Chris Hamblin

Mind the Gap;
Existential Phenomenology and Cognitive Science

***

Imagine one of the world's great pianists has a deaf son. After years of lamentation and frustration, he sees a glimmer of hope. The young boy displays a prodigal capacity for mathematics, which the father sees as a medium through which the son might finally come to understand his music, his life's work. The pianist is quite inept himself at mathematics, and subsequently music theory; he can't read sheet music and he has great difficulty identifying the chords and scales he is playing. The extent of his disability in such regards is so great that most of his contemporaries regard him as a savant. His musical genius comes from his remarkable ear and his creativity; he can recreate instantly anything he hears, it flows from his fingers as naturally as words from our mouths, and he always has so much to say. All the same, his musical colleagues, whom he holds in high esteem, insist that there is a strong mathematical foundation for the structure of music. The son begins rigorous training with these colleges, learning about frequencies, intervals, principles of tonal theory, and even music psychology. The boy is a quick and willing learner, and the father is immensely pleased. He listens with teary eyes when the son sight-reads piano sheet music, one of the many skills he is developing.

One night the boy comes to his father's room and expresses that he is deeply unsatisfied and cannot continue with his musical training. The father is shocked.

“But why!? Is there something you don’t understand, that’s too difficult? You are learning so quickly and improving so much!”

“No . . .” replies the boy, “I understand everything they tell me, I can identify all the concepts, I can do all the work, its just . . . well what’s the point of all this! It’s a made up, meaningless exercise in symbol manipulation, and very loose symbol manipulation at that. It has no direction or motivation. All these definitions and rules, scales, keys, chords, progressions, songs, they fit together . . . but everything is so arbitrary! Like a jigsaw puzzle made from a grey square. Why did we cut it up the way we did? Why did we cut it up at all? I’m sorry, I know it’s important to you, but I just can’t keep pretending to care about this meaningless game.”

The boy leaves and his father is crushed. He sits at his piano bench and channels his sadness into his craft, entering a state common to the masters of music. He does not try, he does not think, he loses himself. The music simply *is,* all around him, his world.

***

The above characters will serve as a dramatized metaphor for two possible characters that emerge in regards to philosophy of mind. Where the subject matter of the characters described above is music, the subject matter of these characters is far more general and elusive; consciousness, or worse; understanding, or worse still; meaning. The choice of a musical metaphor was arbitrary in this regard; it was just a particular manifestation of a general theme—The pianist could be a great painter and his child could be *Mary the Color Scientist* (Jackson 1982)*,* whom he wishes to teach the *true* nature of color. Nonetheless, I will return to this music example frequently,

as a clear instantiation of the dual perspective this essay tries to make sense of and navigate, as something necessarily both a reducible system of patterns and an immediate perceptual holism. To learn music theory is to learn the rules to a game that in some respect you already knew how to play. When it comes to understanding the mind/brain, one must undergo a naturalistic and existential development that mirrors the development of the musician. For the typical musician, music theory develops concurrently with his musical ear, they feed off one another fluidly, his ear calls for a certain theory, and the theory in turn prompts him to listen in a certain way. The budding musician would be greatly hindered if he polarized himself and went forward on the basis of his ear or the theory alone, effectively rendering himself either the pianist or his son. Similarly, I argue that existential phenomenology calls for a certain naturalistic account of the brain, and that developing naturalistic account calls for a certain existential phenomenology in turn. I warn against the parallel polarization, embracing naturalism at the cost of losing existential phenomenology, renouncing our privileged access and denying oneself of its indispensible utility. Neither must we embrace phenomenology at the cost of losing naturalism, denying the possibility that we are highly complex machines.

Such polarized positions emerge in light of philosophy of mind's classic difficulty, the seemingly unbridgeable divide between the mind and the brain. We might view the positions as two faced-off, with both feet planted firmly on one side of the explanatory gap. In 1983, Joseph Levine characterized this gap by noting our difficultly in understanding the claim 'pain is the firing of C-fibers' (Levine 1983). Such a characterization presents the gap in terms of a particularly tricky, seemly

impossible Frege puzzle (Zalta 2014). To use Fregean terms; 'pain' and 'the firing of C fibers' are so drastically different in *cognitive sense* that the identity statement 'pain = the firing of C fibers' seems nonsensical and arbitrary. Where 'heat = the movement of molecules' is a necessary analytic truth (when one does not conflate it with the claim 'warmth = the movement of molecules') true in all possible worlds, pain and the firing of C fibers have such distinct modes of presentation for us that we can easily imagine one without the other. Here, 'pain' and 'the firing of C fibers' are fill-ins for two seemingly incommensurable sets, the first, the immediate subjective features of experience, what have been termed qualia, and the second, subsets of the natural world; systems of objects in causal relations.

It is in response to such identity statements that the characters I am concerned with emerge; they are in effect characters that try to solve this Frege puzzle by denying the possible truth of any identity statements constructed as such. The first way of doing this is by accepting some sort of ontological dualism; there is 'pain' and there is 'the firing of C-fibers', but they are of different kinds and thus 'pain is not the firing of C fibers'. For reasons that should become clear in this paper, I find such a position defeatist and unsatisfying. As such this paper will address more moderate positions. The second way of renouncing the puzzle is again ontological; one simply denies the existence of one of the sets causing the issue. Strict eliminative materialists deny the first-person set; 'pain' becomes an empty term, and we can subsequently dismiss all identity statements that try to reference it. Similarly one could deny the third-person set, a move taken by metaphysical solipsists and panpsychists. Again,

these positions seem radical and unsatisfying; my concern is with something more moderate.

The moderate position is this; if we are to come to a coherent understanding of consciousness, it is necessary that we commit some sort of bracketing in our methodology. Again this bracketing comes in two forms, bracketing of the third-person, objective world, and bracketing of the first-person, subjective world. We see this sort of bracketing in the foundation of the phenomenological discipline, in Husserl's epoché, where assumptions about the existence of the external world are suspended, granting us access to the unadulterated, immediate content of subjective experience (Beyer 2013). Daniel Dennett makes an opposite bracketing in his conviction that we approach consciousness as heterophenomenologists, which he himself likens to the 3rd person parallel to Husserl's epoché (Dennett 2003). The heterophenomenologist regards the subject's first person world as a theorist's fiction, and studies it as an anthropologist might study the god of a forest tribe. He regards everything he is told with an air of skepticism, and is not compelled to explain the features of the god (the phenomenology) directly, but only why the subjects purport the existence of those features. The emphasis is thus placed on verbal report; for ever phenomenal question "why do I experience X?" there is a corresponding heterophenomenal question "why does the subject report that they experience X?" We could see both bracketings as an attempt to define a new project, to work within a closed, coherent system into which such troubling Frege puzzles do not enter, as the entirety of the project stands on one side of 'the gap'. This enables us to move on with

a study of consciousness, without being stonewalled by foundational philosophical difficulties.

I am very interested in 'moving on' in this regard; I am finishing my undergraduate education and will likely step away from philosophy into the fields of neuropsychology or cognitive science. With philosophy I often fear I am simply smashing my head against a truly insurmountable wall at the bottom wrung of the abstraction ladder, that philosophy cannot be overcome and has no solutions, that it is philosophy precisely in virtue of these characteristics. I am suspicious there are fundamental limitations faced by the mind trying to fully understand *itself* (perhaps due to a paradox of self-reference, but that is for another paper), and that the philosophical mind is one obsessed with this limitation, spinning its wheels in a hopeless effort to look over the edge and see how 'it *all* hangs together'. Alas the same sort of problem keeps popping up; we always find ourselves appealing to something mysterious, because reductive explanations must bottom out, so we throw up our hands and say "I'm really not sure what I'm talking about in terms of the system, but thankfully you know what I mean or we never would have been able to get it off the ground!" It would seem a system only has meaning against a *background* of sense, that it must be seen from some vantage point that it does not contain and does not explain. Thus emerges a general sense of duality, between the background and foreground, the interior and exterior, the subject and object, and a principled 'gap' between them. The 'explanatory gap' as I presented it is just one manifestation of this more general difficult, where the system observed is *the mind* viewed as a subsystem of the natural world, and what is left out is *qualitative experience.* Of course there are

other manifestations, as they arise say in philosophy of mathematics and language (to which philosophy of mind is no doubt intimately bound), such as the *rule following paradox* (Kripke 1982). There, it would seem that reason, when conceived as a formal system of rules, only makes sense against some background of normativity that it cannot explain, that the primitive rules on which everything rests amount to nothing more than our innate ability to go on in the *right* way. I find the example of music so helpful because it highlights this parallel between the rule following paradox and the explanatory gap. The musician's son is a character who has been stripped of the background from which music is given a sense, a background that can be easily understood as either the *experience of hearing* the music, or the normative force that motivates and directs the music theoretic rules.

For now, I feel compelled to leave *the difficulty* at that, which perhaps opens me up to the criticism that I've done little more than characterize a vague existential confusion. I could dive in and try to unpack it further, but as I said before, I am concerned with 'moving on'; I don't want to spin my wheels against the wall! Of course in moving on to cognitive science I am ultimately engaged in a rarefied version of the same troublesome project--I'm a brain trying to understand itself--and as such philosophy and the gap will loom ever close at hand. What then is a fruitful methodology for the cognitive scientist in light of this gap? In the first part of this essay I will argue against a methodological bracketing such as those suggested by Husserl and Dennett, which run the risk of 'losing naturalism' and 'losing phenomenology' I brought up earlier. We cannot avoid a certain degree of *perspectival*

dualism[1], but more important to this essay, I suggest that we *should not* try and avoid it. When we close ourselves off in a single discipline, what we gain in coherence we may sacrifice in creativity, by losing sight of modes of thought that can work to pull our own discipline forward.

In part two I will present my own methodology, mostly demonstratively, which embraces this dual perspective. It is a project I feel would be disabled, or even disallowed, by the bracketing suggested in part one. On the one hand I will take very seriously purely functional machine schematics that are supposed to underlie mentality. On the other I will take very seriously the phenomenologist's characterization of lived experience, of our *being-in-the-world*. Furthermore, I will take seriously their justification for such a characterization, which is validated only to the extent that I live it out, that their words resonate with me and I find myself anew through them. A fruitful method will depend then on the extent to which these two disparate projects can influence each other. The approaches are obviously far afield, the domain of the engineer and the continental philosopher respectively, and thus constructive discourse has often proved difficult, or even been openly shunned. At the heyday of the Artificial Intelligence project, Hubert Dreyfus, drawing on the work of Heidegger and Merleau-Ponty, put forth a sweeping criticism of AI, sparking

---

[1] As opposed to ontological dualism--which is very problematic and seems to disregard the possibility of a *cognitive science* from the start--the perspectival dualist may very well agree that more or less the mind is the brain (or if swayed by externalism that it at least has some naturalistic counterpart), but as we are so curiously positioned, *being* that very mind/brain, we inevitably conceptualize the mind and brain as distinct.

much hostility on both sides of the issue. Dreyfus thought the project faced fundamental limitations in its quest to emulate the human subject, as it rested on the wrong foundations. Indeed, at first glance a computer may seem like the perfect candidate for the human subject; it takes in discrete bits of information about the external world (perceives), feeds them forward to a CPU which performs logical operations on that data (thinks) with the help of an extra-reserve of further symbolic strings (memory), and then outputs the appropriate motor operations (acts). Dreyfus observed that once one undergoes minimal phenomenological development such an account of the human condition falls apart on a number of grounds. Intelligent understanding is highly context dependent, and thoroughly devoid of discrete and determinate units of meaning. My natural engagement with the world is not a matter of following explicit rules, but rather my ability to immediately perceive salient features and 'skillfully cope' in an unmediated way (Dreyfus 1979). Time proved the insight of Dreyfus' critique, and at this point GOFAI (Good Old-Fashioned AI) is effectively dead; the most natural and effortless of human practices proved remarkably difficult to emulate (Andler 2007).

To what extent can Dreyfus' critique and the broader phenomenological perspective be put to a constructive rather than destructive end, and serve to guide artificial intelligence and cognitive science in the right direction? It would seem in many respects, whether openly or tacitly, cognitive science has begun to embrace many of the central tenants of the phenomenological discipline (to the extent that it can, given it's broader naturalistic foundations), such as the necessity that consciousness be embodied and world engaging, the inadequacy of sense-data

theories, the contiguity of perception, cognition, and action, the indeterminacy of percepts and meaning, etc. Such grounds are the basis for a broad conglomerate of new approaches in cognitive science collectively referred to as *post-cognitivist* (Potter 2000). I will focus on one model suggested through these approaches, the Bayesian predictive processing framework, that seems particularly well equipped to address certain phenomenological concerns. This should not be taken as an outright endorsement of the framework and its capacity to capture the phenomena, but rather as an opportunity to engage in/demonstrate a truly interesting project open to cognitive science, some kind of neurophenomenology (Gallagher 2009), in which one tries to breath life into the machine and ascribe their phenomenology to it, or equivalently recognize traces of the machine in their own lived experience, to get a sense of *what it is like to be a machine*. Thus part two of this paper can be simply put as a *Phenomenology of the Bayesian predictive machine.*

In this way the explanatory gap will finally be addressed, if only indirectly, receding or looming large along with the relative success or failure of this and similar approaches. In this regard I must reemphasize the importance of Phenomenology (with a capitol P). Phenomenology gives us a truly first-person account of lived experience, but prompts us to regard ourselves in a novel way, pushing back against the natural impulse to surrender to experience's immediacy and opacity; to regard ourselves as a constellation of irreducible, ineffable, intrinsic and wholly inert qualia. As Merleau-Ponty states at the very beginning of *Phenomenology of Perception, "*In beginning the study of perception, we find in language the seemingly clear and straightforward notion of sensation: I sense red or blue, hot or cold. We will see,

however, that this is the most confused notion there is, and that, for having accepted it, classical analyses have missed the phenomenon of perception." Of course it is precisely this account of experience—hot, cold, red, blue--that suggests an explanatory gap in its most menacing, insurmountable form, as qualia are by their very definition impervious to any sort of explanation or naturalistic account. As such those who stand on the opposite side of the gap—those convinced of the full explanatory power of the natural sciences--have a vested interest in qualia's removal from the discourse on consciousness and present their own arguments against them as conventionally conceived (Dennett 1988). In any case, the *phenomenology of the machine* presented here will bear little resemblance to the problematic identity statements we started with (pain=the firing of c fibers).

I

Methodological bracketing such as heterophenomenology could be (and has been) regarded under a range of interpretations; from the strict adherence to a 3rd person, naturalist perspective that shuns introspection, to something so permissive it is unclear as to whether or not anything is being bracketed at all. Dennett notes "the difficulties that people have had trying to see whether heterophenomenology is a trivial redescription of familiar practices, or a restatement Husserl with nothing original in it, or a betrayal of Husserl, or a revolutionary proposal on how to study

consciousness, or a thinly disguised attempt to turn back the clock and make us all behaviorists, or an outrageous assault on common sense, or something else."

As such Dennett has returned to the subject again and again in an effort to clarify his position (Dennett 1991, 2003, 2007), each time placing greater emphasis on a permissive interpretation. I will reflect this movement towards the center in my assessment of methodological bracketing, starting somewhere far to the left of the mark, a thoroughly 3rd person account, from which the mind is regarded exclusively in the terms of reduced natural science. I will then allow for more and more direct reflection on subjective experience to account for what is left opaque, ultimately leading us into the domain of study I am interested in.

So where does pure, reductive, natural science as it is conventionally conceived leave us? What if we were to have the full scientific image of the mind? Lets suppose that the naturalistic reduction of the mind is hugely successful, and some 200 years from now we start building minds ourselves, and not in the old fashioned sexual way. The dream child of strong AI is popping off the assembly line, synthetic people, built from the bottom up, that function just like you and me. These humanoids are not the product of a sporadic evolution of computer and robotics technology, rather they are modeled after us as closely as possible, the final product of a massive *Blue Brain* project (Duncan 2005), which successfully reverse engineers and schematically maps of all the function components of our own brains and bodies. The scientific project is thus more or less over with regards to human beings, just as it is over with regards to cars and personal computers. Nothing mysterious is going on under the hood and there is nothing more to be said or

unpacked, we have unbridled access to every functional part of this machine's schematic.

Now imagine an extremely talented electrical/ computer engineer, John, who up until this point has no exposure to this advanced humanoid technology. He's always wanted to know how his brain works, so he goes down to the factory to find out. "You've come to right place'' a technician tells him, "We can tell you exactly how these machines work. Exactly how you work. We won't give you any hand waving, appeals to mental states, normativity, teleology, or raw feels, just a highly complex, technical schematic, all the relevant circuitry for this sophisticated robotic system presented in purely naturalistic terms." He hands over a massive tome. "Go home and read up. This is what we build here, this is man, no more, no less."

Night after night John pours over the text, which while highly technical and demanding, is in no way above his capabilities. All the same, as he gets closer and closer to the end, as all the dark corners of potential mystery are illuminated, he is filled with a growing confusion. He is working towards an impossible end, made all the worse by the fact that he understands each step along the way. It's reminiscent of a time he followed his friend's proof that the summation of all the natural numbers, $1+2+3+4+5+6 \ldots = -1/12$. But how? The proof in no way illuminated the result. His friend reminded him that it's only the analytic continuation over the complex numbers by the Riemann Zeta function that equals -1/12, but it only seems to push his confusion back. Why/how is the analytic continuation -1/12? Is there a

deeper mystery, or perhaps another mathematical avenue from which the result might be illuminated?[2]

John has a question he's not sure how to ask, a childish philosophical itch that has up until this point receded with time. He used to try and ask earnestly and coherently such questions, but they were hopeless attempts dismissed as a toddler's stupidity. Now he has his own children to dismiss, and nothing left of such questions but an occasional fleeting angst when he catches his own eyes in the bathroom mirror. Reading this tome has forced the philosopher in him up from the depths once more, his cry of confusion cannot be ignored. I see the machine, it is stripped open before me, every part. But I do not see myself in it. Where are the pains, where is red, the smell of rancid meat, the sound of a song? Where are the choices made and emotions felt? Where is my wandering train of thought? Where does it all come together as it is for me? What is it like to be this machine? It doesn't seem to be like

---

[2] Mathematician Terence Tao notes this result can be viewed from three mathematical levels of maturity, the pre-rigorous, rigorous, and post rigorous. When one reaches the post rigorous stage they have not only explicated and formalized their mathematical concepts, but they have rediscovered their intuitions, they rediscover the 'big picture', and it is from this perspective that they conduct their work. From the post-rigorous perspective Tao thinks the counterintuitive result gets some conceptual meat, that 1+2+3+4= -1/12 + . . . "where '. . .' on the right side denotes terms [. . .] 'orthogonal' to that application." We might characterize our project then as coming to the post-rigorous level of maturity in cognitive science, where one can think fluidly about their personal level of experience, all against on the backdrop of the machine (Tao 2014).

anything! Maybe it isn't like anything at all, maybe we've made a mistake and it's not conscious, maybe it's just a zombie . . .

At this point there are two primary ways one might dismiss the thought experiment. The first is to hold it to be incoherent from the start; there cannot be a schematic blue print such that if we 'build this' we will build a human being. The second is to hold that after reading the tome there is no room left for coherent confusion, that solely in virtue of understanding the technical description all mental content will be explained away[3]. Such responses are those of the 'phenomenologist who has lost naturalism' and the 'naturalist who has lost phenomenology' respectively, the very positions this essay is ultimately warning against, so if you find yourself so disposed bear with me, allow me to go on, and come back to this example's shaky origins after everything is on the table.

---

[3] I am inclined to dismiss this second objection outright on the grounds that it is simply empirical false. In my personal correspondence with friends who work directly in machine learning, one at Lawrence Livermore National Labs and the other at Wolfram Alpha, I observe clearly the disconnect that troubles our hero John. My friends have a far better grasp than I do of the pure technical content behind the systems I will investigate in part 2, the sort of content that John would presumably be exposed to in his readings, the purely naturalistic content. With no background in philosophy however, my friends have no real inkling as to how these systems could possibly pertain to consciousness. They have left their own consciousness at the door, and working entirely from the ground up it would seem there is no hope of them recovering it.

So, it would seem our hero is standing on the brink of the explanatory gap, and if we want to pull him back from the edge, where do we turn? We cannot appeal to the tome alone, the scientific image, we have already exhausted that, that's what got us here in the first place. Perhaps John should speak to a philosopher, since, to use Dennett's conception, "at least a large part of philosophy's task consists in negotiating the traffic back and forth between the scientific and manifest images" (Dennett 2004, Sellars 1962). Indeed perhaps he should turn to Dennett himself, who has taken up the role of being the demystifier of consciousness in these regards. Unfortunately in this case Dennett's usual diagnosis--that the gap stems from an ignorance of the neurobiology--simply will not do. The scientific image is completely laid out and presumably understood. All the 'easy problems' are accounted for, but is there some lingering 'hard problem' (Chalmers 1995)? Luckily, we *can* appeal to more than just the tome, we can look beyond the strict scientific image and appeal to the manifest image, we can appeal to ourselves. Perhaps there is still a great deal of learning to be done; perhaps John still needs to learn about what it is like to be John.

But such a proposition only confuses John. After all, what could this learning possibly consist in? Doesn't he have a full grip on his own experience by default? John has lived his whole life in *the natural attitude* (Beyer 2013) and thus does not see the possibility of phenomenological reflection. He wonders, "But how then could I have not reflected? How could the inspection of the mind, how could the operation of my own thought, have been hidden from me, given that my thought is by definition for-itself?" He has yet to undergo that curious existential shift, where his

experience remains entirely unchanged and yet he sees it as it truly is and has always been for the first time, where reflection does "not limit itself to replacing one view of the world by another . . . [but rather] . . . shows us how the naïve view of the world is included and transcended in the reflective view (Merleau-Ponty 2012, pg 221)." Returning to our musical metaphor might help him see what this reflection consists in, and how it pertains to him resolving his explanatory gap difficulty.

John's musical analog is a blank slate with no prior musical exposure of any kind in ideal experimental conditions. He is placed in an empty room with nothing but speakers that play a wide variety of musical scores. He is quickly entranced, and after a few hours he's taken out of the room and enters a 'cool down' period for a couple of days so the music isn't so fresh in his mind. Then he begins a rigorous musical training, but it's a noiseless training, the exact same sort of training the pianist's son under went at the beginning of the paper. It culminates ultimately in a series of scores, which the student understands from many theoretical perspectives; he understands the underlying music theoretic concepts that make up the scores, as well as the physics theory, that ultimately these scores are just describing complex waveforms. When he is then told that those waveforms 'just are' the sound waves he was listening to days before he is baffled at first. But those sounds . . . they were infused with this ineffable feeling and meaning, there were none of these awkward abstract concepts and rules, they are two different worlds, incommensurable. But perhaps . . . he listens to the music again, this time with an informed ear, this time he knows what to listen for. The pieces play and they all sound exactly the same in one respect, but different in another. The music carries something new with it, and

unexpected isomorphisms keep popping out at him. There's the major scale, the 2-5-1 progression, the salsa rhythm. He realizes he can follow along to the music with the score in his hands. The music comes to a section he recalled from his first listening, this uncomfortable wavering shriek, viscerally dissonant. When he was first told of the identity between the waveform and the music, it was this sound his skeptical doubt first brought to mind, as he had been jarred by its expressive, qualitative nature. Looking at his score, he sees the sound corresponds to two notes a half step apart being played simultaneously, which he knows from his lessons is a small enough interval to generate a beat frequency, a fluctuation in volume which overlays the wave at a much lower frequency than the pitch, kind of like . . . the wavering shriek he heard. hmmm, that's interesting. Where at first he felt himself faced with a stark explanatory gap, he is suddenly finding it very difficult to formulate his worry. After all, the identity between the wavering shriek he heard and this beat frequency seems anything but arbitrary. Can he really imagine one without the other, or is each just a description of the other? From there it's easy to generalize to the tone and timbre of the instruments; given their waveforms; could they really sound any different[4]?

---

[4] One could object that this isomorphism between qualitative sound and waveform is only possible in light of *some* primitive sound for which no isomorphism can be drawn, something to establish sound in general, the sound of a pure sine wave lets say (but really it could be any sound, we just have to *start* somewhere). I am very sympathetic to this view, it is a particular manifestation of our original difficulty, we need a background of sense, a sound axiom, something must be left out. I can only respond by saying that we can recognize interesting isomorphisms nonetheless, if

Obviously the above thought experiment is quite an unnatural construction, as I said before the most natural and fruitful music education involves a constant interplay such that the theory is never truly separate from the sound. Similarly, I positioned John as an outsider to the development of AI as a means of establishing a sharp contrast between the scientific and manifest images that otherwise might have been muddled. Those within the field present during the development of the technology would, like the typical musician, already be engaged in a natural interplay between both halves of the project simultaneously, the scientific and manifest images both becoming illuminated by the others light, the machine being modified towards capturing the phenomena, the phenomena towards capturing the machine, in a gradual process of convergence. Or at least, that is the hope, that is the approach we must be wary not to disallow ourselves.

In any case, it would seem John must 'listen again' to the content of his manifest image, to his phenomenal world. Just as with the music, in the immediate sense everything will remain the same; it will look, sound, feel, and emote as it always has. In another sense, it may drastically shift; he may recognize what was latent in the structure of his world from the start. Up until this point Dennett and myself are very much in agreement. He suggests that perhaps the "ultimate nature" of phenomena is drastically different than we thought, and that introspection is not simply "a matter of just 'looking and seeing'" (Dennett 1991). However, what we

---

we allow ourselves the notion of *the sound of the waveform*. What isomorphisms might we recognize if we allow ourselves the notion of *the phenomenology of the machine*?

each prescribe next, what will enable us to move past folk psychology and the natural attitude, is very different. Dennett of course suggests the heterophenomenological method, which I worry makes precisely the wrong emphasis, that we must *distance* ourselves from our experiences, rather than engage them ever deeper still.

When asked to clarify the boundaries of the heterophenomenological bracket, Dennett put it thusly, "Lone-wolf autophenomenology, in which the subject and experimenter are one and the same person, is a foul, not because you can't do it, but because it isn't science until you turn your self-administered pilot studies into heterophenomenological experiments." (Dennett 2003) In one sense I very much agree, indeed, *it isn't science*, but perhaps that's precisely the point. After all, in the case of helping John with his explanatory gap difficulty, we are not in the business of furthering science, the work of science is already well and done, the final model written up in that giant book John's been reading. Surely, then, if we are working towards curbing John's existential confusion we can recognize the value of phenomenology over heterophenomenology. Any number of experimental circumstances might help shift his intuitions in the right direction, and in each case him sitting in the subject's chair and living out the circumstance will have a much stronger and clearer effect than observing others talk about what sort of experiences they are having.

Of course even if our end is to further science, we must recognize the breadth of *this* project, that it involves a great deal more than collecting empirical data. Cognitive Science is still quite young, and has yet to produced anything close to a

grand unified theory of consciousness that stands out against the fray. It would seem a very large portion of the work in need of doing is on the side of wholly creative theorizing, analogous to the work of the theoretical physicist, as opposed to the experimental physicist. For the theorist the experiment *result* is the start as much as it is the end, it is meant to spark him in the right direction, to sculpt his perspective. Just as with John, we must not underestimate the value of intimacy, of engaging our perceptions directly and working from them. So it goes with all creative endeavors; we must foster the right intellectual climate such that something truly novel is most likely to bubble out of the imagination. A collection of interesting experiences themselves is a much more potent concoction for sparking innovative mental chemistry than the "scientific data" of people's belief statements about those experience.

Here it might be useful to proceed with another story, a story of strict heterophenomenologists studying the effects of LSD use. They proceed entirely from the 3rd person, and describe their project as such; 'We administer LSD to the machine, and take measurements on a vast array of physiological changes that then take place. We take very seriously observations regarding the machine's self reports; just like all the other empirical data presented to us, these reports require an explanation.' They conduct their research true to their word for many weeks on a test subject, John. During the course of a usual interview session, John is displaying one of his usual symptoms, self-reports of the ineffable nature of his experience. John starts to get frustrated with his inability to convey his condition, and proposes, "Look, there's no clear way for me to say this. Maybe there is, and I'm just not a speaker up to the task.

I can't tell you what its like, but I can show you. All you have to do is take the drugs, enter my world, and everything I mean will become abundantly clearer. Perhaps you'll be able to describe what it's like better than I can."

"John I think you misunderstand our project." The researchers respond. "We are researching the effects of LSD from the objective perspective; we are interested with its effects on a *machine.* It is very important that we maintain our neutrality; for us to take the drugs and proceed directly from the first person would be ungrounded, poor scientific practice, it would be to proceed on the basis of an *illusion."*

John's frustration only increases, "No, I understand your project and your ultimate aims, I want to help you! Don't you see the limits you are placing on yourself, that you are actively suppressing an abundant source of true, potential insights in regards to *your project.* Can't you at least appreciate the utility of taking the drug, let alone my further conviction of the *necessity* of taking the drug? Granted, I don't think you can truly appreciate this necessity until you are in my shoes . . . but still! Put yourself in my shoes! Tell me again why you are interested in my subjective reports at all? Hell, tell me why you are interested in the entirety of your research, in *LSD.* I just don't see why you insist on shooting yourself in the foot about this!"

The researchers wrote furiously in their notepads; the subject was in quite a state, displaying a lot of interesting behavior that would need to be explained.

Dennett would surely respond to this story by claiming it is a false portrayal of heterophenomenology, that the heterophenomenologist is free to take the drug. He has made such responses elsewhere; as I said before, he supports a very permissive interpretation of heterophenomenology. Would he agree with the

further claim that it is useful to the researcher's project to take the drug, or the further claim still, that it is in fact so useful it should be seen as methodologically indispensible, a necessary part of the researcher's project? Why should we content ourselves with mere linguistic descriptions of LSD, or of Crane and Piantanida's impossible color (1983), when they fall well short of capturing the phenomena and we can bypass a lot of opacity and confusing by simply sitting in the subjects chair ourselves? If Dennett does agree, then perhaps our positions are not so different after all, and I'm just another of the many critics who attacks the method, only to "then go on to describe what they take to be a defensible alternative methodology that turns out to be. . . .heterophenomenology!" (Dennett 2007) Of course in this case I must insist that 'heterophenomenology' is an empty title, that he's 'putting a paper crown on the king' (Wittgenstein 1958), as it is entirely unclear if anything is being bracketed at all, save some ludicrous 'Lone-wolf autophenomenology at the expense of all other modes of investigation'.

Afterall, heterophenomenology must not (cannot) mean 'heterophenomenology at the expensive of autophenomenology', it is autophenomenology, phenomenology of one's self, that we must inevitably return to if we are to have a meaningful language. Above, LSD is used as a particular mode of consciousness, a particular phenomenal world in which the research subject could legitimately exist while the researchers do not. But Dennett's project is concerned with *all* modes of consciousness, with all phenomenal worlds. As such it is impossible for the general heterophenomenologist to position himself in the way these LSD researchers are positioned; he cannot will himself into a zombie along with his

subject matter. If he could, then John's difficulty in conveying meaning, as presented above, would be infinitely compounded, as would his insistence that the researchers open their eyes. After all, in light of what do researchers *adopt the intentional stance* and *interpret speech acts*? The subject tells them, "I see a red dot in the middle of my visual field, slowly growing larger." The words carry an immediate significance; they prompt the researcher to see for himself, in his mind's eye, a red dot slowly growing larger. We've interpreted the speech act, which seems to amount to nothing more than following a cumbersome road right back to first person phenomenology.

Here I should note the parallels between heterophenomenology and Stroud's response to transcendental arguments. What I have essentially stated above--that our own phenomenology, our immediate personal world, is a necessary condition on the possibility of a meaningful language--can be quite naturally grounded in a transcendental argument. Indeed it is perhaps the most basic transcendental argument--that which underlies the entire Phenomenological discipline, phenomenology as precisely that which gives meaning (See White 2007 for a more analytic example of a transcendental argument for phenomenology). Stroud responds to the transcendental argument as follows, by observing that 'it is enough to make language possible if we *believe* that *S* is true, or that it looks for all the world as if it is, but that *S* needn't actually be true' (Stroud 1968). So it goes with heterophenomenology; the only 'real world' is the external natural world, the inner world is characterized as a theorist fiction or an illusion, an illusion that constitutes our very being and necessarily ensnares us. A response to Dennett then goes hand in hand with a response to Stroud; can we really have it both ways? Can we

simultaneously not *really* believe in the existence of the phenomena and necessarily believe in it to ground a meaningful language? How do we coherently doubt precisely that which we cannot doubt?

Let's now move on to the second form of bracketing I referenced at the beginning, a mirrored bracketing that launched the phenomenological movement, the bracketing of the 3rd person, objective world. There is a strong argument to be made against the bracketing in its strongest form, a true and thorough *epoche* where one suspends *all* their external world commitments simultaneously. Again, one might put forward a transcendental argument against the possibility of such a bracketing, that it disallows the possibility of a meaningful language (White 2007). Later phenomenologists such as Merleau-Ponty and Heidegger recognized this and distanced themselves from this aspect of Husserl's work. It would seem however that something of the bracketing carried through, that the classic phenomenologist is naturally dismissive of certain approaches concerning mechanical, natural systems. Learning about machines isn't going to get us anywhere in regards to learning about the phenomenology, in fact such considerations run the risk of leading us astray! When cognitive scientists complain that phenomenology itself doesn't get us anywhere in regards to how the machine works, the phenomenologist shrugs off the criticism as besides the point, and as presupposing such things are a matter of machines at all. Suppose John is a phenomenologist in this sense, and a phenomenologist interested in the effects of LSD. While marveling in the wonder of his obscure new experiences, John is approached by LSD researchers who want him to consider their work (work regarding a complex machine that undergoes changes

and begins to behave differently). He might respond as such, "I think you researchers are misunderstanding my project. I am concerned with the nature of my *experience,* of my immediately present 1st person world. To proceed on basis of the considerations you present, considerations about a mechanical world, would be to proceed on the basis of the illusion."

"No, we understand your project and your ultimate aims." Respond the researchers. "We want to help you! Don't you see the limits you are placing on yourself, that you are actively suppressing an abundant source of potential insights in regards to *your project?* Can you not see the potential utility in hearing us out? That it might serve as a catalyst, that it might stir the pot?"

We see John's concern reflected in this statement of Merleau-Ponty who notes that if we appeal to the machinery of the eye in our effort to understand vision, "We ought to thus perceive a sharply delimited segment of the world . . . But experience offers nothing of the sort, and we will never understand what a *visual field* is by beginning from the world. (Merleau-Ponty 2012, pg 6)" But perhaps the machinery of the eye is simply the wrong machinery to appeal to. Vision science has come along way, sparked precisely by this sort of insight, and has proposed machinery that is much more provocative as to the nature of *visual fields* and their phenomenal ilk. Such a machine is precisely the topic of part 2, so perhaps we should simply move on and allow the subsequent investigation to serve as evidence for the claim that machines can prompt phenomenal insight.

II

Here I will begin by briefly outlining the Bayesian predictive processing framework (for a more thorough treatment see Clark 2012). The model characterizes perception as the result of a predictive processing hierarchy of neural subgroups, in which top-down predictions are made by higher-level groups concerning the states of the groups below them, and bottom up information is sent back from lower-level groups in the form of error signals, or the discrepancy between predictions about those groups and their actual states. To get the predictive processing model off the ground, it helps to consider the utility of predictive processing in the problem of unsupervised learning (see Hilton et al. 1999). The problem of unsupervised learning arises when a machine must find patterns or structure in a set of unlabeled data. The difference between unsupervised and supervised learning can be roughly characterized as such; in supervised learning the machine is shown various pictures of animals along with the appropriate description, "dog. . . cat . . . horse", which help guide the machine as it tries to establish such categorizations on its own. In unsupervised learning, the machine is simply given a series of images and must generate its own categorizations. The brain is then in the business of unsupervised learning. It is in the dark, it only has immediate access to information about itself, to the unfolding firings of its own neural populations. Predictive processing offers up an elegant solution; one neural population, via dumb mechanical processes, is tasked with predicting the next state of another neural population (a population lower in the hierarchy). The population being predicted sends information concerning the accuracy, or rather the *inaccuracy* (i.e. prediction error), back to the predicting

population, which adjusts its next prediction accordingly and gradually reduces error.

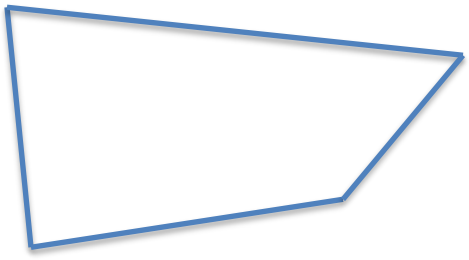
Over time the predicting population gets better and better at generating accurate predictions, and can be thought of as having a generative model for the unfolding firings of the lower neural population. For having a model or theory as to how a system works is closely linked to making predictions about how that system will behave. Consider the scientific method; we make hypotheses about how a system with unknown causal variables will behave and compare them to what we empirically observe. While its conventional to view good scientific models as the generators of good hypotheses, we must appreciate this relation as bidirectional. Generating a model of a system can be understood as establishing a methodology for making good predictions, and thus if we concern ourselves solely with making good predictions and succeed we will likely have developed an accurate model in the process. Empirical Bayes is a purely mechanical statistical procedure that does exactly this; it develops the priors that will drive its future predictions (its generative model) concurrently with making predictions and comparing them with the empirical data. Using hierarchical Bayesian predictive coding as a means of developing visual models is not just a theoretical possibility, but has actually been put into practice, where an unsupervised learning machine generates from the top down, with increased efficiency and accuracy, natural scenes it is presented with (Rao and Ballard 1999). Thus empirical Bayes gives us a tractable computational grounding to work with.

The model here presented has a number of happy side effects. First, predictive coding is a great way to compress data, as only prediction error needs to be transmitted and all predicted information can be left out. In fact it was through an effort to compress data that predictive coding was primarily developed, it's the tool behind JPEG and lossless audio compression (Shi and Sun 1999). Second, the model seems like a good way to deal with the multiple-realizability of sensory signals. Bare sensory signals could represent a vast multitude of real world states that would bring them about (Helmholtz 1860/1933), and a brain that uses probabilistic inference to generate models is well equipped to differentiate these states. Lastly, the single mechanism of predictive processing can account for a wide spread of mental faculties. It explains perception as the top down cascade of predictions tuned to explain away upward flowing error signals driven by sensory input. Such a system working offline, detached from sensory input, gives us an explanation of our imaginative faculties and dreaming. As shown above, the 'bootstrapping' effect of predictive processing is inherently linked to learning and cognition. Predictive processing even gives us an account of action as yet another product of the elimination of error signals (Friston et al 2010). Where perception arises as our predictions conform to error signals sourced in the world, action arises when higher levels in the hierarchy model some course of action and generate predictions about the unfolding states of lower levels in the hierarchy. Error signals thus emerge, as we are not performing the desired action, but this time the body itself adjusts rather than the predictions, thus generating the sensory inputs that will fit the predictions and eliminate the error signals.

Now suppose our hero John, 200 years in the future, is reading the machine schematic and comes across this sort of architecture. Its one of the things he finds deeply puzzling, it's just not how perception seems for him; why does it only feed forward error signals, why doesn't it just feed forward the sensory input, that's what he sees after all isn't it? He is discomforted by the distance predictive processing puts between perception and raw, visual, sensory data. According to the model, sensory input only gets into the brain indirectly, by driving the prediction error. These error signals inform the development of probabilistic, internal models of the world, a sort of 'guided hallucination' in where perception must lie, (if anywhere). Yet intuition tells him that the bulk of visual phenomenology is captured neatly and succinctly by visual sensory data. He conceptualizes that data as an image, a projection of the external world onto the two-dimensional plane of the retina. He conceptualize his own visual phenomenology in much the same way, as it is difficult for him to distinguish the pure content of a picture and what the picture looks like, what it is to *see* the picture. J. J. Gibson notes, 'Painting can reach a degree of perfection, we are told, such that a viewer cannot tell whether what he sees is a canvas treated with pigments or the real surfaces that the painter saw, viewed as if through a window (Gibson 1978 pg. 231).' Thus we open ourselves to error, that the content of visual phenomenology is the same as the content of the painting, the paint. We are the rational agent standing in front of our perceptions, looking at the canvas of our sense data. We do not feel as if we are making up a world and consulting our eyes only to ensure we are getting the picture right.

John's difficulty stems, in part, from the fact that he's getting his own phenomenology wrong, a sophisticated phenomenological account is much more amenable to the Bayesian conception of vision. Intuition Pump One, '*Drawing a Windshield*' (White 2007), will hopefully shift the intuition that visual sensory data and visual phenomenology have highly similar content. Get out a piece of paper and draw with four straight lines the apparent outline of a windshield as viewed from the driver's seat of a car. Most draw a very simple shape, minor variations on a standard isosceles trapezoid;

Rarely does one draw an image that captures all the important features of the correct shape, this shape;



The top is longer than the base and both lines converge to the right because the bottom and right edges are farthest away from the viewer. The angle of the line

on the left side is different than that on the right because the left edge of the windshield is mostly inline with our gaze, and the right edge is not.

In drawing normal isosceles trapezoids, we are not drawing the apparent image produced by a windshield from a certain angle, rather we are drawing what we think we see (what we do see?), namely, a windshield. Generalizing, I can safely say that most of us have a very poor grasp of drawing anything we see in our visual world, even when we are free to stare directly at the objects we are attempting to capture; I often stare back and forth between the world and the page trying to see the appropriate direction or length of the line I'm drawing. To learn to do so requires a great deal of attention, effort, and training. Such a difficulty seems quite strange in the light of our original intuition, that visual phenomenology bears the information of sensory data in an immediate way. For the visual sensory data *is only* the 'apparent image' of things, which now seems like something that must be extracted from our visual phenomenology, and not without difficulty.

In the context of this example, I should return to heterophenomenology for a moment. It is indeed good form that such an experiment be conducted in accordance with scientific methodology, that we do not recklessly assert we all lack direct perceptual access to the apparent shape of the windshield, we should collect empirical data from the populous. Nonetheless, the power of the example, what will enable us to shift our intuitions and *see things differently,* comes from living out the difficultly oneself.

Such is an example of *deflationary* phenomenology relative to sense data; that we find less in our visual phenomenology than would be supposed under a

sense-data theory. There are also countless examples of our phenomenology being *inflationary,* or much richer than what we are given by sense data. Imagine you are the person (crudely drawn) standing in Figures 1 and 2 below.

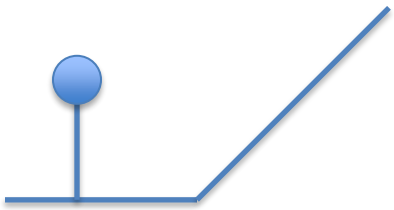


Figure 1

Figure 2

In both instances you are standing on a road in a minimalist setting, 10 feet from the line at which the plane of the landscape shifts by 45 degrees. However in one instance you are standing on flat ground and staring at an upward incline, and in the other you are standing on a downward incline and staring at the ground leveling out (You are attached to a rope and thus maintain the same angle with the ground). The data entering your eyes, the perceived image, is the same in both Figure 1 and 2 (a particular instance of Hemholtz's insight described earlier). Yet we can still ask the question as to whether or not these two circumstances would *look*

the same. Further still, it seems quite reasonable that the answer to this question is *no.* Figure one *looks* like an incline rising up, and figure two *looks* like a downward slope leveling out. Information from other sensory modalities seems to penetrate visual phenomenology in a way that is ineffable, but only if we limit ourselves to the 'feed-forward' model of visual perception. If we liken perception to a model of the world generated from the top down, one that places us in the spatial context that best predicts the unfolding of sensory states across all modalities, it seems quite reasonable that a distinction between 'leveling out' and 'inclining' would work its way into visual perception.

Rich phenomenology might get us away from the idea that perceiving is simply a matter of perceiving the sense data, but we will have to say more about the phenomenology of perception if it is to call for anything particularly Bayesian in nature. The phenomenology of Merleau-Ponty centers on a key insight similar to Helmholtz's. Perception is necessarily between intellectualism and empiricism in the sense that it involves a passive power of organization. Perception becomes an act of sorts; it necessitates getting a grip on the world. This is best demonstrated by his example *coming upon the boat.* Merleau Ponty describes walking down the beach towards a boat sitting in front of a forest. As he gets closer he "felt that the appearance of the object was about to change, that something was imminent in this tension" (Merleau Ponty pg 18). Where at first the mast merges with the trees, "these details suddenly reunite with the boat and become welded to it." (pg 17) The question is this; in virtue of what does this shift in perception occur? Certainly it is not purely in virtue of the sense data, as those are more or less the same in the

moments right before and after the transformation. At the same time, it isn't in virtue of any higher order mental faculties; it isn't in virtue of an intellectual act of *our* doing. We do not make an active judgment, we do not call upon our memories or try to sort out the image in different ways, rather we are helpless to the transformation as it is inherent in the act of perception itself. All such higher order analysis is only possible in light of the fact that the perception came to us in the way that it did.

Like Helmholtz, Merleau Ponty observed that this circumstance is the norm rather than the exception; the primacy of the perceived unity of the boat extends to the simplest of perceptual objects such as the contour of a line. However for Merleau-Ponty the task faced by perception is much deeper than that faced by a Helmholtz statistical machine that must decide between the infinitude of 'real worlds' that would result in those particular sense data, for this act of deciding, of getting a grip and constituting the world, is itself constitutive of all these possible 'real worlds'. As he puts it, "'Good form' is not achieved because it would be good in itself in some metaphysical heaven; rather, it is good because it is realized in our experience."(pg 17). Of course, this account does not refute Bayesianism; if anything it is simply a more radical argument for it's necessity, or the necessity of something like it, something above and beyond a 'feed-forward sense data' model. After all, 'good form' for the *unsupervised* learning machine is not given in the data itself, but is rather a construction realized through the Bayesian prediction process.

In his discussion of color sensation, an act that we intuitively characterize as wholly inert, the direct perception of an indescribable *quale,* Merleua Ponty puts

forth a description that bears uncanny resemblance to the Bayesian account, despite being completely unfamiliar with the framework.

> *Thus, a sensible that is about to be sensed poses to my body a sort of confused problem. I must find the attitude that will provide it with the means to become determinate and to become blue; I must find the response to a poorly formulated question. And yet, I only do this in response to its solicitation. My attitude is never sufficient to make me truly see blue or truly touch a hard surface. The sensible gives back to me what I had lent to it, but I received it from the sensible in the first place.* (pg 222)

Of course the context and intent of his words is very different than I am implying, but nonetheless one can't help wonder if he is catching the glimpse of some deeper mechanism, coming upon and living out it's existential manifestation.

Something very important to note in this regard—the underlying mechanism manifesting at the personal level—is that as long as the Bayesian machine (and a lots of similar systems) is working properly, it necessarily becomes hidden; the process of error correction falls away precisely to the extent that the machine is getting a grip on the world. Thus the machine and the phenomenology reveal themselves most clearly when something is amiss (an insight Merleau-Ponty was keen on. We will draw upon various cases of abnormality frequently). The neurological disorder agnosia demonstrates the drastic perceptual deficiencies that come when this passive power of organization/construction is impaired. Lets start

with the case of visual agnosia and the remarkable patient of Dr. Oliver Sacks, *The Man Who Mistook His Wife for a Hat.* This patient, Dr. P, began to visually deteriorate in his old age due to a brain tumor, first in the case of faces, which became increasingly difficult for him to recognize. What's more, he began seeing faces that weren't there at all, genuinely mistaking parking meters and doorknobs for real people. As the years went by his conditioned worsened, and generalized to other objects of perception. After consulting an ophthalmologist, who concluded there was nothing wrong with his eyes, he was referred to neurologist Dr. Sacks. Sacks characterized Dr. P's impairment most generally as an inability to perceive *the whole,* noting, "his eyes would dart from one thing to another, picking up tiny features, individual features [. . .] a striking brightness, a color would arrest his attention and elicit comment—but in no case did he get the scene-as-a-whole." It was clear that in one sense Dr. P was 'all there'; he was not suffering from dementia, he was charming, articulate, he could make sophisticated, intelligent judgments. In fact in some sense this was precisely his problem, as it was these higher order intellectual powers that Dr. P was relying on to try and *solve* what he was given in perception. He had to make the judgments his perception should have been making for him. When asked to identify a glove, he responded, "A continuous surface, infolded on itself [. . .] it appears to have five outpouchings." Yes, and what might it be? "A container of some sort? [. . .] It could be a change purse, for example, for coins of five sizes." When shown a minimalist desert scene with no details to pick out, Dr. P's imagination seemed to take over; "I see a river [. . .] and a little guest house with its terrace on the water. I see coloured parasols here and there."

Under a phenomenal analysis it is clear that Dr. P has lost the ability to *perceive* as Merleau-Ponty presents it, that he functions normally as far as empiricist and rationalist are concerned, but a visual world nonetheless never manifests concretely. Under a Bayesian analysis, while it would be too ambitious to offer an explicit diagnosis, we see the general markers one would expect from a degenerative predictive hierarchy; Dr. P has the perceptual capacities of a much weaker system, such as those (and in some respects much worse than those) currently being developed by computer scientists. As we would expect with a Bayesian system, the degenerative process is marked not only by an increasing inability to *get a grip,* for the generative model to get ahead of the sense-data, eliminate error signals and generate the whole, but also the tendency to go forward on the basis of the wrong model (seeing faces where there weren't any). Dr. P confabulates a river and a guesthouse in place of a desert, which is another symptom one might expect from a defunct Bayesian system. Just as the functional system imagines and dreams when working offline, unguided by bottom up error signals, Dr. P creates false imagery when presented with a more or less blank slate, and his brain is not muddled by bottom up error signals.

How might we characterize the Bayesian visual field? Of course I can only understand what a visual field is in relation to my own, so the extent to which I can characterize the Bayesian visual field *at all* rest on the extent to which I can ascribe features of my own visual field to it. Here I will present my own vision of the visual field influenced by Phenomenology and Bayesian perception alike. The visual field is characterized by the ever-increasing indeterminacy emanating from its center, some

object of focus. My imaginative powers play a larger and larger role as we work towards the horizon of the visual field, not in the sense that they are being used up and occupied by this space—quite the opposite—but in the sense that this area has the vague presence of something imagined. It is not wholly ground in the world. These imaginative percepts extend into the world beyond the horizon of my visual field, operating on there own, separate from any 'inbound' sensation.

The opportunity for exchange between phenomenology and the Bayesian predictive framework is all the more apparent in regards to the interplay and intimacy between perception and action. In the Bayesian framework perception and action are realized by the same mechanism working in the opposite direction of fit. Thus action becomes the means of realizing an expected perception. This aligns nicely with a variety of phenomenal characterizations, mainly that well habituated actions are directed all the way out to the desired perception. When we speak or write, we do not manipulate our tongue and lips or our fingers and wrist, rather we simply produce the words (perhaps we could go even further and say we simply produce the meaning of the words!). The experienced driver changes the speed and direction of his car directly, he does not tell his arms to twist and his ankle to flex. The experienced pianist is in the business of producing music, he is not in the business of moving his fingers. A baseball play chasing down a fly ball does nothing more than keep the ball in the center of his visual field. The list goes on. In this sense action is an act of perception.

The sophisticated phenomenologist tightens these bonds further; not only is action perceptive, but ordinary perceptions themselves have an agential character.

To use J.J. Gibson's term we perceive *affordances* for action laden in the perceived world. The soccer ball is seen as *to be kicked,* the door as *to be opened*, the trail as *to be followed.* Strip such perceptions of the affordances they offer and something is lost; I lose the trail in the woods when a clear line ceases to jump out of the bramble and beckon me through. Merleau Ponty observes that perception of ones own body often takes the form of a potentiality for action. The phantom limb patient persists in perceiving the same world after his accident; the piano keys are still *to be played,* the pen is still *to be grasped,* and thus his hand continues to be sensed despite its absence, as the perception of his hand is realized in the perception of these affordances. Such a perception of affordances should be expected in a Bayesian system generating imaginative projections of action with its perceptive machinery. The machine has masterfully tuned itself through a constant interplay of perception and action to engage the external world, and it is precisely through this engagement that the world reveals itself. It is a machine designed precisely to 'skillfully cope' with the world.

Again it would seem the machine is so good at what it does, we necessarily fail to notice it. That is, unless something goes wrong. Here is another case of Oliver Sacks[5] that hints towards a curious convergence of Merleau Ponty's phenomenology

[5] In regards to my *meta* point, which concerns the broader issue of appropriate method and style for phenomenology in cognitive science, I should note that my inclusion of ample Oliver Sacks is no accident. I am drawn to his work because it is an exemplar of the style I am advocating for; the work of a neurologist who's thought is steeped in lived experience, in the phenomenal world. The story I present here is a prime example of the scientist unafraid to work from the subject's chair.

and the Bayesian framework. It concerns the phenomenal status of spatiality, which Merleau-Ponty grounds in the habituated phenomenal body; space is dependent on my having a grip on my body schema and the ways in which it can engage the world. This time Sacks' patient is himself, as he describes his recovery from a fracture that stripped him of the use of his leg, such that it became wholly alien to him. Upon trying to walk again Sack's encounters great difficulty;

> *The floor seemed miles away, and then a few inches; the room suddenly tilted and turned on its axis. [. . .] Then I perceived the source of the commotion. The source was my leg—or rather, that thing, that featureless cylinder of chalk [. . .] It was constantly changing in size and shape, in position and angle, the changes occurring four or five times a second. The extent of the transformation and change was immense [. . .] Within a minute or two the changes became less wild and erratic [. . .] the conformations and transformations were being modulated and damped [. . .] But what could produce such an explosion in my mind? The perceptions had the quality of constructs, and not of raw sensations or sense-data. They had the quality of hypotheses, of space itself [. . .] I felt I was bearing witness, even as I was undergoing it, to the very foundations of measure, of mensuration, of a world.* (Sacks 1998)

Reading such passages I'm often filled with a strange envy, an envy grounded precisely in the spirit of this essay; if only I could break my leg and bear witness to that process of world creation that, in my existential mastery, I have all but lost access. I want to see for myself! But there are small ways of disrupting the usual feedback between perception and action, thereby revealing the difficult task at hand, our typical, unnoticed mastery of that task, and the bootstrapping process by which we establish that mastery. Try wearing inverted glasses, at first it is deeply

disorienting, the usual actions produce inverted perceptions, and something inevitably breaks down such that I can neither act nor perceive with confidence. Bayesian predictive processing gives us an account of how, after several days, we learn to adjust and reveal once again a cohesive world, how perception and action find each other once again. A similar effect can be found in wearing a 'speech-jammer', a recording device that plays back your own voice into your ears with a slight delay, rendering it very difficult to speak. I don't know if anyone has worn the device for a prolonged period and learned to cope, rediscovering normal speech. It would certainly be an interesting experiment.

If we allow ourselves to be a bit more speculative and apply the predictive coding model to the case of language, again a picture emerges amenable to the existential phenomenologist's view, who tends towards supporting a conception of language in line with the ordinary language philosopher (Weinzweig 1977). He claims language to be a loose normative process, of which the meaningful content is holistic and highly context dependent, much like perception. As such he is resistant to the notion that our use of language reduces to a program carrying out the formal rules of some complex predicate logic. But the Bayesian account of language suggests nothing of the sort. A very simple Bayesian account would characterize our grasping of language as the process of a Bayesian machine implementing strategies to best predict the next word in a sentence, or predict the context in which a certain word will be used. It is clearly reminiscent of the Wittgensteinian tenant 'meaning as use'. Bayesianism seems to suggest a new phenomenological avenue in general, that we might understand 'getting a grip' on the world better as 'getting one step ahead

of the word'. Indeed such a description seems to capture getting a grip on language quite nicely, as with language we often quite naturally get ahead of ourselves. Take these amusing examples;

> *What I if told you,*
> *You the read first line wrong,*
> *Same with the second.*

I cnduo't bvleiee taht I culod aulaclty uesdtannrd waht I was rdnaieg. Unisg the icndeblire pweor of the hmuan mnid, aocdcrnig to rseecrah at Cmabrigde Uinervtisy, it dseno't mttaer in waht oderr the lterets in a wrod are, the olny irpoamtnt tihng is taht the frsit and lsat ltteer be in the rhgit pclae. The rset can be a taotl mses and you can sitll raed it whoutit a pboerlm. Tihs is bucseae the huamn mnid deos not raed ervey ltteer by istlef, but the wrod as a wlohe. Aaznmig, huh? Yaeh and I awlyas tghhuot slelinpg was ipmorantt! See if yuor fdreins can raed tihs too.

It would seem there is a Phenomenological motivation for the Bayesian framework in many respects, but the subjective perspective also suggests clear limitations. The Bayesian framework has been criticized on the following grounds; if all the brain is trying to do is minimize prediction error, shouldn't we all regress into dark corners and sit motionless so as to have nothing to predict? Moreover, attempts have been made to defend the Bayesian framework against such criticism (Friston 2010). Minimal phenomenological reflection shows that such a worry confabulates what the brain is *trying to do* at the level of prediction error minimization, and what *trying* consists of at the personal level. While Bayesianism

gives an account of how a course of action might be carried out in our direct unmediated engagement with the world, of 'zombie action' (Clark 2001) or to use Merleau-Ponty's terms, 'motor intentionality', specifically as it pertains to 'concrete movement', it does not give an account of agency in the sense of deciding upon that course and exercising ones freedom. This example should highlight the distinction. When I was a little boy I was quite frightened of standing on subway platforms. I would stare down into the dark subway tunnel as it began to rumble with the approach of a train. It would get louder and louder, the light of the train's high beams getting brighter and brighter, closer and closer, bigger and bigger. I was entranced like a bug by a night-light; the approaching train filled my world fully, it absorbed all my imaginative faculties, demanding a certain continuation. All I could think about was walking forward and jumping into the gap, onto the tracks in front of the oncoming train. The thought absorbed me so fully that I would become frightened, because the imaginative projection seemed to pull me forward in a real way, it came with the immanent potential for real action. I felt helpless to my dramatic perceptions. So what did I do? I laid down on the platform, and waited for the train to arrive. Luckily, when faced with what might be interpreted as my Bayesian mechanisms trying to reduce prediction error, I still had my agency to fall back on.

There are of course countless other ways in which the Bayesianism predictive machine seems to fall short of capturing the phenomena, but these largely fall under the more general difficulty lurking down at the bottom of the abstraction ladder that we have carried with us from the start; namely, how does

the Bayesian predictive machine lend itself in principle to any phenomenology *at all*? Indeed when I look at the system as plainly as possible, as a system of objects in causal relations, when I make no effort to attribute anything extra to it, when I don't go actively searching for myself in it, I am lost. When I look at the system as naturalism demands I must, as dumb, lifeless, and without phenomena, it would seem I can't help but see precisely that; something dumb, lifeless, without phenomena. At such moments it's truly tempting to renounce the phenomenal in some way, to bracket it away and thereby rectify science. But upon a moments reflection I recognize something misguided in my impulse. The aim of science is, after all, to uncover and explicate the workings of the natural world. It is not the aim of science to vindicate the explanatory power of science. The fact that at some level I am deeply confused does not disable me in regards to furthering a science of consciousness. In fact in many respects it is precisely what enables me, in the sense that it comes with the territory of being an intelligent, thinking, creative mind. So I will continue pursuing the explanatory gap in the fashion outlined above, making progress in spite of myself, in spite of the fact that at some level I am entirely unsure as to how that progress is possible.

Works Cited

Andler, Daniel (2007) “Phenomenology in Artificial Intelligence and Cognitive Science” *The Blackwell Companion to Phenomenology and Existentialism*: 377-393

Beyer, Christian "Edmund Husserl", *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2013/entries/husserl/>.

Berndt, Bruce C. (1985), “Ramanujan’s Notebooks: Part 1” Springer-Verlag: 135–136

Chalmers, David (1995) “Facing Up to the Problem of Consciousness” *JCS*, 2 (3): 200-19.

Clark, Andy (2013) “Whatever Next: Predictive Brains, Situated Agents, and the Future of Cognitive Science” *Behavioral and Brain Sciences*

Clark, Andy (2001) “Visual experience and motor action: Are the bonds too tight?” *Philosophical Review* 110: 495-519

Clark, Andy (1997) “Being There: Putting Brain, Body, and World Together Again.” Cambridge, MA: MIT Press

Dennett, Daniel (1988) “Quining Qualia”, *Consciousness in Modern Science.* Oxford University Press

Dennett, Daniel (1991) “Consciousness Explained”, The Penguin Press: 281

Dennett, Daniel (1996) "Facing backwards on the problem of consciousness". *Journal of Consciousness Studies* 3 (1): 4–6.

Dennett, Daniel (2003) "Who’s On First? Heterophenomenology Explained" *Journal of Consciousness Studies, Special Issue: Trusting the Subject?* (Part 1); 19-30

Dennett, Daniel (2007) “Heterophenomenology Reconsidered” *Phenomenal Cognitive Science*: 1-20

Dennett, Daniel (2013) ”Kinds of Things—Towards a Bestiary of the Manifest Image,” *Scientific Metaphysics*, D.Ross, J. Ladyman and H. Kincaid, eds., Oxford University Press: 96-107.

Graham-Rowe, Duncan (2005) "Mission to build a simulated brain begins" *NewScientist*

Dreyfus, Hubert (1979) “What Computers Can't Do”, New York: MIT Press

Friston, K. J., Daunizeau, J., Kilner, J. & Kiebel, S. J. (2010) Action and behavior: A free-energy formulation. *Biological Cybernetics* 102(3):227–60

Friston K. (2010) The free-energy principle: a unified brain theory? Nature Reviews Neurosciences 11(2):127-38

Gallagher, Shaun (2009) "Neurophenomenology". *Oxford Companion to Consciousness:* 470-472

Geoffrey Hinton, Terrence J. Sejnowski (editors) (1999) "Unsupervised Learning: Foundations of Neural Computation", *MIT Press*

Gibson, James (1977) "The Theory of Affordances. *Perceiving, Acting, and Knowing"*, ed. by Robert Shaw and John Bransford, ISBN 0-470-99014-7.

Gibson, J. J. (1978) "The Ecological Approach to the Visual Perception of Pictures." *Leonardo*, Vol. 11, pp. 227-235

Helmholtz, H. (1860/1962) *Handbuch der physiologischen optik*, ed. J. P. C. Southall, English trans., Vol. 3. Dover.

Jackson, Frank (1982). "Epiphenomenal Qualia". *Philosophical Quarterly* (32): 127–136

Kripke, Saul (1982). *Wittgenstein on Rules and Private Language*. Harvard University Press: 7-54

Levine, Joseph (1983). "Materialism and qualia: the explanatory gap". *Pacific Philosophical Quarterly* (64): 354-361.

Merleau-Ponty, Maurice, trans. Donald Landes (2012) "Phenomenology of Perception" Routledge

Moravec, Hans (1988), *Mind Children*, Harvard University Press

O'Regan, J.K., & Noë, A. (2001). "What it is like to see: A sensorimotor theory of visual experience." *Synthèse*, 129(1), 79-103

Potter, J. (2000). "Post cognitivist psychology", *Theory and Psychology, 10*, 31-37.

Rao, R. P. N. & Ballard, D. H. (1999) "Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. " *Nature Neuroscience* 2(1):79–87.

Sacks, Oliver (1998). “A Leg to Stand On” New York: Simon and Schuster

Sellars, Wilfrid (1962) “Philosophy and the Scientific Image of Man” *Frontiers of Science and Philosophy,* University of Pittsburg Press: 1-40

Stroud, Barry (1968) “Transcendental arguments,” *Journal of Philosophy*, 65: 241–56

Tao, Terence (2014) “There’s more to Mathematics than Rigor and Proofs”. URL= https://terrytao.wordpress.com/career-advice/there%E2%80%99s-more-to-mathematics-than-rigour-and-proofs/

Weinzweig, M. (1977), “Phenomenology and Ordinary Language Philosophy.” *Metaphilosophy*, 8: 116–146. doi: 10.1111/j.1467-9973.1977.tb00267.x

Wittgenstein, Ludwig (1958) “Philosophical Investigations” Basil Blackwell Ltd

White, Stephen (2007) “Psychology, Transcendental Phenomenology, and The Self. “ *Cartographies of the Mind: Philosophy and Psychology in Intersection*: 143-153

White, Stephen (2007) “The Transcendental Significance of Phenomenology” *Psyche* 13(1)

Yun Q. Shi and Huifang Sun (1999) “Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards.” *CRC Press*

Zalta, Edward "Gottlob Frege", *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward Zalta (ed.) URL=<http://plato.stanford.edu/archives/fall2014/entries/frege/>.