Phenom Cogn Sci (2007) 6:247–270
DOI 10.1007/s11097-006-9044-9

Heterophenomenology reconsidered

Daniel C. Dennett

Published online: 15 February 2007

Key words heterophenomenology · autophenomenology · cognitive science

> Descartes' Method of Radical Doubt was not radical enough. –A. Marcel (2003, 181)
>
> In short, heterophenomenology is nothing new; it is nothing other than the method that has been used by psychophysicists, cognitive psychologists, clinical neuropsychologists, and just about everybody who has ever purported to study human consciousness in a serious, scientific way. –D. Dennett (2003, 22)

I am grateful to Alva Noë for organizing this most stimulating and informative congregation of essays. They have opened my eyes to aspects of my own work, and the different contexts into which it must be shoehorned, and forced me to articulate, and revise, points about which I have been less than clear. Instead of providing seriatim answers to each essay, I am running my reactions together, taking advantage of the contexts they provide for each other, and concentrating on a few themes that emerged again and again. I apologize to those whose essays are given at most a glancing response; typically I found much to agree with in them, and nothing that *needed* discussing here. This essay threatened to grow much too large in the making, and I felt it was better to try to do justice to the most perplexing points raised at whatever length was required, at the cost of postponing other responses to some other occasion.

A bridge too far?

My epigraphs allude to the difficulties that people have had trying to see whether heterophenomenology is a trivial redescription of familiar practices, or a restatement

D. C. Dennett (✉)
Department of Philosophy, Tufts University, Medford, MA 02155-7059, USA
e-mail: daniel.dennett@tufts.edu

of 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. Although I had anticipated the sources of resistance to my proposal, it is clear that I underestimated them in several regards. And the level of disagreement among the commentators persuades me that my *buffet approach* to Husserl (take what I like and leave the rest untouched) was tactically the right way to go – even if some of my choices might be properly criticized on their own terms. Charles Siewert, for one, joins me in this tactic, seeing the need to detach his "plain" phenomenology from "dubious methodological claims associated with certain brand-name phenomenologists" (Siewert, 2007) and making choices not very different from mine, and Jean-Michel Roy usefully surveys my "dual attitude of dismissal and acceptance," and judges that my notion of Husserl's autophenomenology "does indeed capture the essential idea.. . and there is no reason to dismiss his criticisms as directed at a strawman. . ." (Roy, 2007). Others, such as Hubert Dreyfus and Sean Kelly, would disagree with this gratifying verdict, finding my attempt to salvage the best from the brand names "heavy-handed sleight-of-hand." Dan Zahavi is puzzled by my claim that if my reading of Husserl is wrong, so much the worse for Husserl. I was not defending the *accuracy* of my interpretation, as he supposes, but just noting that for me the question is always: *is my Husserl-inspired proposal right?* not *is this what Husserl meant*? I am happy to give credit where credit is due, so if my attempts at *improvements* on Husserl actually reinvent some of his wheels, I'll gladly concede this, but Husserl scholarship is not my primary aim. (Drummond's analyses of problems and controversies in Brentano and Husserl scholarship – see also Siewert – go some way to showing, to the uninitiated, why I prefer to disengage from that project.)

You can't please everybody, but I should have done better. To me the most persuasive evidence that I have failed heretofore to give a sufficiently clear explication and defense of heterophenomenology is the frequency with which commentators criticize it – and then go on to describe what they take to be a defensible *alternative* methodology that turns out to be. . . . heterophenomenology![1]

[1] I discuss this pattern of misconstrual in my essay, "Who's on First? Heterophenomenology Explained," in the special issue of JCS (2003) on "Trusting the Subject," (subsequently published as a book of that title), responding to Anthony Jack's challenge to say what *isn't* heterophenomenology and there I cited Chalmers, Goldman, and Velmans as instances of the misconstrual. In fact two other essays in that issue – essays I had not seen when I wrote mine – are even more vivid examples of the unwitting re-invention of heterophenomenology. Piccini (2003), after labeling me an "introspection agnostic" who is seen as "rejecting introspective reports as sources of evidence" (142), insists that "we do have means to evaluate the accuracy of introspective reports" (147) and then goes on, in a section entitled "The Epistemic Role of Introspective Reports in Science" to describe, quite accurately, the assumptions that heterophenomenologists use to turn the raw data of verbal reports into data about what it is like for the subject. He quotes Jack and Roepstorff (2002) as saying we should adopt a "second-person" perspective, in which subjects are treated "as responsible conscious agents capable of understanding and acting out the role intended" (149) as if this contrasted with heterophenomenology. Gallagher (2003) surmises that my motivation for heterophenomenology is 'longstanding suspicions about introspection as a psychological method" while his motivation is to address the problem of "how the use of introspection might be made methodologically secure" (91) but here we are in agreement. I also discuss this gravitation back to heterophenomenology in chapter 2 of *Sweet Dreams*, 2005), and the pattern continues, with variations, in the papers in this volume by Thompson, Roy, Zahavi and Velmans once more.

Let me begin, then, with something of a bird's-eye view of what I take heterophenomenology to be: a bridge – *the* bridge – between the subjectivity of human consciousness and the natural sciences. This bridge must stretch across quite a chasm, and there are plenty of ideological pressures, encouraging some to deny (or fuzz over) some of the more daunting requirements, and encouraging others to inflate the phenomena or otherwise exaggerate the span of the chasm. So a key element of my proposal is to insist that one side of the bridge be firmly anchored to the same objective perspective that reigns in the sciences that aspire to explain such unproblematically physical phenomena as continental drift, biodiversity and metabolism. This is what I call the third-person perspective, but Evan Thompson suggests it is more properly called *impersonal* (Thompson, 2007). Well, impersonal regarding atoms and molecules and land masses, maybe, but third-personal when dealing with the digestive tract, the muscle-tone and the center-of-mass of a person, for instance.

Moreover – and this is a sort of 'constructivist' point whose importance I haven't sufficiently highlighted in the past – the physical side of the bridge must be anchored in *conservative* physics (and biology): no morphic resonances, vitalistic vibrations, hitherto unidentified force fields or radically emergent bottom–up causal powers (such as those dimly imagined by John Searle) are to be introduced without a thorough articulation and defense. There is a simple way of enforcing this rule: whenever suspicions arise, translate the case into robot-talk. One of the great virtues of robots (whose innards – by definition, if you like – do not include any living subcomponents) is that *all* the causation involved is garden-variety causation, well-behaved at the micro-level, however startling the macro-level products are. When a robot amazes us with some new anthropomorphic competence, we can be sure it has nothing up its sleeve. No sub-process within it exploits mysterious affinities or action-at-a-distance or anything like that. Whatever mix of control theory (Eliasmith, 2003) and computer science we settle on, we can be confident that the cascade of hardware and software levels takes us transparently down to the basic micro-mechanical operations of the CPUs and the mechanical operations of all the moving parts that matter – and just *how* they matter will be well understood because the robot was designed and built, not somehow grown by a process somewhat outside our ken – this too can be considered true *by definition* if need be. This insistence on robot-friendly thought experiments and examples is, I have learned, a reliable way of smoking out the mysterians and "biochemistry romantics" who find it comforting to interpose a presumably physical but utterly undelineated *je ne sais quoi*, – a bit of wonder-tissue, as I have called it – whenever the need for a little more. . . *elasticity* in the bridge is felt. I will exploit this demand several times in this essay.

Roy considers the "most important" claim of heterophenomenology to be that ". . . a level of first person description of the conscious dimension of cognitive properties, corresponding to the phenomenological properties of cognition, should be introduced into contemporary cognitive science in order to overcome the explanatory gap problem that it faces in virtue of its commitment to naturalism." Exactly, and he develops this idea of phenomenology independently of the Husserlian tradition first, so that we can assess the extent to which it departs from, contradicts, or is simply orthogonal to, Husserl's vision. The most obvious (in retrospect) departure is that Husserl was strongly opposed to the sort of naturalism I espouse, and any

phenomenologist who finds the project of phenomenology naturalized repugnant is not on the same page. I am heartened to learn, though, that there is serious division within the ranks of those who call themselves phenomenologists, and that my naturalism is not anathema to all of them. (Indeed, when Zahavi says "Phenomenology is a philosophical enterprise; it is not an empirical discipline" he is not at all rejecting the burden of relating this non-empirical enterprise of phenomenology to cognitive science. More on this below.)

> The notion of (cognitive) phenomenological property is taken to designate what there is for a cognitive system in virtue of its cognitive activity and as it is apprehended by this system; in other words to what there is for a cognitive system from the point of view of this cognitive system itself, by contrast with the point of view of an external observer. (Roy, 2007)

I agree with this exactly, including the contrast with the point of view of the external observer, the heterophenomenologist. It is precisely the point of heterophenomenology to honor that contrast, and preserve and protect the point of view of the subject, and then to *convey* the point of view of the subject, the cognitive system, to the scientific enterprise. Sellars' (1963) contrast between the *manifest image* and the *scientific image* can be put to good ecumenical use here: What phenomenology should do is adumbrate each individual subject's manifest image of what's going on with them. The ontology is the *manifest* ontology of that subject. It can be contrasted with the ontology that is devised by the cognitive scientist in an effort to devise models of the underlying cognitive processes. Consider, for instance, the debates in AI over the proper ontology of event representations: the frame problem, opportunities, states and events, etc.[2] *The ontology of a robot* is no oxymoron at all – even if the robot is not intended to be a conscious robot. There is still the level of description in which the robot's data structures are described, and their semantics explained. *What is represented* by these data structures? Compare the classic question: what does the frog's eye tell the frog's brain? (Lettvin et al., 1959) This is *not* a question of frog phenomenology; it is sub-personal, and so is the exploration by Jackendoff (e.g., 2002) and other linguists of the ontology presupposed by the semantics of (ordinary) language. The *manifest* ontology of the subject will surely track the ontologies of these other investigations quite closely – that can be seen by the degree of familiarity or recognition we feel when we read the proposals of Jackendoff, for instance. But their enterprises are logically independent of phenomenology. It is open to them to uncover a counter-intuitive and alien ontology that nevertheless handles human competences better than the 'folk' ontology of the

[2] There are two distinct ontologies to consider at these subpersonal levels: what you might call the *parts catalogue* – the discrimination-mechanisms, buffer memories, data-structures, operations, sub-routines, and the like that compose the model-*maker's* ontology of real things and events in the brain, and the *dramatis personae*, the set of things, events, states, relationships, etc., that these parts are best seen as being *about* – the *model's* ontology of events, opportunities, changes, affordances, objects, situations, agents, beliefs, desires, intentions, nouns, verbs, adjectives, phonemes, ... The latter is close kin to the conscious subject's *manifest* ontology, but we should leave room for the discovery that in some regards *false consciousness* reigns; subject's only *think* that their environment is filled with the things they are aware of, while at the subpersonal level, their brains and bodies are dealing with a *different world*.

manifest image. Roy recognizes that this prospect is possible, and to some degree threatening but discounts it:

> Indeed, the fact that there might be a systematic distortion between the objective reality of cognitive processes and what they subjectively appear to be in consciousness is in no way harmful to the possibility of giving a faithful first person description of these appearances. (Roy, 2007)

This is true, but such systematic distortions would threaten our conviction that we know our own minds, that we are not spectacularly deluded. This is a hot-button theme that has been aired in the recent discussion of the Grand Illusion (Noë, 2002) and the provocatively entitled book, *The Illusion of Conscious Will*, by Daniel Wegner (2002), which inspired one commentator to call him a "cryptobehaviorist" who provided "terrifying interpretations of his experiments" (Bernard Baars, comments made at ASSC7 Memphis, 2004). Heterophenomenology deals with this *political* issue by insisting that cognitive theory is not complete until it takes the phenomenology seriously, so that science can eventually *settle* whether or not any of it is deluded, but one gathers that some defenders of Phenomenology[3] would like to achieve a happy outcome of this investigation by fiat: *as Phenomenologists, we already know the denizens of our minds; they are not phantoms*. Yes, there is a way of securing one's authority to make this claim, but the cost is acknowledging that we *don't* know whether our minds, *thus* characterized, are attached to our bodies in the way we would want.

As Evan Thompson points out, the enlargement of third-personal objective science to include the methods of heterophenomenology is "no mere extension, because it employs methods fundamentally different from the methods of the natural sciences." (Thompson, 2007). That is why I went to such lengths to describe those methods in terms that the natural sciences could – and ought to – honor. It is certainly important that you cannot identify first-person reports and deposit them in the category of useful data (and then turn that data into evidence, as my colleague George Smith likes to insist) without *interpretation* – that is, without adopting the intentional stance. The adoption of the intentional stance is *not,* I argued, an ineliminably subjective and relativistic affair. Rules of interpretation can be articulated, standards of intersubjective agreement on interpretation can be set and met; deviations can be identified; the unavoidable assumption of rationality can be cautiously couched, and treated as a defeasible, adjustable, defensible and evolutionarily explicable assumption. This blunts – and was intended to blunt – the all-too-familiar claim from the humanities and humanistic social sciences that *because* they are "interpretive" enterprises, they are not – and should not be aligned with, held to the standards of, or made to traffic with – the natural sciences.

The allure of the various versions of that protectionist line is strong, and it seems to me that some of the resistance I encounter in these essays to my claim that heterophenomenology makes first-person methods available to the natural sciences is flavored by the misguided desire (of others – not the authors here) to man the

[3] I will persist in my habit of capitalizing "Phenomenology" to refer to the "brand-name" (as Siewert puts it) phenomenology championed by Husserlians.

barricades and keep natural science away from the mind. For instance, I was surprised to see Thompson yielding to the temptation to engage in a little implied guilt-by-association. Does heterophenomenology, as he warns, "amount to a kind of positivism" (Thompson, 2007)? He is entirely right that my proposal is in the spirit of positivism, as I have just explained, but if heterophenomenology is a kind of positivism, it is an as yet untarnished, unrefuted kind of positivism, a kind that is alive and well and deservedly respected wherever science is taken seriously. (At least he didn't accuse of me "scientism.") Yes, this is just what we need if we are to build a sound bridge between subjectivity and natural science.

And am I wrong to see similar ideological motivations behind the various claims that the trouble with heterophenomenology is its *third*-person perspective, when what we need is a *first*- or *second*-person perspective? These criticisms, echoing Jack and Roepstorff (2003), from Thompson, Marbach, Siewert, and Zahavi are otherwise somewhat baffling to me, since they amount – so far as I can see – to bickering over labels for a process of interpersonal interrogation and investigation that is accepted on all sides. Heterophenomenology, I argue, is a cautious, controlled way of *taking subjects seriously*, as seriously as they could possibly be taken without granting them something akin to papal infallibility, while maintaining (contrary to everyday interpersonal communicative practice) a deliberate bracketing of the issue of whether what they are saying is literally true, metaphorically true, true under-an-imposed-interpretation, or systematically-false-in-a-way-we-must-explain. It is respectful of subjects, and can be conducted with as much or as little *I-thou* informality and give-and-take as the circumstances suggest. So heterophenomenology could just as well have been called – by me – *first-person science of consciousness* or the *second-person method of gathering data.* I chose instead to stress its continuity with the objective standards of the natural sciences: "intersubjectively available contents which can be investigated as to truth and falsity." as Alva Noë puts it, so I called it a third-person perspective, and thereby created a target to snipe at – but the critiques are directed at the label, not the method.

I wish I had gone to greater lengths to stress the real-time, individually tailored, interactive possibilities of heterophenomenological method, since these are correctly stressed by several writers. My exposition left readers with a somewhat impoverished sense of the scope of heterophenomenology, but nothing rules these "first-person" methods out. Shaun Gallagher develops the charge most explicitly: "In heterophenomenology, first-person data are averaged out in statistical summaries" (2003, 90), I have certainly never discussed this, so it must be an inference from my silence about it that leads him, and others, to make this (false) claim.[4]

[4] Thompson cites Lutz and Thompson (2003) as an instance of proper first-personal phenomenology and neurophenomenology, but the fact that their method is not 'lone wolf autophenomenology' but rather involves using subjects other than oneself is enough to establish it as third-person science in the sense that I intended. Zahavi (this volume) applauds the *rapprochement* of his kind of phenomenology and cognitive science, but does not distinguish the approach he favors from heterophenomenology in any way that I can see. Gallagher (2003) cites Lutz and Varela with approval, as well as experiments by Decety and Frith and their colleagues, and Braddock, but all of this work falls crisply into the methodology of heterophenomenology. No one, so far as I know, has advocated dropping the ("third-person") prohibition against relying on oneself as the sole subject. Marcel (2003) usefully clarifies the historical background on first-person, second-person, and third-person approaches to science, and the grounds for resisting Cartesian presumptions, for asking when, and why, we can trust ourselves.

As Thompson candidly notes, "On the one hand, there seems to be nothing in the heterophenomenological method itself that disallows using the first-person perspective in this way." But he goes on to say "On the other hand, given its resolutely third-person attitude, there is nothing in heterophenomenology that would lead it ever to envision – let alone take the step – of working with experience in this direct and first-personal phenomenological way." (Thompson, 2007). Well, I don't see why, but if the charge can thus be reduced to something no worse than misdirection, I'll plead guilty so we can get on with more important matters.

Zahavi raises another sort of misdirection that I need to correct before we go on. (Variations on this presumption also affect the essays by Dokic and Pacherie, and Dreyfus and Kelly, but Zahavi's articulation of it permits the most straightforward response.)

> Why does Dennett consider the heterophenomenological world a *theoretical* posit? Presumably because he advocates a version of the theory-theory of mind and considers experiencing a form of theorizing and experiential states such as emotions, perceptions, and intentions, theoretically postulated entities. (Zahavi, 2007)

Not at all. When a subject says "There is a tree in front of a house" he is not *reporting a belief*, he is *expressing a belief* (see the discussion of reporting and expressing in Dennett 1991, 303–9). His sentence is *about* the tree, not his inner state–but of course we can learn something about his inner state from his utterance, just as we can learn something about a person's inner state when he *asks* "Where's the men's room?" or *requests* "Pass the mustard, please." If the subject says "I see a tree in front of a house" he is, officially, reporting a mental state (his current vision state) by expressing a belief *about* it. Subjects in vision experiments–for instance–typically do express beliefs about their current mental states, and this is what experimenters want – they already know what is out in the world, the stimulus or distal cause of the subject's behavior, and want to know something about the proximal causes of that behavior. But if subjects instead express beliefs about distal objects – "The one on the left is brighter." – this doesn't interfere with the experimenter's quest for information about proximal causes. When a subject expresses a belief about *anything*, that belief is a (salient, crucial part of the) proximal cause of the expression. If the investigator wants to answer a question along the lines of "How did the subject come to believe *that then*?" heterophenomenology is the methodology in play. Sometimes the beliefs of interest are *about* such items as mental images, or sensations, or dreams, or other "mental" items. Heterophenomenology no more presumes that these are beliefs than that trees are beliefs, or that Feenoman is a belief (the belief in Feenoman is a belief – Feenoman may be a trio of tricksters or a real god). As we shall see below, our ability to have beliefs about mental items, and not just about their distal causes, is particularly fraught with confusion and puzzlement.

Why resort to fiction?

I learned a lot from Eric Schwitzgebel's essay, a fine example of constructive philosophical criticism. His analysis of what I have said and what baffles him about

it is so searchingly fair and sympathetic that I must grant that the problem is mine, not his. I have failed to make my claim clear, and it is high time I correct this situation.

Schwitzgebel notes that I often make use of the analogy with the interpretation of fiction, and he finds this unhelpful. He quotes the passage where I say that subjects are "unwitting creators of fiction" but he finds this "uncharitable" – it gives him a headache, in fact. "If so, that seems to imply that besides facts about conscious experience there are, in addition, distinct and sometimes contrary facts about 'what it is like'."

> What would be the difference between the two species of fact? And why should we be so uncharitable as to interpret subjects' claims as inaccurate attempts to convey facts of the first sort, or even mere fictions, rather than accurate attempts to convey facts of the second sort? (Schwitzgebel, 2007)

How can I insist that it is phenomenology that I am proposing to explain when I go on to say that phenomenology is *fiction*? ("But what about the actual phenomenology? There is no such thing." Dennett 1991, 365, see also 95). Schwitzgebel sees that I acknowledge that people don't *take themselves* to be creating fiction. And the only interpretation he can find is the minimalist – throwaway, you might say – version in which what people get to be dictatorial authorities on is just the *wording* of their report. Paltry indeed, as he says. And that is all he can see.

"Subjects do not mean to be writing fiction, and it is distortive to reinterpret what they are doing as creating fiction." (Schwitzgebel, 2007) He looks closely at what I say, and comes up with what is apparently a mistake: my proposal "seems preposterous: How could having dictatorial authority over an *account* of something be tantamount to having dictatorial authority over 'what it is like to be you', that is to say, over one's phenomenology or conscious experience?" (Schwitzgebel, 2007) How indeed? Here is how: by granting the subject dictatorial authority over the (unwitting) metaphors in which that account is ineluctably composed. *That* is the fictive element.

Imagine, if you can, a primitive people with language, but with no experience of technology at all – not even spears and fishhooks. (There have almost certainly never been any such people – since simple hand-axes go back a million years, and language probably doesn't, but never mind – this is just an intuition pump.) As far as these people are concerned, there are just three kinds of things: plants, animals, and unliving things: "rocks". This unholy trinity of basic types is built into their language: if something isn't a plant or an animal it *has* to be a rock – no other options are "conceivable" to them. Now let some of these people be brought to contemporary technological civilization for a day, and be shown radios, cars, blenders, television sets, but also hammers and pencils and other simpler artifacts. They return to tell their comrades about the amazing plants and animals they have seen – slender woody plants with black centers that mark the oddly shaped white leaves that are everywhere to be found, and larger woody plants with amazingly hard flowers with which one can smash things – and then there are the animals – large smelly animals that the people actually get inside and ride, noise-making animals –

including some that also have amazing moving patterns on their skin, like cuttlefish only more so, etc.

What I am asking you to imagine is that this is the best these people can do. In their own minds, these are not metaphors but as accurate a literal description as they can muster. They are trying to tell the truth about an amazing part of the real world, not trying to create fictions, and they don't even recognize that they are unwittingly availing themselves of some rather apt metaphors. If we want to know what this new world was like to them, we had better not translate away all their "misdescriptions" and "metaphors". That *is* what it seemed like to them; that is how they chose to *express* their *beliefs*.

Why do I go to such outlandish lengths to describe heterophenomenological subjects as in a similar bind? Because the phenomena *are* outlandish, far from our everyday ken – in ways we tend to overlook because we cushion our ignorance with a false model. We start by reminding ourselves of something familiar and well understood: a reporter sent out to observe some part of the external world – a nearby house, let's say – and report back to us by telephone. He tells us that there are four windows on the front of the house, and when we ask him how he knows; he responds "because I *see* them, plain as day!" We typically don't think of going on to ask him how the fact that he sees them explains how he knows this fact. We tacitly take the unknown pathways between open eyeballs and speaking lips to be secure. Because we all can do it (those of us who are not blind) we don't scratch our heads in bafflement over how we can just open our eyes and then answer questions, with high reliability, about what is positioned in front of them in the light. Amazing! How does it work? (Imagine if you could just spread your toes and thereby come to have breathtakingly accurate convictions about what was happening in Chicago. And imagine not being curious about how this was possible. How do you do it? Not a clue, but it works, doesn't it? Faced with your accounts of what it was like to have these convictions, we would be wise to adopt the agnosticism of heterophenomenology; we wouldn't automatically export our standard presumptions about your reliability in other matters of reportage to this new setting.)

The relative accessibility and familiarity of the *outer* part of the process of telling people what I can see – I know my eyes have to be open, and focused, and I have to attend, and there has to be light – conceals from us the utter blank (from the perspective of introspection or simple self-examination) of the rest of the process. How do you know there's a tree beside the house? Well, there it is, and I can see that it looks just like a tree! How do you know it looks like a tree? Well, I just do! Do you compare what it looks like to many other things in the world before settling upon the idea that it's a tree? Not consciously. Is it labeled "tree"? No, I don't *need* to 'see' a label; besides, if there were a label I'd have to *read* it, and know that it labeled the thing it was on. I just know it's a tree. Explanation has to stop somewhere, and at the personal level it stops here, with brute abilities couched in the familiar intentionalistic language of knowing and seeing, noticing and recognizing and the like. Phenomenologists may enrich the vocabulary of the personal level, and may tease out aspects of the patterns of competences, inabilities,

needs and methods of persons in illuminating ways, but this is all just setting the specs – the competence model – for the subpersonal level account of how the performances are achieved.[5]

The standard presumption is that "I know because I can see it" is an acceptably complete reply when we challenge a reporter, but when we import the same presumption into the case where a subject is reporting on mental imagery or memory imagery, for instance, we create an artifact.[6] We ask a subject to tell us how many windows there were in the front of the house he grew up in, and he closes his eyes for a moment and replies "four." We ask: How do you know? "Because I just 'looked'. . . and I 'saw' them!" But he didn't *literally* look. His eyes were closed (or were staring unfocused into the middle distance). The "eyes part" of the seeing process wasn't engaged but, it seems, a lot of the rest of the process was–the part that we normally don't question. It's sort of like seeing and sort of *not* like seeing, but just how this works is not really very accessible to folk-psychological exploration, to introspection and simple self-manipulation. When we confront this familiar vacuum, there is an almost irresistible temptation to postulate a surrogate world – a mental image – to stand in for the part of the real world that a reporter observes. And we can be sure of the existence of such a surrogate world in one extremely strained sense: there has to be *something in there* that reliably and robustly maintains information on the topic in question since we can readily confirm that information can be extracted from "it" almost as reliably as from real-world observation of a thing out there.[7]

The "recollected image" of the house has a certain richness and accuracy that can be checked, and its limits gauged. These limits give us important clues about how

[5] Drummond presents a lucid and sympathetic account of the Phenomenologists' self-imposed exile from naturalism and causal explanation, and suggests, ingeniously, that since I have maintained that natural science needs to posit theoretical fictions – beliefs, selves, and the other items revealed to the intentional stance – in order to make sense of significance, my own physicalist metaphysics begins to look like a bit of "ill-envisaged dogma" (quoting me). This permits me to highlight the value of my aligning *my* theoretical fictions with those of, say, physics – centers of mass, equators, parallelograms of forces – since it is not just the complexity of the mind (or significance) that encourages and justifies the adoption of such useful fictions. Drummond says that "for Dennett the relation between the intentional and physical accounts remains obscure, whereas the phenomenological program has a specific way of locating the scientific or empirical within the phenomenological" (Drummond, 2007). Does it? It "locates the logical space within which the empirical account has its validity" (Drummond, 2007), but I do not see how it addresses any of the "specific" problems that must be solved for us to move comfortably back and forth between that "logical space" (descriptions of the world of subpersonal processes) and the logical space of phenomenological descriptions of experience. He quotes and rejects Carr's "paradoxical" opinion (1999, p135) that these two descriptions are "equally necessary and essentially incompatible" (Drummond, 2007) but he says nothing in detail about how his "location of the logical space" resolves the problems that inspired the opinion.

[6] Siewert describes this process in his excellent criticism of Brentano's theory of consciousness as inner perception. He notes that in "the outer case" we can draw a distinction between "a *presentation of the object*, and both: *the object that is presented,* and a *judgment about the object*" but no analogue of these distinctions can be drawn in "the inner case." (Siewert, 2007). I commend Siewert's discussion of these issues, and see little disagreement between us. Such as remains would take a lengthy exposition, and I have decided to devote this essay to the more damaging disagreements and misunderstandings.

[7] Cf. my discussions of Popperian creatures, who try out their hypotheses against an inner model of the world, thus allowing them safer passage than mere Skinnerian creatures, who don't get to "look before they leap," Dennett (1995, 1996).

the information is *actually* embodied in the brain, however much it *seems* to be embodied in an "image" that can be consulted. This is where the experimental work by Shepard, Kosslyn, Pylyshyn and many others comes into play.

From this perspective, our utter inability to say what we're doing when we do what we call framing mental images is not so surprising. Aside from the peripheral parts about what we're doing with our eyes, we are just as unable to say what we're doing in a case of seeing the external world. We just look and learn, and that's all we know. Consider the subpersonal process of normal vision and note that at some point it has to account for the fact that internal cortical states suffice to guide a speaking-subsystem in the framing of descriptive speech acts. We are making steady progress on this subpersonal story, even if large parts of it remain quite baffling today, and we can be confident that there *will be* a subpersonal story that gets all the way from eyeballs to reports and in that story there *will not be* a second presentation process with an inner witness observing an inner screen and then composing a report.[8] As I never tire of saying, the work done by the imagined homunculus has to be broken up and distributed around (in space *and time*) to lesser agencies in the brain. Well then, what diminutions, what truncations, of the observing reporter might do the trick? An agent that was full of convictions but clueless about how he came by them – rather like an oracle, perhaps, beset with judgments but with nothing to tell us (or himself) about how he arrives at that state of belief. Or a would-be reporter who has been blinded but doesn't realize it, because, *mirabile dictu*, he has a seeing-eye companion who *tells* him plenty to report, so much so fast that he is tricked into thinking he can actually still see. But these are just crutches for the theorist's imagination, impressionistic ways of easing the passage from personal level humanity to subpersonal level machinery by creating intermediate levels. More promising and realistic subpersonal agencies would be less like us persons and more like neural machines.

Recall Shakey, the robot that moves boxes and pyramids around. I imagined giving Shakey the capacity to tell us how it tells the boxes from the pyramids, and it did so by saying it drew line drawings and then examined the vertices of these drawings looking for the tell-tale signs of boxes.[9] But in fact, Shakey was not *actually* but only *virtually* drawing line drawings. Shakey's way of talking is either false (a fiction, however illuminating) or metaphorical. If a human subject says she is rotating a figure in her mind's eye, this is a claim about what it is like to be her,

[8] There could have been; we could have discovered, surgically, that there is a control room in the brain, inhabited by an inner agent, like the tiny green alien sitting in the control room behind the hinged face of a bald "corpse" in the morgue in the movie *Men in Black*. In that case, there would literally have been a Cartesian Theater. But we know, from empirical investigation, that there is no such place in the brain, and nothing functionally equivalent to it, either.

[9] I slid past the technical question of just how one might extend Shakey's very limited capacity to compose "speech acts" to include reports that were informed by one or another level of its visual operations (Dennett 2001, 92). Why didn't I go into the details? Because although the existing techniques for controlling human–computer linguistic interfaces at the time were hugely unrealistic as models of human speech act production (so an explicit account would bog the reader down in irrelevant details), there was no clear sign – to me – of any principled barrier or obstacle to improving and extending the techniques into the human (or at least informatively humanoid) range of competence. This lack of explicitness on my part might, of course, harbor a fatal flaw in my example, but to my knowledge no one has developed this possible objection.

and indeed she intends to describe, unmetaphorically and accurately, a part of the real world (a part that is somehow inside her) as it seems to her. But she is speaking at best metaphorically and at worst creating an unwitting fiction, for we can be quite sure that the process going on in her head that inspires and guides her report (to put it as neutrally as possible) is *not* a process of actual image rotation. It is, perhaps, *virtual* image rotation. Now there are several logical possibilities here.

(1) Her speech acts are not real speech acts at all, and have no interpretation, let alone a truth-preserving one – she is just babbling. Or
(2) Her speech acts are deliberate lies. Or
(3) Her speech acts are *intended* to be factual and non-metaphorical, but are utterly without truth-preserving interpretation – she is telling a bald fiction, an *unwitting* lie, and there is nothing in the real world that they are even *arguably* about. Or
(4) She is telling us something that has a truth-preserving interpretation *cum grano salis*, if we just squint a little and let metaphor pass for literal truth.[10] Or, of course,
(5) She is telling us the truth about some *other* realm, not the goings on in her brain, but the goings on in some other medium. (Dualism, in short.)

Heterophenomenology is officially neutral at the outset, ready to discard some vocalizations and other gestures via (1), and leaving to third-person science the discovery of which of (2–5) might be the case about specific cases. Our only way of being neutral at the outset is to take her at her word, as best we can interpret it. She, the subject, is not authoritative *at all* about which of 3–5 might be the case. (I'll allow her to be authoritative about whether she is deliberately lying or not.)

When Evan Thompson discusses this issue, he says my insistence on interpreting a subject's remarks as somehow *about subpersonal brain events* belies the neutrality I claim: "The bias of this approach is that it demands we interpret subjects as expressing beliefs not simply about 'what is going on inside them' but about 'what is going on inside them *subpersonally*' " (Thompson, 2007). As he notes,

> Descriptive reports carry no *particular* [my emphasis – DCD] commitments on the part of the subject about what is going on in his or her brain. . . . .One is describing one's subjectivity at the personal level in a way that is completely noncommittal about the subpersonal workings of one's brain. (Thompson, 2007)

[10] This is standard practice in computer circles, where virtual machine talk is allowed to pass for the truth with no raised eyebrows, no need for the reminder that it is, officially, just metaphorical. A close kin to this interpretation is to treat the speech acts as making *topic-neutral* claims, which can be considered candidates for *literal* truth if there is a mutual understanding that they are to be interpreted *functionalistically*. Do people speak the truth when they say they are rotating images in their mind's eyes? It will depend on a judgment call about the fit between their claims and what is subsequently learned. (This is like the ultimately political question of whether the shaman was *right* after all when he said his patient was inhabited by demons – the demons turn out to be *bacilli* or *protozoa*.) Thanks to Andrew Jewell for pressing this alternative.

True enough, but that word 'particular' is papering over a crack. There are just three ways to fill that crack.

(1) If you yourself are a materialist, then your words do carry an extremely general and open-ended commitment about what is going on in your brain: *something* is going on in your brain that bears some sort of striking resemblance to the events you've just described, because you firmly believe that you were caused to have your current reportorial intentions (to express your current experiential beliefs) by brain events–not liver events or events happening in some other realm. Once we find out just which events in your brain did drive your experiential beliefs, you will discover *in particular* what you were talking about. If it turns out that materialism is false, you will be proven to have deluded yourself about this – you have *in fact* been talking all along about events occurring in the *whoosis* realm (name your poison). But your theoretical delusion was curiously non-destructive of your competence as a heterophenomenological subject. It cancelled out neatly. (1b) There is a mirror-image story to be told about the committed dualist, who may or may not prove to have been right about whatever kind of events *not* in the brain he thought he was talking about. In either case, no harm done. Live and learn.

(2) If you are resolutely agnostic about materialism and dualism, then at least your reports carry some minimal *reality* commitment. You are implying, by your serious attempt at phenomenological reportage, that you are telling about something that actually has just happened, that you have done, that really occurred – you're quite sure you are not making it up, for instance. You may insist that you haven't a clue about where this happening happened ("well, it happened *in my mind* – that's all I can say."), and what kind of stuff was involved in this happening, but this has the implication that if somebody can come up with a plausible candidate for what you were actually talking about (and you mustn't complain about that – you are insisting that you really don't know what you were talking about!), you will be in the delicate position of one who might be thus instructed: unbeknownst to you, you were *in fact* talking about events in your brain. See how nicely the details of these information-transformations we have plotted in your brain fit the details of your phenomenological reports. We've *found out* what you were talking about! (And nothing approaching "analytical isomorphism" need obtain for this verdict to be sustained.)

(3) If you are convinced that you know jolly well – incorrigibly even – what you are talking about, and you are definitely *not* talking about brain events at all, but *rather* about events that have a different ontological/metaphysical status altogether, as *intentional objects constituted at the personal level,* you insist that it would be a category mistake to identify them with either brain events or ectoplasmic events. Except when you happen to be thinking about, or inspecting (with an autocerebroscope) what is happening in your brain, your *noemata* are not even candidates for identification with brain things. In this case, you are in the somewhat different delicate position: you may be instructed that while you were talking about *those* things (we'll let you be the authority on their ontological status – see Drummond for some of the problems), we have found your talk to be gratifyingly informative about some *other* things that we have discovered in your brain, and this is no coincidence. From our point of view, you have been creating a really very useful theorist's fiction, a heterophenomenological world. (Imagine that Truman Capote suffers from

Multiple Personality Disorder – now known as Dissociative Identity Disorder. Tru, the novelist, writes *In Cold Blood* (1965), sincerely believing it to be entirely and straightforwardly a work of fiction, just like *Breakfast at Tiffany's* (1958). He supposes that Perry Smith and Dick Hickock and the Clutters are pure creatures of his fancy, just like Holly Golightly. We discover that Tru has an alter, Man, who has done all the research, interviewing the good citizens of Kansas, visiting the convicts in prison, attending the execution. Unbeknownst to Tru, Man has guided the 'novel' writing in many or all of its details. Is the resulting book truth or fiction? Tru may sincerely insist that he was never talking about the real Clutters, the real Dick Hickock, and in a sense he is right. And in a sense he is wrong.)

Now heterophenomenology is neutral with regard to all three categories. It takes down the details of the heterophenomenological worlds of subjects, and lets the further work in science settle which way to enlighten the various subjects once we get a good theory. As Thompson says, it is important to "keep in clear view the conceptual difference between experiential content at the personal level and representational format at the subpersonal level" (Thompson, 2007) but it is also important to remember that the experiential content at the personal level concerns real events, to put it as neutrally as possible, and we will not have an explanatory theory of consciousness until we can relate that personal level content to the subpersonal events that are responsible for it.[11]

Heterophenomenology by itself is neutral about whether materialism will be vindicated, but as a part of the natural sciences, it starts with the same defeasible bias that is built into *mermaidology,* for instance, which demands that we interpret mermaid reports as accounts of natural phenomena if we can. The only way of showing that mermaids are *not* natural phenomena is to try to account for all the sightings as natural phenomena and fail systematically. The default presumption of materialism is not an objectionable bias.

As noted by Thompson, a somewhat different approach to this problem of interpretation has been suggested by Alvin Goldman (2004), in his latest critique of heterophenomenology. He proposes that cognitive scientists should adopt the "rule of thumb: 'When considering an introspective report, and a choice is available between an *architecturally loaded* interpretation and a *architecturally neutral* interpretation, always prefer the latter.' " (Thompson, 2007) Goldman describes my practice is 'just the opposite" but this is misleading. As I have said, heterophenomenology is neutral between (1–3) but it is not neutral about supposing that the subjects are purporting to express themselves about something really happening. According to Goldman, lay subjects' descriptions are typically "much less fine-grained than those of interest to cognitive science." (Thompson, 2007). I wonder what he can have in mind. When subjects say (in their debriefing in a classic Shepard experiment, for instance) that they were rotating *the figure on the left* in their mental image, the interpretation of these words Goldman recommends is apparently so neutral it doesn't even require that there be, somewhere in the universe, something

[11] Thompson (Thompson, 2007) draws attention to Georges Rey's related distinction between what he calls the "phenomenal mental image" on the one hand and the 'functional mental image' or "depictive structures in the brain" (ms 26) on the other. Heterophenomenology treats these phenomenal mental images as the intentional objects of the subject's reports.

"answering to" (not necessarily referred to by) the definite description "the figure on the left." If so, then his rule of thumb is tantamount to putting everything subjects say into scare quotes – and *not* taking them seriously. One of the phenomena, according to heterophenomenology, that needs to be explained by a scientific theory of consciousness is the subject's ability and inclination to *refer to items and features in mental images*, and any interpretation that excuses itself from this obligation is too neutral by half. Heterophenomenology takes such purported references *almost* at face value – the way we take a novelist's references – by using the category of a theoretical fiction that stands in, *pro tempore*, for face value reference until the science is in.

Heterophenomenology's caution about such reference is further illustrated by contrasting it with Uriah Kriegel's bold attempt to cut through the "footstomping" that besets those who are unapologetic realists about phenomenology in spite of the fact that they are utterly unable to reach agreement about what is, and is not, "phenomenologically manifest." He proposes to tie phenomenology to what is "first-person knowable" and I endorse the cautious account of first-person knowability that he derives from David Pitt (2004). But then he goes on to assert that "there is good reason to believe that it is strictly phenomenologically manifest properties that can be first-person knowable." (Kriegel, 2007), and I cannot find any interpretation of this that doesn't imply that he wants to *explain* the first-person knowability of some item by noting that it is *phenomenologically manifest* – in just the same way that we can explain the reporter-knowability of the fact that the house has four windows by citing their manifest presence in daylight in front of his open eyes. This is to endorse the Cartesian Theater, a place where the manifesting happens and *thereby* informs the knower. We have to turn this picture inside out: when we are struck by first-person knowability (or its deficiencies, in some cases) we need to resist the temptation to postulate what Ryle might call a "paramechanical" explanation – which is no explanation at all. I am reminded of the time Ned Block told me about being a subject in a laterality experiment, doing word/nonword judgments with the target either left or right of fixation. If you are strongly lateralized for language in the left hemisphere (which is normal, unless you are left-handed, in which case the story is more complicated) then you take slightly longer to identify words when they are presented in your left visual field – primarily processed in the right hemisphere – rather than the right. "The words on the left seemed sort of blurry" Ned said, as if this "explained" his longer latency. "Did the words seem blurry because you had difficulty identifying them, or did you have difficulty identifying them because they were blurry?" I asked. Ned realized that he had no experiential or first-person resources for resolving that question. He was caused to believe that the words on the left seemed blurry, but he had no privileged access into the source of this particular bit of first-person knowability and hence couldn't really shed light on the causal mechanisms behind the demonstrated lag in first-person knowability for items on the left. We definitely want to include the blurriness of those words in Block's heterophenomenological world, but we do well to bracket it as a theorist's fiction for the time being.

This may help resolve a further difficulty commentators have had with my use of the fiction trope. Carman, Dokic and Pacherie, and Marbach (if I understand them) all object to my declaration that there are no *real seemings*. Here is what I now want

to say, informed by their reactions: judgments are about the qualia of experiences in the same way novels are about their characters. Rabbit Angstrom sure seems like a real person, but he isn't. He's a fictional character in Updike's tetralogy. Updike's words are perfectly real, but what they are about is not. Dualism (option 1b above) is parallel to the curious view that when novelists write novels they somehow bring into existence (or discover) characters and events and places inhabiting another, non-physical "dimension," the *real* universe of fiction – populated with real seemings. Do these commentators wish to endorse such a view? If materialism is true, there are no real seemings – unless we adopt the Procrustean tactic of overriding subjects' demurrals and *identifying* the subpersonal brain states as those seemings.[12] The alternative is to treat seemings as denizens of a theoretical fiction, characters in the subject's autobiographical novel, the default position of heterophenomenology until we do the science.

Finally, before leaving this topic, I let me make explicit how these points abrogate Max Velmans' claims, since it may not be obvious to everyone how he has misinterpreted me. His main mistake is confusing my base camp with my destination. Heterophenomenology is the neutral standpoint from which I then develop and deploy my occasionally "eliminativist" views, drawing on further considerations and discoveries. Eliminativism is not built into heterophenomenology as a method. "While Dennett is willing to listen to what people have to say about their experiences, he is not prepared to believe what they say." (Velmans, 2007) This is partly true and mainly false. I am quite eager to find an interpretation where I can believe what they say, but this must be a matter of some further discovery and negotiation. Nobody gets to be pope, so I am not "prepared" to believe what they say in advance, and if I were to "believe what they say" while imposing an "architecturally neutral" interpretation on what they say, I would be merely paying lip service– acknowledging their sincerity, which I am happy to do in almost every case. (I certainly don't presume that subjects typically lie.) Velmans also says that I think that "since their subjective worlds are not real, subjects' beliefs about their qualitative nature are false." Again, partly true, and mainly false. The subjective world is not to be confused with the real world, but that does not mean that it is not by and large composed of truths, in two senses. There are the truths-in-fiction (analogous to the truth that Sherlock Holmes is a man who lives in Baker Street, London) and then there are the truths, embedded in fictions, about the real world. (For instance, E. L. Doctorow's book *Ragtime* contains lots of truths about 20th century America, but it is a novel. Jane Austen's novels contain a bounty of truths about the human condition.) Suppose you have just seen an afterimage of an American flag, caused by staring at a green, black and yellow flag image for a few seconds. Just as the fictional Sherlock Holmes can be correctly described as taller than the real Max Beerbohm, the fictional red stripes on your afterimage can be correctly described by you as somewhat more orange than the real red stripes on the

[12] There are definitely reasons to identify a host of utterly unrecognized subpersonal brain events as *unconscious* seemings of a sort–they are content discriminations that lead to behavioral adjustments without ever achieving cerebral celebrity–but these are not the real seemings defended by these authors. See Siewert's long footnote 9 (this volume) on this topic, which expresses some disagreements that *may* be dissolved by these remarks.

flag flying outside the window. Fiction is parasitic on fact, and the afterimage stripe is red in exactly the same way that Holmes is tall.[13]

Putting the squeeze on autophenomenology

Schwitzgebel sees another problem with my insistence that subjects *do* have a limited incorrigibility about how it seems to them. He notes my own campaign to show how wrong people often are about their own consciousness, and concludes – correctly – "It turns out ordinary people aren't such great authorities on what it's like to see," and *that* conclusion is in flat contradiction with the claim that people have *any* domain of Rortian incorrigibility, is it not?

No, I don't think so. A point I make – but not clearly enough, obviously – is that when people *generalize* or *theorize* about what it's like, they cede their authority altogether. Alva Noë puts it well: "I *can* be mistaken about the nature of my experience – about how I, in experience, take things to be. . . . . But it would be a different kind of mistake for me also to be mistaken about how I *take* my experience to be. I can be wrong then about how things seem but not wrong about how I take things to seem." (Noë, 2007) If somebody says her visual field *seems* detailed all the way out to the periphery, which lacks a perceptible boundary, there is no gainsaying her claim, but if she goes on to theorize about "the background" (or as Searle would say, "the Background"), and claims – for instance – that *there are* lots of details in this background, she becomes an entirely fallible theorist, no longer to be taken at her word. But then how do we (how does she – how do the heterophenomenologists studying her) distinguish theorizing from the more naive or at least theoretically neutral attempt to say what it is like now? I don't think there is a good method for drawing this line. I don't think it is possible for there to be perfectly neutral, perfectly theory-free testimony from subjects or theory-free inquiry from researchers.

The ideal of utterly neutral, utterly bracketed heterophenomenology is as unreachable, practically, in the case of hetero- as in the case of auto-. One brings one's current sense of what is unremarkable to the table. In heterophenomenology, the unavoidable practice of using one's own reactions to what the subject says as a backdrop against which to highlight abnormality (which then provides the targets for the next round of interactions) runs a serious risk of distortion, but it also provides the leverage without which one is a merely aimless data-gatherer. The fact that the experimenter has to start with whatever biases structure the quality space of her own experience shapes the trajectory of the chemist and the geologist just as much as it does the heterophenomenologist. One can purify the methods as one goes, if one can develop a model of normal functioning that goes beneath or beyond heterophenomenology and gives one reasons for trusting various aspects of one's 'normal'

[13] Briefly, two other corrections: in his attempt to distinguish his "critical" phenomenology from my heterophenomenology, Velmans says that his view "does not assume that subjects are necessarily deluded and scientifically naive about their experiences." Nor does heterophenomenology. Critical phenomenology is "reflexive," he says: about others *and about oneself*. Heterophenomenology is no different; one can certainly adopt the heterophenomenological method towards oneself, treating oneself as an experimental subject, indirectly.

reactions. If, for instance, one has no sense at all about what a subject might prefer to keep secret, to conceal from the inquirer, one will be unlikely to guess the truth about a blip in reaction times in a word stem completion task for such word stems as *cun-* or *shi-*. There is no apparent end to the way in which shared knowledge – and the special case of common knowledge – might enhance or impede or distort the task of extracting a heterophenomenological world from a subject, but that does not mean that *auto*phenomenology is in any better position. I'm reminded of the comedy riff by George Carlin that begins "You know that cool blast of air in the middle of your brain when an axe splits your head? I love that feeling...." – and then he catches himself and says something like "Oh, perhaps you haven't shared that experience. . ." It certainly helps when experimenter and subject share a lot: not just language but gender, age, socio-economic status, familiarity with baseball. . . . But it has not been shown by any of the critics of heterophenomenology that these impediments of difference cannot be overcome by *mutually cooperative and interactive exploration* – and if it were shown, then the critics of heterophenomenology would discover that they had proven more than they wanted to prove: that a single, unified *first-person* science of consciousness was flat impossible: we'd have to settle for *solipsistic science*. If that is the only alternative, heterophenomenology may look more inviting to them; it should, since I proposed it as the maximally open-minded intersubjective science of consciousness.

The problem with autophenomenology is not that it is (always, or typically) *victim* to illusion and distortion but that it is (always) *vulnerable* to illusion and distortion. That is why it must be quarantined behind brackets. As Roy says, "the problem is not that autophenomenology takes consciousness to be a purely passive form of observation, but that it fails to appreciate its real limits." But he also says: "Heterophenomenology is phenomenology only inasmuch as it is a closet autophenomenology." Well, yes, and I am asking autophenomenologists to come out of the closet and become an accredited part of the scientific enterprise. You don't have to abandon *anything* of value, since the widespread conviction that you have to defend the citadel of the first-person is simply a mistake. And after all, as autophenomenologists you have all along had the burden of making your soliloquies comprehensible to an audience aside from yourself. Phenomenologists don't want to be solipsists, do they? I am making that burden more explicit, and proposing a distribution of labor that should satisfy everyone. In effect, I have been asking phenomenologists: How would *you* align your research with the researches in cognitive science? And when they rise to the challenge of uniting their enterprise with cognitive science, they tend to reinvent heterophenomenology.

But not always. Eduard Marbach takes on my challenge with utter clarity: he proposes a "data-driven answer" to my challenge to point to a variety of data that are inaccessible to heterophenomenology. (Marbach, 2007) He leads with an example, based on one of my own: imagining a purple cow. My subject says (in Marbach's telling) ". . . not only is the re-presented (*purple cow*) consciously given to me as being *not actually present*, but at the same time the conscious experience of so referring to something absent contains *within its very structure an experiential component of not actually performed perceiving (seeing)*." Can I, the heterophenomenologist, handle such a case? Marbach says that he doesn't see how I could begin to understand this experience of non-actuality without abandoning heterophenomenology in favor of

autophenomenology, but I am not persuaded. I don't have any trouble comprehending this subject's report, of course. Put into somewhat more rustic language, I take my subject to be telling me something like this: "It's not as if I was hallucinating a purple cow (in which case, I might worry about its stepping on me and so on) – I know I'm just imagining it, not actually seeing it, even though it does strike me as rather *like* seeing a purple cow; for instance, the cow's facing left, and about ten feet away from me. . ." As the heterophenomenologist, I myself have no difficulty recognizing this difference my subject speaks of. I have noted it myself, on many occasions. It's an unavoidable part of growing up knowing how to use words like "imagine" and "hallucinate." Aha! But then my own first-person point of view is *presupposed* in my ability to be a heterophenomenologist!

That is Marbach's point, I take it, but I don't think it accomplishes the task of supporting his final conclusion: "no heterophenomenology without autophenomenology." For after all, talking robots would have to have undergone roughly the same *Bildung* for them to be able to converse with us about such matters, and a talking robot could be a fine surrogate heterophenomenologist. In virtue of its capacity to converse with us, it would have it's own "first-person point of view" (I would say) that it used in the course of its heterophenomenological exploration of subjects' mental lives. But would its *so-called* first-person point of view be anything like ours? Could it really conduct heterophenomenological investigations of us if it wasn't conscious *the way we are*? Notice that this challenge begs the question, invoking an imagined solution to the problem of other minds. How do we know that *we* are all conscious in the same way? The childhood taunt "It takes one to know one" can be put to new use here, as the tacit assumption that must be made explicit if it is to do any work.

One way to make this challenge clearer is to imagine a situation in which you are engaged in heterophenomenological inquiry and are asked to explain what you are doing by visiting "Martians" or "robot" scientists. What, they ask, is the point of your exercise? What kind of reverse engineering is this? It is no doubt hard to imagine how you could couch your answer without assuming that the Martians (or robots) are conscious in something *rather* like the way we are, but this is – so far as I can see – a negligible sociological or psychological fact, on a par with the fact that if they worked entirely in binary arithmetic, it might be hard for them to imagine why we were so comfortable with, and insisted on using, base-10 arithmetic. We'd have to work around the mismatch in habits.[14]

Marbach's discussion shows that it would be no small undertaking to create a robot that could look at pictures (*as* pictures) and use them – or ignore them, when it made sense to do so – as representations of parts of the world in which it operated. It would have to see the pictures both as straightforward objects, part of the furniture of the world, you might say, while also seeing them as representing other parts of the furniture of the world, real or fictitious. Marbach points to some subtleties that would have to be incorporated into its understanding. Consider, for instance one of my favorite impressionist paintings, Gustave Caillebotte's *Les raboteurs de parquet*, the floor-planers (Fig. 1).

[14] For more on this, see "Scientists from Mars" in *Sweet Dreams*, chapter 2.

Fig. 1 Gustave Caillebotte, *Les raboteurs de parquet*

We see the stripes as alternating between bare whitish wood and varnished dark wood, almost black. Where the light reflects off the varnish the "black" stripes are actually whiter than their neighboring "white" stripes – see especially at the bottom center of the painting, where the "light" stripes are actually much darker – on the canvas – than the "shiny black" stripes they alternate with. A wonderful inversion. Paraphrasing Marbach parallel claim (about a "yellow" cow in a painting, Marbach, 2007), I see the "white" stripes *in the mode of non-actuality*. . . . the stripe's appearing white–in contradistinction to the perceptually appearing stripe's being white – only appears so because of my consciously taking the pictorial object "stripe" to be a representation of an absent (real or imaginary) strip of planed wood on a floor. A robot that was capable of (or susceptible to) this effect, and that could reflect on it and report on it would be a marvel, but not an impossibility, and we'd confirm it all without any abandonment of the third-person perspective of science (and engineering).

But Marbach has a further point to press: He says (Marbach, 2007) that whereas a perceived cow's appearing yellow to him

> depends for its appearing [yellow] on my perceptual apparatus in ways that science elaborates in detail, it is *not* the case that the real cow's appearing to be yellow is appearing so only because of *my taking it as appearing yellow* in virtue of an appropriately structured conscious experience.

This denies what I assert (if I understand him), but I don't see what grounds Marbach gives for this adamant assertion. He says "the manifest presence of the sensory or phenomenal quality 'yellow' in a corresponding visual experience is a given, a datum, in my waking life." and I am tempted to diagnose here a lapse into theory on Marbach's part. This is like Otto insisting on real seeming (Dennett 1991, 362ff), and when Marbach plays the role of subject I mark his insistence as a bit of heterophenomenological report that may itself be quarantined thus: "It seems to Marbach–it seems very very vividly to Marbach – that this is a 'given, a datum, in his waking life.'" (Compare this with "It seems very vividly to Jones that his visual field is about equally detailed, all the way out.") In spite of his assertion, Marbach and I are really very close to agreement now. He deftly anticipates that this is how I

would respond to his claim, incorporating it into my heterophenomenological account of him, and says he doesn't object to my claim that the truth of his conviction "must not be presupposed by science". But this leads him to "a crucial point": the data he thus provides me are "not owed to naive everyday introspection." I agree – up to a point: I agree that there are data I have been ignoring that "must be acquired through reflection by each and everyone upon the structures of his or her conscious experiences.. .. " because "certain kinds of consciousness can only be understood from the reflective perspective itself." (Marbach, 2007) This as a significant criticism of my earlier accounts of heterophenomenology. I didn't say, or even suggest or imply, that the heterophenomenologist might well avail himself (or itself!) of the subtle distinctions elaborated by Husserl and others, and then *use these to direct* the investigation, getting subjects to reflect for themselves on how their experiences unfold. I seemed to be saying, on the contrary, that the probing of subjects by heterophenomenologists could be well-conducted by rather passive and untutored inquirers. Not so. I am happy not just to concede but to insist that many of the brilliant reflections of Husserl and Husserlians ought to be exploited to the full in heterophenomenological research. I just want to strip them of the anti-naturalistic ideology that has – for the most part – weighed them down (see my discussion of Roy in the first section). But I would say that this shows that we can salvage all the good ideas of Phenomenology and incorporate them into heterophenomenology. That is part of my project, and that is why I call the result heterophenomenology. The fact that Marbach quotes Gregory and Gibson on picture-seeing, shows, ironically, that you don't *have* to be a practicing Phenomenologist of the Husserlian school to appreciate these points.

Finally, I would like to comment on Marbach's apparent metaphysical commitments. Early in his essay he asserts that "what consciousness in itself consists in, as lived conscious experience of something of one kind or another, is nothing to be discovered *out there* in the objective world." The implication seems to be that only its *causes* and *effects* can be found (out there) by heterophenomenology, which can thus never shed light on the real thing in itself. Robert Kaplan (2000), in *The Nothing that Is: A Natural History of Zero*, celebrates the power of what he likes to think of as a Newtonian point of view:

> For in working on gravitation, Newton decided to stop asking what it *is* (a fluid, a substance, a force?) and ask instead *how it behaves*. By shifting his focus from the medieval question to a much more abstract and dynamic one, he was able to discover that bodies under gravity's influence, attracted on another inversely as the square of the distance between them. This proved in the end to be much more useful for understanding the world and predicting the positions of bodies in space. (p 141)

Gravity is still in some measure mysterious, but it is not as mysterious as it used to be, thanks to Newton. I think that consciousness too needs this Newtonian point of view. By asking *how it behaves*, by examining its indisputable causes and effects "*out there* in the objective world," we can escape our medieval stalemate about "what consciousness in itself consists in" and actually explain consciousness. This can be illustrated by considering Thompson's discussion (Thompson, 2007) of the Husserlian concept of *prereflective self-consciousness*, "that feature constitutive of

subjective experience" also known as "implicit awareness". Thompson observes: "In my visual experience of the wine bottle, I am explicitly aware of the bottle, but also implicitly aware of my visual experience of the bottle." What *is* this remarkable implicit awareness or prereflective self-consciousness? Thompson doesn't say, and neither does Marbach. I want to substitute "Newtonian" questions: What does implicit awareness *do?* What does prereflective self-consciousness *do?* What does the presence of this feature enable in a subject that would otherwise not be enabled? What kinds of things can a subject do that she wouldn't be able to do if it weren't for the gift of prereflective self-consciousness? For instance, might a robot have everything *except* pre-reflective self-consciousness? And would this deficit be somehow manifest in the robot's performance, in its inability to do something, notice something, remember something, infer something? Wouldn't *some* sort of oblivion have to be imputed to something that sadly *lacked* implicit awareness or prereflective self-consciousness? But once we know what sorts of competences are supposedly enabled by this feature, we will have a guide to how to add it to the robot's competences.

Critics will pounce on this as "behaviorism" and indeed it *is* a sort of behaviorism – the Newtonian sort, which handsomely sets the pace for physics, astronomy, meteorology, geology and biology, for instance, accounting for all the "behavior" of all the phenomena and their smallest and most inaccessible parts and declaring that this is all that needs to be explained – or could ever be explained. None of the well-known critiques and "refutations" of psychological – e.g., Skinnerian or Thorndikean – behaviorism or logical – e.g., Rylean or Wittgensteinian – behaviorism lay a glove on this ideologically bland but anti-mysterian methodological doctrine. The alternative to this behaviorism is, as Kaplan would say, a medieval view. It is the view that prereflective self-consciousness, the "feature constitutive of subjective experience," is a marvelous private gift that leaves no traces in the world, but that sharply – essentially – distinguishes those blessed with it, the conscious seers and hearers and enjoyers, meaners and actors, from the mere zombies or robots that only *seem* to be seeing and hearing, enjoying, meaning and acting.

I want to suggest, moreover, that for all their talk about what "consciousness in itself consists in, as lived conscious experience," Phenomenologists are already committed to this bland sort of behaviorism without realizing it, since they themselves have nothing to say about this feature beyond what it does. As I noted before, they don't have any performance models to implement their competence models. Indeed, a pure Phenomenological account is a particularly noncommittal sort of competence model, leaving *all* the grubby details of implementation to some later investigation that is not even outlined. It can be contrasted with the sort of phenomenological reverse engineering that Douglas Hofstadter and his students have engaged in (Hofstadter, 1995, French, 1995) for instance, in which the goal of implementing the features in a real working model constrains and provokes the imagination of the theorist (Dennett, 1996, expanded in 1998).

Conclusion

The problem of spanning the various explanatory gaps between the (first-)personal level and the subpersonal level of the natural sciences is about as difficult a problem

as science – or philosophy – has ever faced. Part of what makes it difficult is that in addition to the complex factual puzzles about how the brain works, and the conceptual problems about how solutions to *those* puzzles would – or would not – resolve the puzzles about what our experiences are, our confusions are exacerbated by what might best be called political pressures, and these are of several kinds. The least presentable, but still entirely understandable, are the pressures of interdisciplinary protectionism, which can lead to wanton misreading and caricature. I am pleased to see only faint traces of these pressures in the essays in this issue. More defensible, but still to be resisted, are the quite reasonable anxieties about whether we might hate what we eventually learned about our own brains and minds, and these anxieties promote wishful thinking *on all sides.* They help motivate both mysterian declarations (Levine, 1983, 1994, McGinn, 1999) about Hard Problems (Chalmers, 1995, 1996) and counter-declarations of overly optimistic materialism (the Churchlands, but also most cognitive neuroscientists). To me, one of the most interesting reactions to my heterophenomenology has been the frank acknowledgment, by more than a few cognitive scientists, that they hadn't appreciated how attractive dualism was until they saw the elaborate lengths to which I had to go to make room for subjectivity in the material world! The Cartesian vision–Cartesian materialism or the original Cartesian dualism – is undeniably compelling, and once we see that there is no – can be no – Cartesian Theater, we have to find a safe haven for all our potent convictions. It sure seems as if there is a Cartesian Theater. But there isn't. Heterophenomenology is designed to honor these two facts in as neutral a way as possible until we can explain them in detail.

References

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*, 200–219.

Chalmers, D. (1996). *The consciousness mind*. New York: Oxford University Press.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little Brown.

Dennett, D. (1995). *Darwin's dangerous idea: Evolution and the meanings of life*. New York: Simon & Schuster.

Dennett, D. (1996). *Kinds of minds*. New York: Basic Books.

Dennett, D. (2003). Who's on first? Heterophenomenology explained. *Journal of Consciousness Studies, 10*(9–10), 19–30 (reprinted in A. Jack and A. Roepstorff (Eds.), *Trusting the subject?* Volume 1. (19–30). Exeter: Imprint Academic, 2003).

Eliasmith, C. (2003). Moving beyond metaphors: Understanding the mind for what it is. *Journal of Philosophy, 100*, 493–520.

French, R. (1995). *The Subtlety of Sameness*. Cambridge, MA: MIT/Bradford.

Gallagher, S. (2003). Phenomenology and experimental design. *Journal of Consciousness Studies, 10*(9–10), 85–99.

Goldman, A. (2004). Epistemology and the evidential status of introspective reports; (ref in Thompson).

Hofstadter, D. (1995). *Fluid Concepts and Creative Analogies*. New York: Basic Books.

Jack, A. I., & Roepstorff, A. (2002). Introspection and cognitive brain mapping: From stimulus-response to script-report. *Trends in Cognitive Science, 6*, 333–339.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. New York: Oxford University Press.

Kaplan, R. (2000). *The nothing that is: A natural history of zero*. Oxford: Oxford University Press.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE, 47*, 1940–1959.

Levine, J. (1983). Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly, 64*, 354–361.

Levine, J. (1994). Out of the closet: A Qualophile Confronts Qualophobia. *Philosophical Topics, 22*, 107–126.
Marcel, A. J. (2003). Introspective report: Trust, self knowledge and science. *Journal of Consciousness Studies, 10*, 167–186.
McGinn, C. (1999). *The mysterious flame: Conscious minds in a material world*. New York: Basic Books.
Noë, A. (2002). *Is the visual world a grand illusion?* Exeter: Imprint Academic.
Piccini, G. (2003). Data from introspective reports: Upgrading from common sense to science. *Journal of Consciousness Studies, 10*, 141–156.
Pitt, D. (2004). The Phenomenology of Cognition: or what is it like to believe that *p*? *Philosophy and Phenomenological Research, 69*, 1–36.
Roy, J.-M. (2007). Heterophenomenology and phenomenological skepticism. *Phenomenology and the Cognitive Sciences*, *6*(1–2).
Sellars, W. (1963). *Science, perception and reality*. London: Routlege & Kegan Paul.
Siewert, C. (2007). In favor of (plain) phenomenology. *Phenomenology and the Cognitive Sciences*, *6*(1–2).
Thompson, E. (2007). Look again: Phenomenology and mental imagery. *Phenomenology and the Cognitive Sciences*, *6*(1–2).
Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: MIT.