘e’ to stand for ‘hypothesis’ ‘data’, and ‘evidence’ respectively, and restrict ourselves to explanations in which the H’s specify internally or externally perceivable causes of the basic sensory data that – on this account -- they unify, predict, and explain. Going over some points about this will also be useful for relating Helmholtz’s version of Kantian synthesis to the radical interpretation of Wittgenstein, Quine, and Davidson. For, as noted above, the latter all proceed via the notion of hypotheses used to explain data as in the practice of science.

**Causal explanations as consisting of working hypotheses confirmed or disconfirmed by data or evidence.**

We characteristically frame casual explanations by proposing causal hypotheses to explain data – perhaps just to understand the data better, or perhaps because we find the data surprising, in the sense that we cannot see, and want to understand, why the data are as they are. Such hypotheses explain and unify data by integrating them into larger but hypothetical causal patterns – these, of course, being patterns produced by the causal mechanisms or processes hypothesized for explanation. In this way the hypotheses explain why the data are as they are, and so render the data unsurprising, in the sense that they enable us to see that the data are as we should expect them to be, given the causal mechanisms or processes hypothesized to explain them.

To say this, however, is to say that such hypothesis predict the data they serve to explain. A degree of prediction is inherent in showing that the data are as we should expect them to be, given the hypothesis – for in this, the data are represented as expectable in a way they were not before we brought the hypothesis to bear on them, and again in a way they would not be, given the negation of that hypothesis. In consequence we can regard an hypothesis is explanatory in the sense with which we are concerned just if the probability of the data given the hypothesis is greater than the probability of the data given the negation of the hypothesis, which we can abbreviate P(d given H) > P(d given not-H).

**Bayesian predictive confirmation.**

This, however, means that data so explained become evidence by which the hypothesis can be confirmed or disconfirmed. For by Bayes’ theorem (Joyce 2008)[[17]](#endnote-17) the credibility (probability) we assign to an hypothesis given data that it predicts [P(H given d)] should be greater than the credibility or probability [P(H)] we assign it prior to the explanatory prediction. But for data to increase the credibility (probability) we assign to hypotheses is for the data to constitute evidence that tends to confirm those hypotheses, and comparably for disconfirmation. So in Bayesian terms, and taking data now as evidence, we can regard e as confirming H just if P(H given e) > P(H), which will be so just if H predicts e in the sense that P(e given H) > P(e given not-H).

**Predictive confirmation and the brain**.

Helmholtz took the concepts, models, hypotheses, and data used in tacit abductive inference to be embodied in the nervous system. The data originated in the sensory systems that Kant described in terms of inner and outer sense, as these were affected by causes internal or external to the body. So for example in the case of reading by sight the data might arise in the response of neural cells in the retina of the eye to what Quine describes as ‘certain patterns of irradiation in certain frequencies’ reflected from letters on a printed page, or again generated from a computer screen. These were what Helmholtz called ‘the sensations of the nervous system’, from which our epistemic problem (or that of the brain) was to ‘escape’ by making trustworthy inferences about ‘the world of real things’.

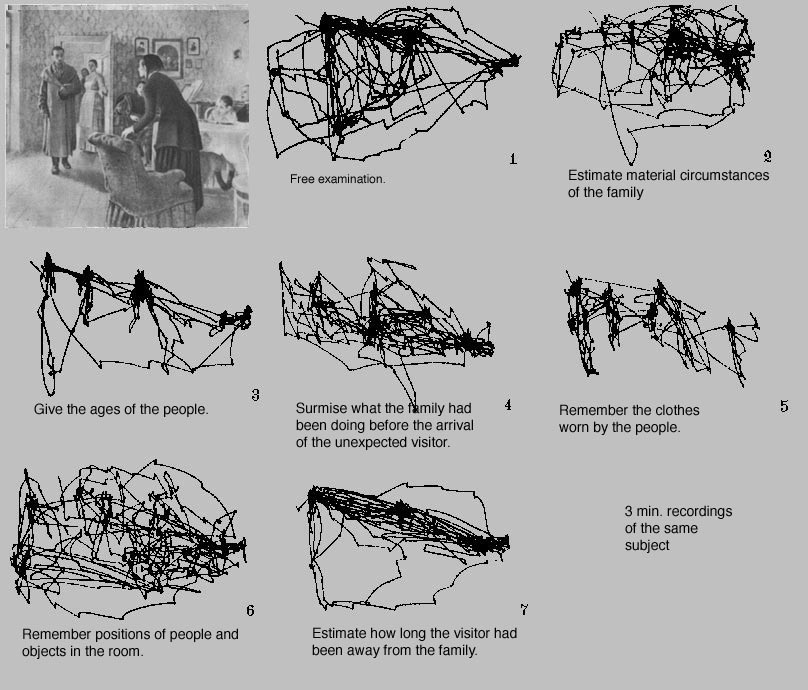
In the case of reading the ‘real things’ would include the page, letters, words, and sentences upon which the eyes were focussed, and also the action of reading in which the reader would take herself as engaged, the book she was reading, the author she took to have written it, the real or imaginary things she took the author to be writing about, and much of the rest of the world, as represented by her concepts. As Frege noted, with ‘a few syllables’ an author can evoke ‘an incalculable number of thoughts’ So in reading, or even glancing at a page, we are cognitively prepared to reckon almost anything as among the causes that might be cited in explaining the exact words (‘epistemology’, ‘photon’, ‘vampire’, ‘galaxy’) upon which our eyes might momentarily fix as we read.

The data Kant described in terms of outer and inner sense are now often discussed in terms of exteroception, proprioception, and interoception[[18]](#endnote-18) Roughly, exteroception gathers data from causes outside the body: from those of seeing, hearing, smelling, tasting, touching, tensions on the skin, etc. By contrast interoception collects data from the interior of the body, and so from homeostatic functioning and the arousal of emotions; and proprioception collects data from the deployment of muscles, and so from movements, posture, etc. The data from these sources are processed in a way that integrates them into a single conscious image of the self in the world – the image available to us in the first-person perspective of current conscious experience – and the skin, which bounds the body and is the source of a multiplicity of data, plays an important role in delineating the boundaries assigned to the self.

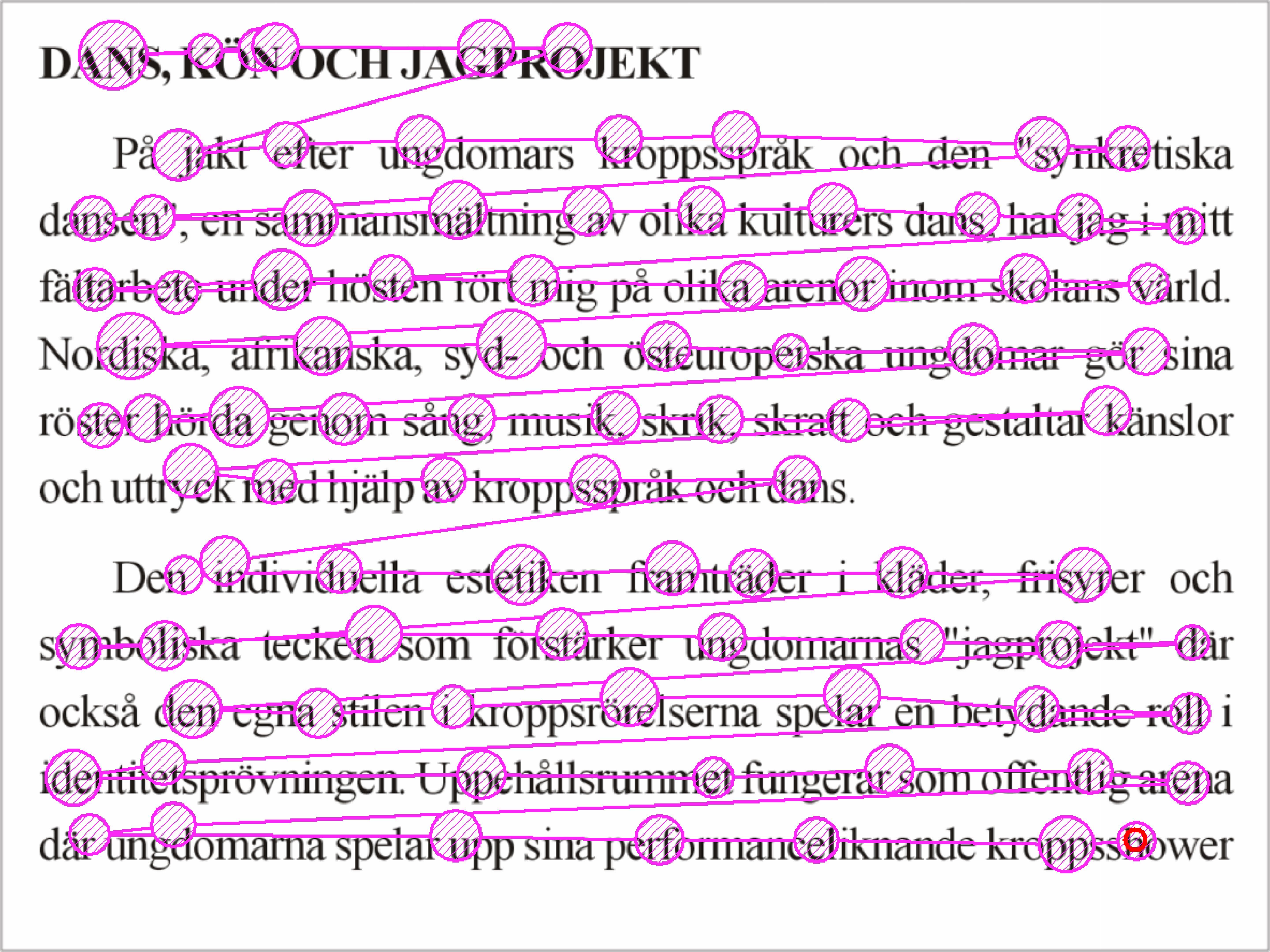
Helmholtz exemplified his claim that exteroceptive data are collected by ‘innervation’ and movement by citing the movements caused by the muscles of the eyes.[[19]](#endnote-19) Although it may seem to us (and many philosophers insist) that when we look we simply take in the visual scene before us, our experiences in visual perception are based on data collected by unconscious but fast and systematic movements of the eyes. These concentrate light reflected by things (e.g. particular letters in words on a page) in a small part of the visible field – corresponding to a thumbnail seen with arm extended, or seven or eight letters of print on a page we are reading -- on the most sensitive part of the retina (the fovea), for about a fifth of a second at a time, during which time, among other things, we unconsciously enlarge or contract the pupils of our eyes in accord with the emotional significance of what we expect or find.

The resulting rapidly changing unconscious mosaic of thumbnail-sized ‘retinal snapshots’, each succeeding others in a fraction of a second, makes up a central part of the stream of visual sensory data. Aspects of the nature of these data are illustrated in the diagrams below. [[20]](#endnote-20)

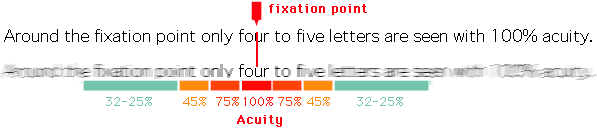
**Conceptual guidance of data-collection by involuntary saccades in viewing a picture**



**Saccades and fixation points in reading**



**Saccadic fixation and visual acuity**



Such conceptually governed and active but involuntary collection of sensory data – the kind in which the reader of this paragraph is now engaged -- accords with Kant’s idea that prior to conscious experience sensory data are synthesized via concepts, as also with Helmholtz’s particular understanding of synthesis by analogy with theoretical unification. For in using concepts as hypotheses to collect, explain and unify fragmentary sensory data the brain does seem, as Kant says, to ‘grasp what is manifold in [the data] in one knowledge’. This ‘one knowledge’ is the interpretation of the data implicit in perceptual experience, which represents – not the fluctuating internal data, but -- their significant distal causes.

**Conceptual hypotheses and sensory data in vision**

Helmholtz describes this in terms of the concepts used in perception by saying

The idea of a body in space, of a table for instance, involves a quantity of separate observations. It comprises the whole series of images which this table would present to me in looking at it from different sides and at different distances; besides the whole series of tactile impressions that would be obtained by touching the surface at various places in succession. Such an idea of a single individual body is, therefore, in fact a *conception* (*Begriff*) which grasps and includes an infinite number of single, successive apperceptions that can all be [unconsciously] deduced from it; just as the species “table” includes all individual tables and expresses their common peculiarities.

Hence, as he says,

The idea of a single individual table I carry in my mind is correct and exact provided [that via unconscious abductive reasoning] I can deduce from it correctly the precise sensations I shall have when my eye and my hand are brought into this or that definite relation with respect to the table…the idea of a thing is correct for him who knows how to determine correctly from it in advance [via unconscious abductive reasoning] what sense-impressions he will get from the thing when he places himself in definite external relations to it… [[21]](#endnote-21)

**Experiences as hypothetical representations of the causes of sensory data**.

The interaction between sensory data, Kant/Helmholtz synthesis, and self-ascriptive judgments of conscious perceptual experience is well illustrated in Dennett’s description of eye-tracking experiments in *Consciousness Explained.[[22]](#endnote-22)*

When your eyes dart about in saccades the muscular contractions that cause the eyeballs to rotate are *ballistic actions…*whose trajectories at lift-off will determine where they will hit ground zero at a new target…if you are reading text on a computer screen your eyes will leap along a few words with each saccade, farther and faster the better a reader you are…a computer equipped with an eye-tracker can detect and analyse the lift-off in the first milliseconds of a saccade, calculate where ground zero will be, and, *before the saccade is over*, erase the word on the screen at ground zero and replace it with a different word of the same length. What do you see? Just the new word, and with no sense at all of something having been changed. As you peruse the text on the screen, it seems to you for all the world as if the words were carved in marble, but to another person reading the same text over your shoulder (and saccading to a different drummer) the screen is aquiver with changes.

The effect is overpowering. When I first encountered an eye-tracker experiment, and saw how oblivious subjects were (apparently) to the changes flickering on the screen, I asked if I could be a subject. I wanted to see for myself…While I waited for the experiment to start I read the text on the screen…’Why don’t you turn it on?’ I asked. ‘It is on’ they replied.

Here two viewers of the same screen – the ‘oblivious’ subject and another watching, say, over his shoulder – see (or experience) the screen in strikingly different ways. Their different visual experiences systematically reflect the differing visual data collected by the saccadic movements of their eyes, and so can be understood in Helmholtz’s terms as tacitly formed abductive hypotheses that explain and unify those perceptual data by reference to the objects or events that cause them. The oblivious subject, whose sensory data are so controlled by the eye-tracking mechanism as to be explained on the hypothesis that he is reading an unchanging text, sees such a text ‘as if the words were carved in marble’. The other, naturally collecting the fuller data available to her, rightly sees the screen as ‘aquiver with changes’.

The differing conscious perceptual experiences here are just as we would expect them to be, on the assumption that their contents were determined by tacit abductive hypotheses -- about successive letters, words, and meanings -- of the kind Helmholtz envisaged. Contrasting such conscious experiences with the data they synthesize, we can see that it is not the data but the *hypotheses realized as perceptual experience* that provide the continuity and order (the carving in marble) that we perceive the world to have. By contrast the manifold unconscious data these hypotheses explain (Helmholtz’s ‘sensations of the nervous system’) fluctuate with a rapidity we scarcely imagine.

Again, we can appreciate the hypothetical character of experience, and also the way it is structured top-down by our concepts, by reflecting that both ‘The words seemed carved in marble’ and ‘The screen was aquiver with changes’ could be reports of what was experienced in a dream. In such a case the dreamer’s (shut) eyes might well have been saccading, as they characteristically do in REM sleep. But there would have been no incoming data from the retina – no bottom-up visual impingements -- for these saccades to collect. The brain apparently produces hypothesis-realizing experiences in both waking and dreaming consciousness – with the former constrained by sensory data but the latter not.[[23]](#endnote-23)

**Perception as conscious experience of causes.**

Taking the example above as characteristic we can say that the brain operates tacitly and abductively to transform sensory data into conscious experience of their probable causes. Helmholtz expressed this by saying that in perceptual experience ‘the concept and cause combine’, so that ‘our impression, consequently, seems to us a pure image of the external state of affairs, reflecting only that condition and depending solely on it.’[[24]](#endnote-24) And in recent years -- as indicated by the title of Karl Friston et al (2013): ‘Perceptions as Hypotheses, Saccades as Experiments’[[25]](#endnote-25) -- Helmholtz’s ideas have become central to a rapidly growing research program in neuroscience. Expositions of this program by Andy Clark and others[[26]](#endnote-26) have begun to attract attention from philosophers and cognitive scientists as well as the neuroscientists to whom it has become familiar.

The beginnings of this program are often traced (e.g. by Friston)[[27]](#endnote-27) to work by Geoffrey Hinton and his colleagues, which transformed connectionism (Garson 2012)[[28]](#endnote-28) into the computational neuroscience of the Bayesian brain (Doya et al 2007).[[29]](#endnote-29) Thus Dayan, Hinton, Neal, and Zemel, began their seminal paper ‘The Helmholtz Machine’[[30]](#endnote-30), by saying

Following Helmholtz, we view the human perceptual system as a statistical inference engine whose function is to infer the probable causes of sensory input. We show that a device of this kind can learn how to perform these inferences without requiring a teacher…

Here, in accord with then-recent findings about primate visual cortices, the authors envisaged the ‘statistical inference engine’ of the brain as working via ‘a complex hierarchical generative model of its inputs’[[31]](#endnote-31) – that is, a hierarchical model that served to predict sensory inputs – realized in the cerebral cortex. In the Helmholtz machine the layers of cortical processing were represented by layers of neuron-like processing units, connected hierarchically by two distinct sets of synaptic weights, alterable by built-in learning algorithms. The bottom layer corresponded to ‘sensory’ input, and weights at higher layers could be adjusted to embody hypothetical representations of causes of input, as abstracted by pattern-recognizing algorithmic learning.

This meant that, like the cortex, the Helmholtz machine could work both top-down (via one set of synaptic connections) and bottom-up (via the other). From the top ‘conceptual’ layer down it acted as a ‘generative model’ implementing inferences predicting input. From the bottom ‘sensory’ layer up it acted as a ‘recognition model’, learning the generative weights and recognizing inputs by inferring their most probable causes. This was done via an important conceptual innovation – that of assuming that the Helmholtz machine, and by analogy the brain, proceeded by minimizing the information-theoretic free energyof the probabilities with which it was concerned.[[32]](#endnote-32)

**Free energy as prediction error**.

As used for the Helmholtz machine, free energy is a quantity associated with a predictive model of the causes of data. This quantity is minimized by maximizing the power of the model to predict the data, and so by minimizing the unpredictability of the data, and therewith errors in predicting it.[[33]](#endnote-33) Crucially, this could be done from *within* the computational system of which both model and data were part, and by tractable computations that approximated Bayesian inference about causes external to it. So apparently the brain could perform the kind of abductive inference Helmholtz had envisaged, by using a generative model such as seemed realized in the cortex, and implementing a mathematically describable process of minimizing free energy. In light of this Karl Friston and his colleagues were able to frame an account of the brain as minimizing free energy that fit with many sources of data from neuroscience[[34]](#endnote-34), as well as from perceptual and Freudian psychology.[[35]](#endnote-35) Howhy, Friston and Roepstorff have recently described this approach in as follows:

There is growing support of the idea that the brain is an inference machine, or hypothesis tester, which approaches sensory data using principles similar to those that govern the interrogation of scientific data…Rather than trying to work backwards from sensory effects to environmental causes, neuronal computational systems work with models, or as we shall say hypotheses, that predict what the sensory input should be, if it were really caused by certain environmental events. The hypothesis that generates the best predictions then determines perceptual content**.[[36]](#endnote-36)**

The predictive hypotheses are generated from the top (concept-realizing) level of the cortical hierarchy, and propagated down – so as to fix hypotheses at a series of lower (and hence successively more sub-personal and less integrated) layers, often working at successively faster temporal scales – until it reaches, and determines the collection of sensory input, at the physiologically separated loci of input for the various sensory systems. (This was illustrated in the first of the diagrams above, in the conceptual direction of fast collection of visual data in saccades.) From the top down, each level works both to predict activity at the level below, and to suppress the activity it succeeds in predicting. At the same time, each level apart from the highest also works bottom-up, to send unpredicted and so unsuppressed activity to the level above. This restricts input to higher levels to data that are so far unpredicted – and this input works to alter the higher-level hypotheses so as to predict it, and so to bring about its own suppression. In this way the model constantly works on input that it is failing to predict, improving its hypotheses (at every level of neural processing) to reduce its ongoing errors.

For such predictive brains, or for persons with brains of this kind, things would be as Helmholtz described. Each movement they made could be construed as an experiment in which they altered sensory data so as to test their conceptual hypotheses, and thereby how well they had understood the causes, including the bodies of which they are part, with which they interact. This locates the brain within the same epistemic horizons as we find ourselves – as part of a world of causes whose nature requires to be inferred from those of their effects that we encounter as sensory data. And while there could be no eliminating the possibility of divergence between our (or our brain’s) best hypotheses and the reality they serve to capture, this could be minimized through abductive practice.

**Prediction as realizing both perception and action.**

Above we used the idea that ‘the hypothesis that generates the best predictions then determines perceptual content’ to describe the content of the experiences of different observers in eye-tracking experiments. But on Friston’s account the brain works not only to understand the world, but also to change it. The model also generates predictions – and further predictive representations including desires and intentions -- about the sensory consequences of motor activity that, if it were initiated, would lessen error in the circumstances as modeled. In this case the brain can minimize error by acting to make its own predictions come true, and it does so, straightway appropriate predictions about the (proprioceptive) consequences of muscular/skeletal activity are generated, by engaging motor reflexes to correct those that are unfulfilled. In this way the brain’s ‘best predictions’ assume the role of sensory goals that motor systems are driven to attain.[[37]](#endnote-37) To put the idea in Davidsonian terms, the brain makes the sentences that describe the agent’s desires true, by making true those that (would) describe the sensory consequences of those desires.

We saw such work on sensory consequences in the first diagram above, in the way saccades – and hence the ocular reflexes by which they implemented -- are ultimately directed by viewers’ desires to gather information the experimenters have requested. Friston accordingly describes such movements together with intentional action as involving ‘active inference’. The predictive role of desires and intentions is signaled (and partly implemented) in commonsense psychology, by our mode of describing them in terms of the conditions of satisfaction or fulfillment that, if acted on, they will work to bring about. As Davidson noted, this is a way of describing causes in terms of (predicted) effects. So here we can see this as amplifying causal theories of action (such as those of Fodor and Millikan), according to which desires are constituted by representations of the situations they operate to cause. On the present account beliefs, desires, and intentions perform this task by setting sensory goals that determine the trajectory of motor activity, often towards what we regard as the conscious experience of satisfying particular desires.

5. **Helmholtz’s influence on Frege and Wittgenstein**.

Helmholtz seems to have influenced both Frege and Wittgenstein. Consider Frege’s introduction of concepts in the *Begriffschrift*:

Even most animals, through their ability to move about, have an influence on their sense-impressions: They can flee some, seek others. And they can even effect changes in things. Now humans have this ability to a much greater degree; but nevertheless, the course of our ideas would still not gain its full freedom from this ability alone…without symbols we would scarcely lift ourselves to conceptual thinking. Thus in applying the same symbol to different but similar things, we actually no longer symbolize the individual thing, but…the concept. This concept is first gained by symbolizing it; for since it is, in itself, imperceptible, it requires a perceptible representative in order to appear to us.[[38]](#endnote-38)

Frege’s final sentence above anticipates Wittgenstein’s claim in *Tractatus* 3.1 that ‘In a proposition a thought finds an expression that can be perceived by the senses’. But he initially focuses on the action- and movement-guiding role of sensory experience sketched just above (‘flee some [sense-impressions], seek others’). So like Helmholtz, and perhaps following his example, Frege took concepts to be linguistically enhanced extensions of natural representations that guide movement via sensory consequences they make it possible to predict. Thus it seems plausible to suppose that when Frege speaks of the way ‘our ideas gain full freedom’, he is extending Helmholtz’s claim that ‘Having learned how to read these symbols we are enabled by their help to adjust our actions so as to bring about the desired result, that is, so that the expected sensations will arise.’[[39]](#endnote-39)

As the term ‘symbols’ suggests, Helmholtz’s influence is also apparent in Wittgenstein’s picture theory. The Viennese physicist Ludwig Boltzmann discussed scientific models in a number of his publications, including an *Encyclopaedia Britannica* article in 1903[[40]](#endnote-40). These discussions make clear that Maxwell and Helmholtz were regarded as leading figures among a whole series of scientists, including Boltzmann himself, who developed the notion of a model or *Bild* for the exposition of physical theory.[[41]](#endnote-41) We can see something of Helmhotlz’s account in his discussion of the idea of a table quoted above. He says of the idea and the table that

…The one [the idea] is the mental symbol of the other [the table]. The kind of symbol was not chosen by me arbitrarily, but was forced on me by the nature of my organ of sense and of my mind. This is what distinguishes the sign-language of our ideas from the arbitrary phonetic signs and alphabetic characters that we use in speaking and writing.[[42]](#endnote-42)

And he later puts this in terms that more explicitly look forward to the conception of the *Tractatus*:

The excitation of the nerves in the brain and the ideas in our consciousness can be considered images of processes in the external world insofar as the former parallel the latter, that is, insofar as they represent the similarity of objects by a similarity of signs and thus represent a lawful order by a lawful order.[[43]](#endnote-43)

Helmholtz’s ideas on this topic were developed further by his student Heinrich Hertz, whom Wittgenstein cites at *Tractatus* 4.04. In his *Principles of Mechanics* Hertz had written

We form for ourselves pictures or symbols of external objects; and we make them in such a way that the necessary consequents of the pictures in thought are always the pictures of the necessary consequents in nature of the things pictured…[[44]](#endnote-44)

Wittgenstein’s source for Hertz, in turn, was most probably Boltzmann himself[[45]](#endnote-45), whose works Wittgenstein had in his library, with whom he had intended to study, and who in his expositions introduced another example that would be elaborated in the *Tractatus*:

Nobody surely ever doubted what Hertz emphasizes in his book, namely that our thoughts are mere pictures of objects (or better, signs for them), which at most have some sort of affinity with them but never coincide with them but are related to them as letters to spoken sounds or written notes to musical sounds[[46]](#endnote-46)

6. **Helhmoltz and analytical philosophy**.

As we have seen, Helmholtz had integrated Kantian epistemology with a scientific understanding of the nervous system by taking unconscious abductive inference to operate on the actual data collected by the sensory systems, and in such a way as to yield conscious experience of the ‘world of things’ that caused these data. Yet despite Helmholtz’s influence on Frege and Wittgenstein, the tradition of analytical philosophy that stemmed from Frege was to forego this progress entirely. As Russell reported in his (1914) *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*,[[47]](#endnote-47) the basic notions to be used by the new ‘logical-analytic method in philosophy’ of which ‘the first complete example is to be found in the writings of Frege’ were those of logic on the one hand, and sense-data on the other.

These sense-data were ‘what is actually given in sense’ and so despite being constitutents of ‘private spaces’ or ‘private worlds’, were ‘the hardest of hard data’ upon which empirical knowledge could be based. Frege likewise espoused ‘an inner world distinct from the outer world…of sense-impressions…of the contents of consciousness’. He indicated the primary epistemic status of these inner contents by saying

I cannot doubt that I have a visual impression of green but it is not so certain that I see a lime-leaf. So, contrary to widespread views, we find certainty in the inner world while doubt never altogether leaves us in our excursions into the outer world.

And he and Russell were of course agreed that this conception might seem to render knowledge of the external world impossible. Thus Frege again:

If we call what happens in our consciousness idea, then we really experience only ideas but not their causes. And if the scientist wants to avoid all mere hypothesis, then only ideas are left for him, everything resolves into ideas, the light-rays, nerve-fibres and ganglion-cells from which he started. So he finally undermines the foundations of his own construction…. [[48]](#endnote-48)

This adhered to an understanding of perceptual experience that went back to Descartes, and that had apparently been refuted twice, first by Kant and secondly by Helmholtz and other sensory physiologists. As Kant had argued, the actual data should be seen as prior to conscious experience, and not as given but rather as actively synthesized by the use of concepts. As Helmholtz had argued, the conscious experiences thus synthesized had the epistemic status (not of data but) of hypothetical abductive conclusions about the probable causes of the data. Moreover, as Helmholtz had shown in his work on illusions, these conclusions were liable to systematic error, explicable in terms of the physiological processes by which they were produced.

All this indicated that the experiences that Russell and Frege regarded as sense-data ‘in our consciousness’ were not in fact ‘given in sense’, and did not have the status of ‘the hardest of hard data’ (or even of data) for the basic epistemic workings of the mind or brain. Rather their traditional empiricist status as person-level conscious data depended upon their integration with the subpersonal neural mechanisms whose working Helmholtz had begun to describe. So when Helmholtz put the matter in traditional empiricist terms, he described ‘our knowledge of the actual world’ as resting ‘upon experience, with constant verifications of its accuracy that we perform with every movement of our body’.[[49]](#endnote-49) In light of this there should have been no question of thinking of such knowledge as ultimately based on data other than those actually collected by the sensory systems; nor of considering, as Frege did, that for the scientist seeking data as opposed to mere hypotheses, ‘everything…the light rays, nerve-fibres, ganglion cells’ resolves into ‘what happens in our consciousness’. But as things went, Russell and Frege’s picture of sense-data – the picture, as Wittgenstein would later say, that ‘forces itself on us at every turn’ -- set the stage for decades of work in analytical philosophy (see Huemer 2011[[50]](#endnote-50)), and for decades of work on the part of Wittgenstein himself. As noted earlier, he finally rejected the conception – and, unlike most philosophers, the picture of the internality of consciousness of which it was a part -- via the account of correctness for the application of concepts that he developed through his own use of radical interpretation.[[51]](#endnote-51)

Over the same period analytical philosophy also ignored Helmholtz’s account of abductive inference in the nervous system. Philosophers began giving serous consideration to such an idea only after Chomsky reintroduced it in describing language-learning as requiring ‘what from a formal point of view is a deep and abstract theory…[the] concepts and principles of which are only remotely related to experience by long and intricate chains of unconscious inferential steps’[[52]](#endnote-52) And finally – over a century after Helmholtz’s work, and as he said, ‘now that we have stopped dreaming of reducing science to sense-data’ – Quine independently introduced the kind of epistemology that Helmholtz had practiced, together with some of the ideas Helmholtz had advanced.

**7. Quine, Davidson, and naturalized epistemology**

According to Quine’s (1969)[[53]](#endnote-53) account

Epistemology still goes on, but…as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz, a physical human subject…This human subject is accorded a certain experimentally controlled input…certain patterns of irradiation in certain frequencies, for example, and in the fullness of time the subject delivers as output a description of the three-dimensional world and its history…

We study the relation between input and output

in order to see how evidence relates to theory, and in what ways one’s theory of nature transcends any available evidence…We are studying how the human subject of our study posits bodies and projects his physics from his data, and we appreciate that our position in the world is just like his…

And this is epistemologically significant because

Our talk of external things, our very notion of things, is just a conceptual apparatus that helps us foresee and control the triggering of our sensory receptors in the light of previous triggering of sensory receptors. The triggering, first and last, is all that we have to go on.

Quine is surely right that sensory input (and as opposed to the philosophical phantasm of sense-data) can be taken to provide a basis for experience, belief, and knowledge. (This is also illustrated by the use of sensory input in ‘brain-in-vat’ scenarios that seek to render sense-data skepticism consistent with acknowledging the metaphysical dependence of conscious experience on the physical brain.) Like Quine, Davidson held that there is ‘a causal bridge that involves the sense organs’ between ‘external events and ordinary beliefs’; and also that it was an error to try to cross this bridge ‘with sense data, uninterpreted givens, or unwritable sentences [experiential protocols] as its impossible spans’.[[54]](#endnote-54) But he was willing ‘to count myself a naturalized epistemologist’ only in the sense that epistemology should be third-person and ‘externalized’,[[55]](#endnote-55) as in his project of radical interpretation.

Davidson regarded thought was holistic because the application of one concept to an object of belief or perception involved the application of many others, and externalistic because in basic cases the contents of thoughts were determined by the objects and events that caused them. So when he compared his own views with ‘what we now know about neurons, neural nets, and the processing of information (so-called) in the brain’, he emphasized how three central and partly *a priori* features of his account – holism, normativity, and externalism – might be thought to distinguish it from ‘serious’ science.[[56]](#endnote-56) But since the neuroscience with which we are concerned describes the processing of (Shannon) information via concepts that determine the contents of perceptual experiences by specifying the objects and events that cause them, it has these features as well.

We can see this in the way the hypothesis-generating conceptual level that dominates cortical processing always meets the disparate sensory data it is gathering with a current hypothesis tacitly involving the whole range of objects and events conceived as possible objects of perception. A recent study of ongoing cortical conceptualization during the perception of a range of familiar objects and events coheres with such a view. Huth, Nashimoto, Vu, and Gallant (2012) [[57]](#endnote-57) mapped a series of short films to 1705 hierarchical categories of the objects and events that appeared in them, as taken from the WordNet lexicon. They then mapped the categories to the cortical activity that arose as viewers recognized (categorized) those objects and events while watching the films. Under statistical analysis the categories could be seen to correspond to locations in an apparently holistic representation – a continuous multidimensional semantic space -- spread smoothly over the cortical surface. Although far short of a precise tracking of the cortical activities involved, this has a *prima facie* claim to be taken as an empirical illustration of Davidsonian holism and externalism, as realized in the processing of sensory data.

**8. Epistemology, evolution, and interpretation.**

Philosophers have often criticized Quine for giving up the attempt to evaluate claims about these topics, and seeking instead merely to describe them. But if we take his suggestion in terms of the program sketched above this will not be so. Rather, the abductive inferences by which a human subject ‘posits bodies and projects his physics from his data’, as studied in the example above, should be evaluable by the standards of Bayesian epistemology (Talbott 2011).[[58]](#endnote-58) Still, and in the same naturalistic perspective, we should also assume that the overall function of brain’s modeling of the world is to promote reproductive success; and this would entail that its activity might further that end, as opposed to approximating the real causes of sensory data.

Thus we might expect a high degree of veridicality for the basic perceptual judgments Helmholtz and Davidson emphasize, and that figure in the study above. But we should also expect the kind of systematic susceptibility to perceptual illusion Helmholtz found for situations that evolution had not prepared us to cope accurately with.[[59]](#endnote-59) Likewise, and turning to the naturalized epistemology of interpretation, we might expect a high degree of accuracy in the cognitive understanding of language, upon which so much of the rest of our thinking depends. But as regards the understanding of persons and their actions, we might also expect something like the kinds of self- and group-serving biases we manifest in understanding ourselves as opposed to others, or again in understanding ingroups with whom we co-operate, as opposed to outgroups with whom our ingroups compete.[[60]](#endnote-60)

**9. Radical interpretation in the perspective of neuroscience.**

Davidson’s project was ‘not to describe how we actually interpret, but to speculate about what it is about thought and language that makes them interpretable’[[61]](#endnote-61). At most, he argued, his account might provide ‘a model [or ‘satisfactory description’] of an interpreter’s linguistic competence’.[[62]](#endnote-62) This, again, was ‘not about the details of the inner workings of some part of the brain’. Still he regarded it as trivially true that ‘if the theory does correctly describe the competence of an interpreter, some mechanism in the interpreter must correspond to the theory’.[[63]](#endnote-63)

Many influenced by Davidson, from Gareth Evans onwards[[64]](#endnote-64), have tried to describe the corresponding mechanisms in more detail, and in sub-personal terms that admit neural realization. The idea that the corresponding neural systems themselves employ models and hypotheses should encourage us to take this further, since it raises the question how far the models and hypotheses tacitly employed by the brain might correspond to those Davidson explicitly provided.

**Commonsense psychology and truth-conditional semantics**

In *Ways of Meaning* Mark had argued, in accord with John McDowell, that a theory of truth ‘will only be part of an overall theory of understanding’ and one whose contribution required to be assed via ascriptions of attitudes like desire and belief.[[65]](#endnote-65) This corresponds to the two main components of Davidson’s final model: a (Bayesian) decision theory, cast in terms of sentences, yielding degrees of belief and strengths of desire from preferences revealed in choice; and a theory of truth for interpreting the sentences ordered by the theory.

In Davidson’s minimalistic and non-question-begging ‘official’ procedure for radical interpretation, the interpreter systematically compared the interpretee’s choices as to which of a pair of uninterpreted sentences the interpretee would prefer to be true. The ordering of preferences over sentence enabled her to assign subjective probabilities and utilities to them, and so to take them as expressions of belief and desire. She could assign semantic content to these by finding ‘what episodes and situations cause [the interpretee] to prefer the truth of one sentence to another’.

This assumed both rationality (e.g. transitive choices) and first-person authority (choices that correlated with non-verbal action) on the part of the interpretee. In the simplest and most basic case, for example, a sentence construed as expressing a desire on which an interpretee would act would be one that would be made true by her acting. Such a sentence would be assigned relatively high utility and low probability immediately prior to action, and relatively high probability and low utility afterwards. The corresponding changes in preferences would be mapped to the events that caused them, that is, to the changes in the interpretee’s nervous system, voluntary movements, and the targets of these in the environment that constituted the satisfaction of her desire.

Davidson later said that this ‘technically rather byzantine’ procedure could be brought ‘closer to psychological reality’ by integration with everyday interpretive understanding. [[66]](#endnote-66) In this we could regard ourselves as cross-checking hypotheses about the truth-conditions of an interpretee’s utterances, taken as authoritative expressions of belief and desire, against hypotheses as to the beliefs and desires explaining her non-verbal acts (those that ‘speak louder than words’). Given an interpretee’s co-operation (required also for the ‘official’ procedure) this would enable us to test our understanding of the meanings of her sentences against our understanding of the beliefs and desires upon which she acted as often, and in as varying circumstances, as we chose to arrange, and thereby also to confirm the extent of the first-person authority upon which we were relying. In this way we might attain the kind of precision in understanding one another’s language that we seem to enjoy.

**Interpretation and neuroscience**

So far we have considered Davidsonian hypotheses about the truth-conditions of sentences on the one hand, and ascriptions of desire and belief in the explanation of action on the other. It seems reasonable to take these (as well as many other accounts) as related by the claim that the use of concepts in framing models and hypotheses about the causes of sensory data is central to the working of the brain. Above we briefly considered a study describing the brain as forming (holistic) conceptual representations of objects and events as they are perceived. Similar studies could refine our understanding of the mental and neural representation of words and sentences. And roughly, insofar as we use the same concepts of things in synthesizing perceptual experiences of them as in forming desires and beliefs about them, then accounts of language that link words with concepts should also link them with their referents (and vice-versa), and the contents of desires and belief about these should coincide with the referential truth-conditions of the sentences we use to ascribe them.

Davidson also emphasized that interpretation was successful just when a hearer understood a speaker’s utterance in the way the speaker had desired that it be understood. He took this to show that a semantic theory construed as a ‘portable interpretation machine’ -- that is, ‘a machine which, when fed an arbitrary utterance (and certain parameters provided by the circumstances of the utterance) produced an interpretation’ – was insufficient to yield the required interpretive matching. [[67]](#endnote-67) This emphasis on interpreters’ co-ordination with interpretees’ desires distinguishes Davidson’s account from many versions of truth-conditional semantics that have followed in his wake,[[68]](#endnote-68) and much work in the philosophy of language has been devoted to understanding how such matching might be achieved.[[69]](#endnote-69) On the present account the matching is produced by the hearer’s tacit abductive modeling of the speaker and her desires in speaking – but working in such a way that the hearer fulfils the speaker’s desires without explicitly interpreting them.[[70]](#endnote-70)

All this relates interpretation, radical and everyday, to the interactive abductive understanding -- of one another’s desires, beliefs, and other motives -- that underpins human group co-operation. Helmholtz/Bayes neuroscience casts light on this as well. We saw above that action (including speech) minimizes prediction error by making predictions about the sensory consequences of action come true – as in speech a Davidsonian speaker makes her proprioceptive (articulatory muscular) predictions about the consequences of vocalization come true by uttering a sentence, and thereby starts to make her desire in speaking (and thereby sentences describing it) come true as well, as she determines by the further sensory consequences of hearing herself saying what she wanted to say, seeing her hearer start to understand to understand or misunderstand what she intended to convey, and so on.

For all actions an agent’s brain must model a series of causes including (i) the motives that are the internal causes of forthcoming bodily movements (in this case the agent’s desires in speaking), (ii) the bodily movements (in this case of speaking) that these are to cause (iii) the perceivable changes in the causes of the sensory input (the sounds of her own voice, the visible signs of uptake on the part of her hearer) these bodily movements are to cause (the speakers experience of the satisfaction of her desires in speaking); and (iv) the changes these produce in the internal causes with which action originates (the pacification of the speaker’s desires by the experience of their satisfaction).

This means that an agent’s brain must predict changes in both internal and external sensory data from internal causes, and changes in both internal and external causes from sensory data. For speech and other actions aimed at causing changes in others, the external causes must include the movements and actions (e.g. those of understanding) of the others to be changed. This requires internal models of the motivational causes of others’ movements which Bayesians sometimes call ‘models of inverse planning’.[[71]](#endnote-71)

The representation of commonsense psychology in these models is continuous with Davidson’s, and like his must be extended to include irrationality. Friston and his colleagues have described how such models can be framed via the brain’s re-using of the predictive capacities required for intentional action. [[72]](#endnote-72) As such neuroscientific ‘inverse models’ are framed via the use of concepts, they combine elements of different philosophical approaches (simulative and theoretical) in a single account. Here again the explanatory framework of neuroscience suggests paths for progress in the philosophy of mind and language.

**10. Life and Death in Philosophy.**

In the perspective sketched above there is no question of the death of the philosophical program of which *Ways of Meaning* is a part. Rather the philosophy practiced in that book, like the work of Quine and Davidson themselves, should be seen as returning analytic philosophy to the Kantian and scientific roots from which it first drew nourishment, by renewing those that were severed by its founder’s adherence to the notion of sense-data. And if this program, and analytical philosophy with it, are increasingly integrated with neuroscience, this is surely all to our good. In this case the fate of our theories will be, as Einstein said, the fairest for which mortal theories can hope: to live on as part of those that succeed them.

**Appendix: on an objection to Helmholtz**

Philosophers frequently criticize Helmoltz and those who take themselves to have learned from him for misusing the notion of prediction. The criticism is that to speak of tacit prediction (or again of ‘the ‘predictive brain’, ‘predictive coding’, ‘predictive processing’, etc.) is to misuse the notion of prediction. This notion, the criticism goes, can rightly be applied only to the conscious activities of human beings, such as the scientists Helmholtz describes as predicting the positions of celestial bodies. It cannot rightly be extended to representation *per se*, nor to hypothetical unconscious activity of the mind, nor again to the brain, the visual system, or any other subpersonal thing or process.

Since this criticism might be directed against the argument above, as well as many other claims in recent scientific discourse, it may be worth disarming it. This can be done by specifying a notion of predictive representation that is free from confusion, and that we can take to underpin the uses in question. Ruth Millikan’s notion of biosemantic representation, which can be brought into the neuroscience and philosophy of mind and language we are considering, serves this purpose well.

**Biological representation**

As is familiar, Millikan takes representations in living things to be concrete particulars – characteristically events in behaviour or physiology -- produced by certain physiological (producer) mechanisms, and used by other (user) mechanisms. These have been selected for co-operating in discharging some further evolutionary function of the users, often in the governance of behaviour, but which differs from case to case. The items so produced and used qualify as *representations* because their role in the overall process requires that they bear a particular kind of mapping – such as might be expressed in a rule of representation like ‘Socrates’ stands for Socrates or ‘man’ is satisfied by men -- to a range of other items, those that, intuitively, they *represent*.

So in this account representation is explicated as a kind of rule-governed mapping. Representations must map in particular ways to the things they represent, and it is in light of such mappings – e.g. those that were discovered via von Frisch’s ‘radical interpretation’ of the dances of honeybees - that we regard the former as representing the latter. We can readily apply this idea to non-biological entities that we take to use representations, such as computers or robots. We might design a mobile robot that used a video camera to produce representations of the space and objects around it, where these were to be used to enable the robot to navigate the objects, say by motors that drove wheels on which it was mounted.

For this we would require the patterns produced by the camera to map to the objects to be navigated, and in the particular ways that would enable the motor mechanisms to use the patterns in successful navigation. The particular representational patterns produced by the camera, that is, would be required to map to motoric mechanisms in ways that enabled the function for which this representation producing and using system was designed. This is in turn would distinguish representations (outputs from the representation-producing camera) that were accurate or correct from those that were not, as an account of representation seems required to do. But in this case the status of the patterns as representations would depend upon our having designed them to fulfil a particular role, and hence upon our thoughts and intentions as expressed in the speech and writing that co-ordinated the project. This would mean that no light would be cast on these more basic forms representation.

In Millikan’s account, by contrast, the representational roles of states of mind like thoughts and intentions is also explained. Representation is seen as a natural biological phenomenon, such as we expect to find in our own and other animal brains. Taking a still broader perspective, we can see Millikanian representation-producing and using devices as having evolved to subserve the most basic processes of life, including the replication, expression, and modification of genetic material by producer and user mechanisms in individual cells. These representational processes ramify throughout the whole functional mesh of communicating cellular systems that make up a complex multicellular organism.[[73]](#endnote-73) Such natural forms of representation – and particularly those realized in the nervous system, as in perception, thought, and action – provide the foundations of the further representational practices established via learning and in culture, and including the production of representational artefacts like written texts, abacuses, and computers, and so on.

**Representation in single-celled organisms.**

Millikan often uses the example of the dances of honeybees, the parameters of which map to the locations of pollen, water, and other resources to be exploited by the hive. Still more basic examples are found in the simplest of living things, organisms that consist of a single cell. Each of these is a distinct individual, with a body bounded by a protective membrane housing the genetic material and the mechanisms of replication, expression, and modification that implement the vital processes by which it maintains and reproduces itself. For survival and reproduction these cells, like all other organisms, must extract and incorporate nutrients from the environment, and also avoid toxins and other dangers that might impinge from there.

Thus consider the *e coli* bacteria that swim in our guts. These are far simpler than the cells of our bodies (about a thousandth of their size, as they lack the mechanisms of differentiation and co-operation required for multicellular life). Their bodies are continually battered by random molecular movements in the fluids around them, so they must move non-randomly to ensure nutrition and safety. They do this by activating tiny ‘flagellator motors’, directed by ‘sensors’ – transmembrane chemoreceptors – that use current chemical gradients in the environment to predict future gradients expected on current motion. These ‘sensors’ release ‘signalling proteins’ that drive the motors and the cell forward towards nutrients, but pause for toxins, letting the cell ‘tumble’ out of dangerous paths.[[74]](#endnote-74)

As their designation implies, the motor-directing ‘signalling proteins’ are representations. Their production maps to, and so represents, the nutrients and toxins ahead. Accuracy in this is required for *e coli* reproductive success, which the organism accomplishes by its simpler version of the mechanisms that replicate and use DNA in ourselves. These enable the cell to duplicate its genome at something like 2000 neucleotides per second, with one ‘mistake’ per billion.[[75]](#endnote-75) This duplication, in turn, can be regarded as proceeding via representations of the kind we have been considering, with the producer and user mechanisms those that manipulate the genetic material inside the cell, and the ‘mistakes’ misrepresentations produced in this process -- tokens that fail to map to what they are supposed to represent in such a way as to enable the mechanisms that use them to perform their replicative functions.

Considered in this light the representation-guided movement of *e coli* indicates that the conception Helmholtz framed to understand human perception and movement applies far more widely. All living organisms must be ‘spatially determined’ in Helmholtz’s sense, because all must do as *e coli* do -- they must extract vital resources from the environment while avoiding dangerous impingements. Very many do this via systems of representation selected for directing their movements to these ends. Where this is so the movements can be thought of as experiments testing (often, as in the case of *e coli*, by life or death) how accurately these representations depict their environments. As Popper remarked, we differ from many other animals in co-operating to produce scientific hypotheses – explicit and shared predictive representations -- that can die in our stead.

**Predictive representation**.

So is it a philosophical mistake to describe *e coli* chemoreceptors as predicting chemical gradients of nutrients and toxins? This would seem analogous to saying that the brain, or the cortex, or representations in these, predict the sensory consequences of movements of the eye, or the movements of intentional actions. If a *reductio* of this view of the brain is possible, it might be thought provided by such examples as *e coli*, whose mechanisms of representation are far removed from those of human thought.

But the notion of prediction seems to be used in the same way in all these different cases – it is used to characterize the working of biological representations that are required to map to *future* events or states of affairs, often in complex and highly co-coordinated ways. In this the function of these representations (like descriptions of weather on the face of a barometer) is reasonably described as predictive. The guidance of biological movement is almost always effected by representations that are predictive in this sense – as these range from *e coli* protein signals through honeybee dances to the concepts that direct human saccades and intentional actions. Neither the coherence not the usefulness of this concept should be denied.

1. On radical interpretation in Wittgenstein see my (1999) 'Wittgenstein, Davidson, and Radical Interpretation' in F. Hahn, ed, *The Library of Living Philosophers: Donald Davidson,* Chicago: Open Court. Many others have also treated interpretation as a matter of scientific hypothesis. Turing, who pioneered the use of Bayesian methods in cryptography, said that 'There is a remarkably close parallel between the problems of the physicist and the cryptographer. The system on which a message is enciphered corresponds to the laws of the universe, the intercepted messages to the evidence available, and the keys for a day or a message to important constants which have to be determined.' (Quoted in Leiber, J. (1991) *Invitation to Cognitive Science*, Oxford: Blackwell) [↑](#endnote-ref-1)
2. Patton, Lydia, "Hermann von Helmholtz", *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/hermann-helmholtz/>. [↑](#endnote-ref-2)
3. Freud, S, (1895/1950) Project for a Scientific Psychology, in J. Strachey (ed.) *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol I, London: Hogarth Press, p. 311 [↑](#endnote-ref-3)
4. Lenoir, T, (2006) ‘Operationalizing Kant: Manifolds, Models, and Mathematics in Helmholtz’s Theories of Perception’, in Friedman, M, and Nordman, A, eds, *The Kantian Legacy in 19th Century Science,* Cambridge: MIT Press,p 141 – 210, quotations from p. 142 [↑](#endnote-ref-4)
5. Helmholtz, H. (1868/1971) ‘Recent Progress in the Theory of Vision, in *The Selected Writings of Hermann von Helmholtz,* Ed. R. Karl, Middletown: Wesleyan University Press, 1971, p 368. [↑](#endnote-ref-5)
6. Helmholtz, H (1855/1903)"Über das Sehen des Menschen (1855)." Vorträge und Reden von Hermann Helmholtz. 5th ed. Vol. 1. Braunschweig: F. Vieweg, 1903, p 88. [↑](#endnote-ref-6)
7. Kant, I (1787/1963) The Critique of Pure Reason, tr. Kemp Smith, N., London, Macmillian, 1963, p 224. [↑](#endnote-ref-7)
8. See Kant’s Critique of Pure Reason, as cited above, pp 308-327. [↑](#endnote-ref-8)
9. I am grateful to Leonardo Nascimento for help with Helmholtz’s language and concepts here. If we consider Helmholtz’s explication of tacit *induktionsschluesses* by reference to the explicit practices of scientists, as appears later in the text, he seems to have in mind a tacit version of the process Pierce was later to describe as abduction. His own phrase, however, also indicates the *conclusion-like* nature of the tacit reasoning he seeks to describe: perceptual experience has *the character of a* *conclusion* as to the causes of the data synthesized in it. [↑](#endnote-ref-9)
10. See Hawthorne, James, "Inductive Logic", *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.), URL =

    <http://plato.stanford.edu/archives/win2012/entries/logic-inductive/> [↑](#endnote-ref-10)
11. Douven, Igor, "Abduction", *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/abduction/>.) [↑](#endnote-ref-11)
12. See Helmholtz (1868/1971) ‘Recent Progress’, cited above, 217 ff.; and for a post-Chomskian comparison Fodor, J, 1968, “The Appeal to Tacit Knowledge in Psychological Explanation,” *The Journal of Philosophy*, 65(20): 627–40. [↑](#endnote-ref-12)
13. Helmhotlz, H (1867/2001) *Treatise on Physiological Optics*, Vol III, available as online digital document copyright University of Pennsylvania 2001, http//psych.upenn.edu/backuslab/Helmholtz, p 23. [↑](#endnote-ref-13)
14. As Lenoir explains in his (1993) ["The Eye as Mathematician: Clinical Practice, Instrumentation, and Helmholtz's Construction of an Empiricist Theory of Vision,"](http://www.stanford.edu/dept/HPST/TimLenoir/Publications/Lenoir_EyeAsMathematician.pdf) in David Cahan, ed., *Hermann von Helmholtz and the Foundations of Nineteenth-Century Science,* Berkeley: University of California Press, Helmholtz used both the principle of least action, and a least-squares account of accuracy, in describing the working of the eye. See section 5 and following. [↑](#endnote-ref-14)
15. Helmholtz, H (1878/1971) 'The Facts of Perception' in *The Selected Writings of Hermann von Helmholtz,* Ed. R. Karl, Middletown: Wesleyan University Press, 1971 p. 384. His stress on experiment appears also in the Treatise, cited above, pp 45-6. [↑](#endnote-ref-15)
16. See Frigg, Roman and Hartmann, Stephan, "Models in Science", *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition), Edward N. Zalta (ed.), URL =

    <http://plato.stanford.edu/archives/fall2012/entries/models-science/>.

    As these authors note models can be either models of phenomena or models of data, and a given scientific theory, or set of hypotheses, can function as both. I will not attempt to disentangle these issues in the text.

    [↑](#endnote-ref-16)
17. See Joyce, James, "Bayes' Theorem", *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition Edward N. Zalta (ed.), URL =

    <http://plato.stanford.edu/archives/fall2008/entries/bayes-theorem/>

    as well as Douven (2011), previously cited on abduction. [↑](#endnote-ref-17)
18. These terms were introduced by Charles Sherrington in his (1906) *The Integrative Action of the Nervous System*, New Haven, NJ : Yale University Press. [↑](#endnote-ref-18)
19. Thus see the discussion in his (1868/1971) 'Recent Progress in the Theory of Vision' in *The Selected Writings*, cited above, p.216: ‘A movement of the eye which causes the retinal image to shift its place upon the retina…In this way we learn to recognize such changes as belonging to the special phenomena which we call changes in space…’ As Lenoir explains in ‘The Eye as Mathematician, cited above, p 147, this passage is part of Helmoltz’s argument that ‘the mind is in effect performing a series of experiments with the eye, testing the hypothesis of an object’ p. 147. (Helmholtz debated the nature of the movements involved with Hering, and they remain a matter of discussion. See Coubard, O. (2013) Saccadic and vergence eye movements: a review of motor and premotor commands, *European Journal of Neuroscience*, doi:10.111/ejn.12356, pp.1 – 14.) Helmholtz applied the same ideas to touch, e.g in ‘The Facts of Perception’ in *The* *Selected Writings*, cited above, p. 377. [↑](#endnote-ref-19)
20. Diagrams gratefully reproduced from Wikimedia commons. The first is from Yarbus, A, (1967) *Eye Movements and Vision*, New York: Plenum, now online at

    <http://wexler.free.fr/library/files/yarbus%20(1967)%20eye%20movements%20and%20vision.pdf>

    The second and third appear in <http://en.wikipedia.org/wiki/Eye_tracking> and

    <http://en.wikipedia.org/wiki/Eye_movement_in_language_reading> respectively. [↑](#endnote-ref-20)
21. Helmhotlz, H (1867, 2001) *Treatise on Physiological Optics*, Vol III, available as online digital document copyright University of Pennsylvania 2001, http//psych.upenn.edu/backuslab/Helmholtz, p 23. [↑](#endnote-ref-21)
22. Dennett, D., *Consciousness Explained*, Penguin Books, 1993, p. 360-61 (Kindle Edition). It is worth noting that the multiple drafts in Dennett’s account partly correspond to the multiple levels of cortical processing in the Helmholtz/Bayes account. [↑](#endnote-ref-22)
23. Dream experience, unlike waking experience, seems constrained mainly by the arousal of memory, and this may prove very important for the understanding of dreams, symptoms, hallucinations, and mental disorder generally. This is discussed in my ‘Extending the Royal Road: from Dreams to Neuroscience’ in Boag, S., Brakel, L., and Talvititie, V. (2013, forthcoming) *Philosophy, Science, and Psychoanalysis*, London: Karnac. [↑](#endnote-ref-23)
24. *Treatise on Physiological Optics*, Vol III, as cited above, p 29,30. This mode of conscious representation seems responsible for the claim often advanced in philosophical accounts of perception, that we *directly* perceive the objects we see. But then direct perception is consistent with (indeed requires) the claim to which it is often opposed, namely that experience *represents* what we see. [↑](#endnote-ref-24)
25. *Frontiers in Psychology*, 28 May 2012, doi: 10.3389/fpsyg.2012.00151 [↑](#endnote-ref-25)
26. Clark, A., (2013) ‘Whatever Next? Predictive brains, situated agents, and the future of cognitive science’ *Behavioral and Brain Sciences* (2013) 36, 181**–**253, doi:10.1017/S0140525X12000477.

    See also the further comments and replies in *Frontiers in Theoretical and Philosophical Psychology* at

    http://www.frontiersin.org/Theoretical\_and\_Philosophical\_Psychology/researchtopics/Forethought\_as\_an\_evolutionary/1031 [↑](#endnote-ref-26)
27. For Friston’s account, and some useful exposition, see Friston, K., (2011) The history of the future of the Bayesian brain, *NeuroImage* (2011), doi:10.1016/j.neuroimage.2011.10.004. [↑](#endnote-ref-27)
28. Garson, James, "Connectionism", *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/connectionism/>. [↑](#endnote-ref-28)
29. Doya , K, Ishi , S, Pouget, A, and Rao , R. ( 2007 ). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. Boston, MA : MIT Press. [↑](#endnote-ref-29)
30. Dayan, P, Hinton, G, Neal , R, and Zemel , R (1995). ‘The Helmholtz machine’, *Neural Computation* 7, 1022 – 1037 [↑](#endnote-ref-30)
31. Dyan, P and Hinton, G, (1996) ‘Varieties of Helmholtz Machine’ *Neural Networks*, V 9, No 8, p 17. [↑](#endnote-ref-31)
32. As Friston describes this:

    …Hinton’s terribly important contribution was to cast the generally intractable problem of Bayesian inference in terms of optimization. The insight here was that the same problems that Richard Feynman (1972) had solved in statistical physics, using path integral formulations and variational calculus could be applied to the problem of Bayesian inference, namely, how to evaluate the evidence for a model. This is where free energy minimization comes in, the sense that minimizing free energy is equivalent to (approximately) maximizing the evidence for a model. Note again the underlying role of optimization, which here finessed a difficult but fundamental problem in Bayesian inference.

    These issues are discussed in ‘The Helmholtz Machine’ and ‘Varieties of Helmholtz Machine’ cited above, and in a more accessible treatment in Chapter 10 of Dyan, P, and Abbott, L, (2001) *Theoretical Neuroscience*, Cambridge: MIT Press. [↑](#endnote-ref-32)
33. Friston has developed a number of technical formulations relating to these topics. See his website at <http://www.fil.ion.ucl.ac.uk/~karl/>. [↑](#endnote-ref-33)
34. This is discussed in a number of papers on Friston’s website noted above. [↑](#endnote-ref-34)
35. In Friston’s account emotions and other subcortical sources of internal input that are *not* currently being modelled remain in a state of Bayesian suppression, and so as sources of prediction error seeking higher-level (and ultimately conscious) representation. In this Friston’s account of free energy Freud’s, as described in Carhart-Harris and Friston (2010, 2012) ‘Free energy and Freud: an update’. This is developed in more Freudian detail in my (2012) ‘Psychoanalysis Representation and Neuroscience: the Freudian unconscious and the Bayesian brain’. Both articles are in Fotopolu, Pfaff, and Conway, eds (2012) *From the Couch to the Lab: Psychoanalysis, Neuroscience and Cognitive Psychology in Dialogue*. Oxford: Oxford University Press. [↑](#endnote-ref-35)
36. Howhy, J, Roepstorff, A, and Friston, K, (2008) ‘Predictive Coding Explains Binocular Rivalry’ *Cognition* 108 (3), 687–701. [↑](#endnote-ref-36)
37. On this general account see Adams, R., Shipp, S., and Friston, K, (2012) Predictions not commands: active inference in the motor system *Brain Struct Funct* DOI 10.1007/s00429-012-0475-5. For an application of sensorimotor prediction to speech see Hickock, G., (2012) ‘Computational Neuroanatomy of Speech Production’ *Nature Reviews Neuroscience* AOP doi:10.1038/nrn3158.

    [↑](#endnote-ref-37)
38. Frege, G. (1879/1972) ‘On the Scientific Justification of a Conceptual Notation’ in Bynum, T, ed. *Conceptual Notation and Related Articles*, Oxford: Clarendon Press, p 83 – 4 [↑](#endnote-ref-38)
39. Helmholtz, Treatise, cited above, p 19 [↑](#endnote-ref-39)
40. Boltzmann, L. (1902), ‘Model’, *Encyclopaedia Britannica*, 10th Edition, Volume 30, pp 788-791. [↑](#endnote-ref-40)
41. On the transition from Helmholtz to Wittgenstein see Patten, L. (2009) ‘Signs, toy models, and the a priori: from Helmholtz to Wittgenstein’, *Studies in History and Philosophy of Science* 40 (2009) 281–289; Heidelberger, M, (1998), “From Helmholtz's Philosophy of Science to Hertz's Picture-Theory,” in Baird, D., Hughes, R, and Nordmann, A, eds*Hertz: Classical Physicist, Modern Philosopher*Dordrecht: Kluwer Academic Publishers; Barker, P (1980) ‘Hertz and Wittgenstein’ Stud. Hist. Phi/. Sci., Vol 1 No. 3, pp. 243-256; and Wilson, A. (1989) ‘Hertz, Boltzmann, and Wittgenstein Reconsidered’, Stud. Hist. Phil. Sci., Vol. 20. No. 2, pp. 245-263; Visser, H. (1999) Boltzmann and Wittgenstein or How Pictures Became Linguistic, Synthese 119: 135–156. [↑](#endnote-ref-41)
42. Helmholtz, *Treatise*, cited above, p 23. [↑](#endnote-ref-42)
43. Helmholtz (1868/1971) ‘Recent Progress in the Theory of Vision, in *The Selected Writings*, cited above, pp 186. [↑](#endnote-ref-43)
44. Hertz, H. (1894/1956). *The principles of mechanics* (R. S. Cohen, Ed.; D. E. Jones, & J. T.

    Walley, Trans.). New York: Dover, p 1. [↑](#endnote-ref-44)
45. On the role of Boltzmann see particularly Wilson (1989) and Visser (1999) cited above. [↑](#endnote-ref-45)
46. Boltzmann, L, (1897/1974) ‘Lectures on the Principles of Mechanics’ in McGuinness, B. (1974) *Theoretical Physics and Philosophical Problems: Selected Writings (Vienna Circle Collection)*, Dordrecht: Reidel, p 225. [↑](#endnote-ref-46)
47. Chicago and London: Open Court, available online via

    <http://en.wikipedia.org/wiki/Bertrand_Russell>

    The phrases quoted are from pages 68, 71-2, 90, and 89 respectively. [↑](#endnote-ref-47)
48. The quoted phrases and sentences are from Frege (1917-8/1956) ‘The Thought’ *Mind*, July 1956 from pages 299, 306, 304, and 299. [↑](#endnote-ref-48)
49. Helmholtz (1868/1971) ‘Recent Progress in the Theory of Vision’, in *The Selected Writings*, cited above, p 222. [↑](#endnote-ref-49)
50. Huemer, Michael, "Sense-Data", *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/sense-data/>. [↑](#endnote-ref-50)
51. On this see my (2012) ’Wittgenstein, Rules, and Privacy’ in Ellis and Guevara, eds, *Wittgenstein and the Philosophy of Mind*, Oxford: Oxford University Press. I have argued elsewhere that the ‘picture’ of consciousness Wittgenstein rejected is a product of evolution, precipitated by the brain’s use of metaphorical processing in representing the mind as a kind of container within the body, and one that misrepresents the genuine physical internality, subjectivity and privacy enjoyed by conscious processes as they are realized in the brain. See my (2007) ‘The Problem of Consciousness and the Innerness of the Mind’ in McCabe, M.M., and Textor, M. ed, *Perspectives on Perception*, Frankfurt: Lancaster Publishers. [↑](#endnote-ref-51)
52. Chomsky, N, (1965) *Aspects of the Theory of Syntax* Cambridge: MIT Press, pp. 57-8. [↑](#endnote-ref-52)
53. Quotations from Quine, W. V. (1969) ‘Epistemology Naturalized’, in *Ontological Relativity and Other Essays* New York: Columbia Universityy Press p. 82,83 And Quine, W.V. (1981) *Theories and Things*, Cambridge, MA: Harvard University Press, p 1. [↑](#endnote-ref-53)
54. Davidson, D. (1982) ‘Empirical Content’ in Davidson, D., (2001) *Subjective, Intersubjective, Objective*, Oxford: Oxford University Press, p 175. [↑](#endnote-ref-54)
55. See Davidson (1990) ‘Epistemology Externalized’ in Davidson (2001), cited above. [↑](#endnote-ref-55)
56. See Davidson, D. (1995) ‘Could There Be a Science of Rationality’, cited above, pp 124-5 [↑](#endnote-ref-56)
57. Huth, A., Nishimoto, S., Vu, A., and Gallant, J. (2012), A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain, *Neuron* 76, 1210–1224, December 20, <http://dx.doi.org/10.1016/j.neuron.2012.10.014>. This url also links to a video about the research, and the cagtegory-to-cortex mappings themselves can be accessed via color coding at http://gallantlab.org/semanticmovies/ [↑](#endnote-ref-57)
58. Talbott, William, "Bayesian Epistemology", *The Stanford Encyclopedia of Philosophy* (Summer 2011 Edition), Edward N. Zalta (ed.), URL =

    <http://plato.stanford.edu/archives/sum2011/entries/epistemology-bayesian/>

    Many standard arguments about Bayesian epistemology are altered by holding (as here) that conscious beliefs are formed by subpersonal abductive reasoning, the parameters governing which have been set by evolution. This may leave intact the idea that decisions will register preferences, but only as adjusted for non-rational factors such as impingements or error signals from emotions not consciously modeled; but not that cognition or the evaluation of evidence will be rational at the personal level. [↑](#endnote-ref-58)
59. For some interesting recent discussions of illusions about the body and self in Bayesian terms see Hohwy J, Paton B (2010) Explaining Away the Body: Experiences of Supernaturally Caused Touch and Touch on Non-Hand Objects within the Rubber Hand Illusion. *PLoS ONE* 5(2): e9416. doi:10.1371/journal.pone.0009416, and Apps, M. and, Tsakiris, M., *The free-energy self: A predictive coding account of self-recognition*, Neuroscience and Biobehavioral Reviews (2013), doi:10.1016/j.neubiorev.2013.01.029. [↑](#endnote-ref-59)
60. Some of these are discussed in evolutionary and psychoanalytic terms in my (2003) ‘Evolution, Emotion, and Conflict’ in M. Chung, ed., *Psychoanalytic Knowledge*, London: Macmillan: Palgrove Press, and (2004) ‘Conscience and Conflict: Darwin, Freud, and the Origins of Human Aggression’ in D. Evans and P. Cruse, ed, *Emotion, Evolution, and Rationality*, Oxford: Oxford University Press. [↑](#endnote-ref-60)
61. Davidson, D., (1995), ‘Could There Be a Science of Rationality’, in Davidson, D., (2004) *Problems of Rationality*, Oxford: Oxford University Press, p 127. [↑](#endnote-ref-61)
62. Davidson, D (1986) ‘A nice derangement of epitaphs’ in Davidson, D., (2006) *The Essential Davidson*, ed Lapore, E., and Ludwig, K., Oxford: Oxford University Press, p 256. [↑](#endnote-ref-62)
63. Davidson, D (1986) ‘A nice derangement…’ cited above, p 256. [↑](#endnote-ref-63)
64. Evans, G. (1981) 'Semantic Theory and Tacit Knowledge', in Holtzmann, S., and Leich, C. Eds, *Wittgenstein: To Follow a Rule*, London: Routledge p. 118-37. [↑](#endnote-ref-64)
65. Platts, M. de B. (1979, 1997) *Ways of Meaning,* 2nd Edition, Cambridge: MIT Press,pp 59, 62 [↑](#endnote-ref-65)
66. ‘technically rather byzatine’ is from Davidson, D., (1995), ‘Could There Be a Science of Rationality’ cited above; ‘closer to psychological reality’ from Davidson, D., (1999) ‘Reply to Jim Hopkins’ in Hahn, F., ed, *The Library of Living Philosophers: Donald Davidson* Chicago: Open Court, p 286. The method for bringing closer is described in the same volume at p 276-80. [↑](#endnote-ref-66)
67. The quoted phrases are from Davidson, D. (1986) A nice derangement…, cited above, pages 256 and 264. The argument about matching interpretation with speakers’ desires is continued in Davidson (1994) ‘The Social Aspect of Language’ in McGuinnes, B, and Oliveri, G., eds, *The Philosophy of Michael Dummett*, Dordrecht: Kluwer. [↑](#endnote-ref-67)
68. For an account of this program as it has evolved since *Ways of Meaning* see Lepore, E, and Ludwig, K (2007) *Donald Davidson’s Truth-theoretic Semantics*, New York: Oxford University Press. For an extension into Chomskian linguistics and cognitive science see Larson, R., and Segal, G. (1995) *Knowledge of Meaning: an Introduction to Semantic Theory*, Cambridge: MIT Press. [↑](#endnote-ref-68)
69. See for example Barba, J, (2009) ‘Formal Semantics in the Age of Pragmatics’, *Linguistics and Philosophy* (2009) 30:637–668 DOI 10.1007/s10988-008-9031-4 and Borg, E. (2012) *Pursuing Meaning*, Oxford: Oxford University Press. [↑](#endnote-ref-69)
70. There is neuroscientific support for accounts of this kind in the case of irony. See Sportono, N., Kourn, E, Prado, J., Van Der Henst, J., and Noveck, I, (2009) Neural evidence that utterance-processing entails mentalizing. The case of irony’ NeuroImage 63, 25**–**39 [↑](#endnote-ref-70)
71. See Baker, C., Saxe, R., and Tenenbaum, J. (2009) Action understanding as inverse planning Cognition, doi:10.1016/j.cognition.2009.07.005 [↑](#endnote-ref-71)
72. See Friston, K, Mattout, J., and Kilner, J., (2011) ‘Action understanding and active inference’ *Biol Cybern* 104:137–160 DOI 10.1007/s00422-011-0424-z, and the previous articles cited there. [↑](#endnote-ref-72)
73. On this see Shapiro, J., (2011) *Evolution: A View from the 21st Century*, New Jersey: FT Press Science, who describes cellular working and intercellular communication in representational terms drawn from symbolic information processing; and the specifically Millikanian but far-reaching argument in Shea, N., (2012), 'New thinking, innateness and inherited representation', *Philosophical Transactions of the Royal Society B*, 367, pp. 2234-44 [[open access pdf]](http://www.philosophy.ox.ac.uk/__data/assets/pdf_file/0014/28103/Shea_New_thinking_innateness_PhilTrans12_OA.pdf) [[link]](http://dx.doi.org/10.1098/rstb.2012.0125), and ‘Representation in the genome and in other inheritance systems’, *Biology and Philosophy* (2007) 22: 313-331 DOI 10.1007/s10539-006-9046-6. [↑](#endnote-ref-73)
74. This is described in interesting detail at

    <http://chemotaxis.biology.utah.edu/Parkinson_Lab/projects/ecolichemotaxis/ecolichemotaxis.html>. [↑](#endnote-ref-74)
75. For this see Shapiro, cited above, p. 10.

    [↑](#endnote-ref-75)