

Chapter 26

The Fantasy of First-Person Science

Daniel C. Dennett

A week ago, I heard James Conant give a talk at Tufts, entitled “Two Varieties of Skepticism” in which he distinguished two oft-confounded questions:

Descartes: How is it possible for me to tell whether a thought of mine is true or false, perception or dream?

Kant: How is it possible for something even to *be* a thought (of mine)? What are the conditions for the possibility of experience (veridical or illusory) at all?

Conant’s excellent point was that in the history of philosophy, up to this very day, we often find philosophers talking past each other because they don’t see the difference between the Cartesian question (or family of questions) and the Kantian question (or family of questions), or because they try to merge the questions. I want to add a third version of the question:

Turing: How could we make a robot that had thoughts, that learned from “experience” (interacting with the world) and used what it learned the way we can do?

There are two main reactions to Turing’s proposal to trade in Kant’s question for his.

- A. Cool! Turing has found a way to actually *answer* Kant’s question!
- B. *Aaaargh!* Don’t fall for it! You’re leaving out ... *experience!*

I’m captain of the A team (along with Quine, Rorty, Hofstadter, the Churchlands, Andy Clark, Lycan, Rosenthal, Harman, and many others). I think the A team wins, but I don’t think it is *obvious*. In fact, I think it takes a rather remarkable exercise of the imagination to see how it might even be possible, but I do think one

(a written version of a debate with David Chalmers, held at Northwestern University, Evanston, IL, supplemented by an email debate with Alvin Goldman). © Dennett 2001, Feb 15.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-319-72478-2_26) contains supplementary material, which is available to authorized users.

D. C. Dennett (✉)

Center for Cognitive Studies, Tufts University, Medford, MA 02155, USA

can present a powerful case for it. As I like to put it, *we* are robots made of robots—we're each composed of some few trillion robotic cells, each one as mindless as the molecules they're composed of, but working together in a gigantic team that creates *all the action* that occurs in a conscious agent. Turing's great contribution was to show us that Kant's question could be recast as an *engineering* question. Turing showed us how we could trade in the first-person perspective of Descartes and Kant for the third-person perspective of the natural sciences and *answer all the questions*—without philosophically significant residue.

David Chalmers is the captain of the B team, (along with Nagel, Searle, Fodor, Levine, Pinker, Harnad and many others). He insists that he just *knows* that the A team leaves out consciousness. It doesn't address what Chalmers calls the Hard Problem. How does he know? He says he just does. He has a gut intuition, something he has sometimes called "direct experience." I know the intuition well. I can feel it myself. When I put up Turing's proposal just now, if you felt a little twinge, a little shock, a sense that your pocket had just been picked, you know the feeling too. I call it the Zombic Hunch (Dennett 2001). I feel it, but I don't credit it. I figure that Turing's genius permitted him to see that we can *leap over* the Zombic Hunch. We can come to see it, in the end, as a misleader, a roadblock to understanding. We've learned to dismiss other such intuitions in the past—the obstacles that so long prevented us from seeing the Earth as revolving around the sun, or seeing that living things were composed of non-living matter. It still *seems* that the sun goes round the earth, and it still *seems* that a living thing has some extra spark, some extra ingredient that sets it apart from all non-living stuff, but we've learned not to credit those intuitions. So now, do you want to join me in leaping over the Zombic Hunch, or do you want to stay put, transfixed by this intuition that won't budge? I will try to show you how to join me in making the leap.

Are You Sure There Is Something Left Out?

In *Consciousness Explained*, (Dennett 1991) I described a method, *heterophenomenology*, which was explicitly designed to be

the *neutral* path leading from objective physical science and its insistence on the third-person point of view, to a method of phenomenological description that can (in principle) do justice to the most private and ineffable subjective experiences, while never abandoning the methodological principles of science. (CE, p72.)

How does it work? We start with recorded raw data. Among these are the vocal sounds people make (what they say, in other words), but to these verbal reports must be added all the other manifestations of belief, conviction, expectation, fear, loathing, disgust, etc., including any and all internal conditions (e.g. brain activities, hormonal diffusion, heart rate changes, etc.) detectable by objective means.

I guess I should take some of the blame for the misapprehension, in some quarters, that heterophenomenology *restricts itself* to verbal reports. Nothing could

be further from the truth. Verbal reports are different from all other sorts of raw data precisely in that they admit of (and require, according to both heterophenomenology and the 1st-person point of view) interpretation as speech acts, and subsequent assessment as expressions of belief about a subject's "private" subjective state. And so my discussion of the methodology focused on such verbal reports in order to show how they are captured within the fold of standard scientific ("3rd-person") data. But all other such data, all behavioral reactions, visceral reactions, hormonal reactions, and other changes in physically detectable state are included within heterophenomenology. I thought that went without saying, but apparently these additional data are often conveniently overlooked by critics of heterophenomenology.

From the recorded verbal utterances, we get transcripts (e.g., in English or French, or whatever), from which in turn we devise interpretations of the subjects' speech acts, which we thus get to treat as (apparent) expressions of their beliefs, on all topics. Thus using the intentional stance (Dennett 1971, 1987), we construct therefrom the subject's heterophenomenological world. We move, that is, from *raw* data to *interpreted* data: a catalogue of the subjects' convictions, beliefs, attitudes, emotional reactions, ... (together with much detail regarding the circumstances in which these intentional states are situated), but then we adopt a special move, which distinguishes heterophenomenology from the normal interpersonal stance: the subjects' beliefs (etc.) *are all bracketed for neutrality*.

Why? Because of two failures of overlap, which we may label false positive and false negative. False positive: Some beliefs that subjects have about their own conscious states are provably false, and hence what needs explanation in these cases is the *etiology of the false belief*.

For instance, most people—naïve people—think their visual fields are roughly uniform in visual detail or grain all the way out to the periphery. Even sophisticated cognitive scientists can be startled when they discover just how poor their capacity is to identify a peripherally located object (such as a playing card held at arm's length). It certainly *seems* as if our visual consciousness is detailed all the way out all the time, but easy experiments show that it isn't. (Our color vision also seems to extend all the way out, but similar experiments show that it doesn't.) So the question posed by the heterophenomenologist is:

Why do people think their visual fields are detailed all the way out? not this question: How come, since people's visual fields are detailed all the way out, they can't identify things parafoveally?

False negative: Some *psychological things that happen in people* (to put it crudely but neutrally) are *unsuspected by those people*. People not only *volunteer* no information on these topics; when provoked to search, they *find* no information on these topics. But a forced choice guess, for instance, reveals that nevertheless, there is something psychological going on. This shows, for instance, that they *are* being influenced by the meaning of the masked word even though they are, as they put it, entirely unaware of any such word. (One might put this by saying that there is a lot of unconscious mental activity—but this is tendentious; to some, it might be held to beg the vexed question of whether people are briefly conscious of these

evanescent and elusive topics, but just hugely and almost instantaneously forgetful of them.)

Now faced with these failures of overlap—people who believe they are conscious of more than is in fact going on in them, and people who do not believe they are conscious of things that are in fact going on in them—heterophenomenology maintains a nice neutrality: it characterizes their beliefs, their heterophenomenological world, without passing judgment, and then investigates to see what could explain *the existence of those beliefs*. Often, indeed typically or normally, the existence of a belief is explained by confirming that it is a *true* belief provoked by the normal operation of the relevant sensory, perceptual, or introspective systems. Less often, beliefs can be seen to be true only under some arguable metaphorical interpretation—the subject claims to have manipulated a mental image, and we’ve found a quasi-imagistic process in his brain that can support that claim, if it is interpreted metaphorically. Less often still, the existence of beliefs is explainable by showing how they are illusory byproducts of the brain’s activities: it only *seems* to subjects that they are reliving an experience they’ve experienced before (*déjà vu*).

In this chapter we have developed a *neutral* method for investigating and describing phenomenology. It involves extracting and purifying *texts* from (apparently) speaking *subjects*, and using those texts to generate a theorist’s fiction, the subject’s *heterophenomenological world*. This fictional world is populated with all the images, events, sounds, smells, hunches, presentiments, and feelings that the subject (apparently) sincerely believes to exist in his or her (or its) stream of consciousness. Maximally extended, it is a neutral portrayal of exactly *what it is like to be* that subject—in the subject’s own terms, given the best interpretation we can muster.... People undoubtedly do believe that they have mental images, pains, perceptual experiences, and all the rest, and *these* facts—the facts about what people believe, and report when they express their beliefs—are phenomena any scientific theory of the mind must account for. (CE, p98)

Is this truly neutral, or does it bias our investigation of consciousness by stopping one step short? Shouldn’t our *data* include not just subject’s subjective *beliefs about their experiences*, but the *experiences themselves*? Levine, a first-string member of the B Team, insists “that conscious experiences themselves, not merely our verbal judgments about them, are the primary data to which a theory must answer.” (Levine 1994)

This is an appealing idea, but it is simply a mistake. First of all, remember that heterophenomenology gives you much more data than just a subject’s verbal judgments; every blush, hesitation, and frown, as well as all the covert, internal reactions and activities that can be detected, are included in our primary data. But what about this concern with leaving the “conscious experiences themselves” out of the primary data? Defenders of the first-person point of view are not entitled to this complaint against heterophenomenology, since *by their own lights*, they should prefer heterophenomenology’s treatment of the primary data to any other. Why? Because it does justice to both possible sources of non-overlap. On the one hand, if some of your conscious experiences occur unbeknownst to you (if they are experiences about which you have no beliefs, and hence can make no “verbal judgments”), then they are just as inaccessible to your first-person point of view as they

are to heterophenomenology. *Ex hypothesi*, you don't even suspect you have them—if you did, you could verbally express those suspicions. So heterophenomenology's list of primary data doesn't leave out any conscious experiences you know of, or even have any first-person inklings about. On the other hand, unless you claim not just reliability but outright infallibility, you should admit that some—just some—of your beliefs (or verbal judgments) about your conscious experiences might be wrong. In all such cases, however rare they are, what has to be explained by theory is not the conscious experience, but your belief in it (or your sincere verbal judgment, etc.). So heterophenomenology doesn't include any spurious "primary data" either, but plays it safe in a way you should approve.

Heterophenomenology is nothing but good old 3rd-person scientific method applied to the particular phenomena of human (and animal) consciousness. Scientists who were interested in taking the first-person point of view seriously figured out how to do just that, bringing the data of the first person into the fold of objective science. I didn't invent the method; I merely described it, and explained its rationale.

Alvin Goldman has recently challenged this claim. In "Science, Publicity and Consciousness" (1997), he says that heterophenomenology is not, as I claim, the standard method of consciousness research, since researchers "rely substantially on subjects' introspective beliefs about their conscious experience (or lack thereof)" (p. 532). In private correspondence (Feb 21, 2001) he has elaborated his claim thus:

The objection lodged in my paper to heterophenomenology is that what cognitive scientists *actually* do in this territory is not to practice agnosticism. Instead, they rely substantially on subjects' introspective beliefs (or reports). So my claim is that the heterophenomenological method is not an accurate description of what cognitive scientists (of consciousness) standardly do. Of course, you can say (and perhaps intended to say, but if so it wasn't entirely clear) that this is what scientists *should* do, not what they *do* do.

I certainly would play the role of reformer if it were necessary, but Goldman is simply mistaken; the adoption of agnosticism is so firmly built into practice these days that it goes without saying, which is perhaps why he missed it. Consider, for instance, the decades-long controversy about mental imagery, starring Shepard, Kosslyn, and Pylyshyn among many others. If agnosticism were not the tacit order of the day, Kosslyn would have never needed to do his well-known experiments to support subjects' claims that what they were doing (at least if described metaphorically) really was a process of image manipulation. (The issues are not settled yet, of course.) In psychophysics, the use of signal detection theory has been part of the canon since the 1960s, and it specifically commands researchers to control for the fact that the response criterion is under the subject's control although the subject is not himself or herself a reliable source on the topic. Or consider the voluminous research literature on illusions, both perceptual and cognitive, which standardly assumes that the data are what subjects *judge* to be the case, and *never* makes the mistake of "relying substantially on subjects' introspective beliefs." The diagnosis of Goldman's error is particularly clear here: of course experimenters on illusions rely on subjects' introspective beliefs (as expressed in their judgments)

about *how it seems to them*, but that *is* the agnosticism of heterophenomenology; to go beyond it would be, for instance, to assume that in size illusions *there really were visual images of different sizes somewhere in subjects' brains (or minds)*, which of course no researcher would dream of doing. Finally, consider such phenomena as *déjà vu*. Sober research on this topic has never made the mistake of abandoning agnosticism about subjects' claims to be reliving previous experiences. See, e.g., Bower and Clapper, in 1989, for instance, or any good textbook on methods in cognitive science for the details. (Goldman has responded to this paragraph in a series of emails to me, which I have included in an Appendix.)

A bounty of excellent heterophenomenological research has been done, is being done, on consciousness. See, e.g., the forthcoming special issue of *Cognition*, edited by Stanislas Dehaene, on the cognitive neuroscience of consciousness. It contains a wealth of recent experiments all conducted within the methodological strictures of heterophenomenology, whose resolutely 3rd-person treatment of belief attribution squares perfectly with standard scientific method: when we assess the attributions of belief relied upon by experimenters (in preparing and debriefing subjects, for instance) we use precisely the principles of the intentional stance to settle what it is reasonable to postulate regarding the subjects' beliefs and desires. Now Chalmers has objected (in the debate) that this "behavioristic" treatment of belief is itself question-begging against an alternative vision of belief in which, for instance, "having a phenomenological belief doesn't involve just a pattern of responses, but often requires having certain experiences." (personal correspondence, 2/19/01). On the contrary, heterophenomenology is neutral on just this score, for surely we mustn't *assume* that Chalmers is right that there *is* a special category of "phenomenological" beliefs—that there is a kind of belief that is off-limits to "zombies" but not to us conscious folks. Heterophenomenology allows us to proceed with our catalogue of a subject's beliefs leaving it open whether any or all of them are Chalmers-style phenomenological beliefs or mere zombie-beliefs. (More on this later.) In fact, heterophenomenology permits science to get on with the business of accounting for the patterns in all these subjective beliefs without stopping to settle this imponderable issue. And surely Chalmers must admit that the patterns in these beliefs are among the phenomena that any theory of consciousness must explain.

Let's look at a few cases of heterophenomenology in action (Please refer to the accompanying video "Ramachandran's Video" present at the link mentioned at the beginning of the article). Do you see the motion? You see apparent motion. Does the yellow blob really move? The blob *on the screen* doesn't move. Ah, but does the *subjective* yellow blob in your experience move? Does it *really* move, or do you just *judge* that it moves? Well, it sure seems to move! That is what you judge, right? Now perhaps there are differences in how you would word your judgments. And perhaps there are other differences. Perhaps some of you not only judge that it seems to move, but are made slightly dizzy or nauseated by the apparent motion. Perhaps some people get motion sickness from motion capture and others don't. Perhaps some of you don't even experience the apparent motion at all. Perhaps some of you can use such apparent motion just like real motion, to help

disambiguate shapes, for instance, and perhaps you can't. We can explore these variations in as much detail as you like, and can come back to you again and again with further inquiries, further tests, further suggested distinctions.

You are *not* authoritative about what is happening in you, but only about what *seems* to be happening in you, and we are giving you total, dictatorial authority over the account of how it seems to you, about *what it is like to be you*. And if you complain that some parts of how it seems to you are ineffable, we heterophenomenologists will grant that too. What better grounds could we have for believing that you are unable to describe something than that (1) you don't describe it, and (2) confess that you cannot? Of course you might be lying, but we'll give you the benefit of the doubt. (*CE*, p96-7)

Is there anything about your experience of this motion capture phenomenon that is not explorable by heterophenomenology? I'd like to know what. This is a fascinating and surprising phenomenon, *predicted* from the 3rd-person point of view, and eminently studiable via heterophenomenology. (Tom Nagel once claimed that 3rd-person science might provide us with brute correlations between subjective experiences and objective conditions in the brain, but could never *explain* those correlations, in the way that chemists can explain the correlation between the liquidity of water and its molecular structure. I asked him if he considered the capacity of industrial chemists to predict the molar properties of novel artificial polymers in advance of creating them as the epitome of such explanatory correlation, and he agreed that it was. Ramachandran and Gregory predicted this motion capture phenomenon, an entirely novel and artificial subjective experience, on the basis of their knowledge of how the brain processes vision.)

See next Rensink's change blindness (Please refer to the accompanying video "Rensink's Video" present at the link mentioned at the beginning of the article). (By the way, this is an effect I predicted in *CE*, much to the disbelief of many readers.)

Were your qualia changing before you noticed the flashing white cupboard door? You saw each picture several dozen times, and *eventually* you saw a change that was "swift and enormous" (Dennett 1999; Palmer 1999) but that swift, enormous change was going on for a dozen times and more before you noticed it. Does it count as a change in color qualia?

The possible answers:

A. Yes.

B. No.

C. I don't know

1. because I now realize I never knew quite what I meant by "qualia" all along.
2. because although I know just what I have always meant by "qualia", I have no first-person access to my own qualia in this case.

a. and 3rd-person science can't get access to qualia either!

Let's start with option C first. Many people discover, when they confront this case, that since they never imagined such a phenomenon was possible, they never considered how their use of the term "qualia" should describe it. They discover a

heretofore unimagined flaw in their concept of qualia—rather like the flaw that physicists discovered in their concept of *weight* when they first distinguished weight from mass. The philosophers' concept of qualia is a mess. Philosophers don't even agree on how to apply it in dramatic cases like this. I hate to be an old I-told-you-so but I told you so ("Quining Qualia"). This should be at least mildly embarrassing to our field, since so many scientists have recently been persuaded by philosophers that they should take qualia seriously—only to discover that philosophers don't come close to agreeing among themselves about *when* qualia—*whatever* they are—are present. (I have noticed that many scientists who think they are newfound friends of qualia turn out to use the term in ways no self-respecting qualophile will countenance.)

But although some philosophers may now concede that they aren't so sure what they meant by "qualia" all along, others are very sure what concept of qualia they've been using all along, so let's consider what they say. Some of them, I have learned, have *no problem* with the idea that their very own qualia could change radically without their noticing. *They* mean by "qualia" something to which their 1st-person access is variable and problematic. If you are one of those, then heterophenomenology is your preferred method, since it, unlike the first-person point of view, can actually study the question of whether qualia change in this situation. It is going to be a matter of some delicacy, however, how to *decide* which brain events count for what. In this phenomenon of change blindness for color changes, for instance, we know that the color-sensitive cones in the relevant region of your retina were flashing back and forth, in perfect synchrony with the white/brown quadrangle, and presumably (we should check) other, later areas of your color vision system were also shifting in time with the external color shift. But if we keep looking, we will also presumably find yet other areas of the visual system that only come into synchrony after you've noticed. (such effects have been found in similar fMRI studies, e.g. O'Craven et al. 1997).

The hard part will be deciding (on what grounds?) which features of which states to declare to be qualia and why. I am not saying there can't be grounds for this. I can readily imagine there being *good* grounds, but if so, then those will be grounds for adopting/endorsing a 3rd-person concept of qualia (cf. the discussion of Chase and Sanborn in Dennett 1988, or the beer-drinkers in CE 395-6). The price you have to pay for obtaining the *support* of 3rd-person science for your conviction about how it is/was with you is straightforward: you have to grant that *what you mean* by how it is/was with you is something that 3rd-person science could either support or show to be mistaken. Once we adopt any such concept of qualia, for instance, we will be in a position to answer the question of whether color qualia shift during change blindness. And if some subjects in our apparatus tell us that their qualia do shift, while our brain-scanner data shows clearly that they don't, we'll treat these subjects as simply wrong *about their own qualia*, and we'll explain why and how they come to have this false belief.

Some people find this prospect inconceivable. For just this reason, some people may want to settle for option B: No, my qualia don't change—*couldn't* change—until I notice the change. This decision *guarantees* that qualia, tied thus to noticing, are securely within the heterophenomenological worlds of subjects, are indeed constitutive features of their heterophenomenological worlds. On option B, what subjects can say about their qualia fixes the data.¹

By a process of elimination, that leaves option A, YES, to consider. If you think your qualia did change (though you didn't notice it at the time) *why* do you think this? Is this a theory of yours? If so, it needs evaluation like any other theory. If not, did it just come to you? A gut intuition? Either way, your conviction is a prime candidate for heterophenomenological diagnosis: what has to be explained is how you came to have this belief. The last thing we want to do is to treat your claim as incorrigible. Right?

Here is the dilemma for the B Team, and Captain Chalmers. If you eschew incorrigibility claims, and especially if you acknowledge the competence of 3rd-person science to answer questions that can't be answered from the 1st-person point of view, your position collapses into heterophenomenology. The only remaining alternative, C(2a), is unattractive for a different reason. You can protect qualia from heterophenomenological appropriation, but only at the cost of declaring them outside science altogether. If qualia are so shy they are not even accessible from the 1st-person point of view, then no 1st-person science of qualia is possible either.

I will not contest the existence of first-person *facts* that are unstudiable by heterophenomenology and other 3rd-person approaches. As Steve White has reminded me, these would be like the humdrum "inert historical facts" I have spoken of elsewhere—like the fact that some of the gold in my teeth once belonged to Julius Caesar, or the fact that none of it did. One of those is a fact, and I daresay no possible extension of science will ever be able to say which is the truth. But if 1st-person facts are like inert historical facts, they are no challenge to the claim that heterophenomenology is the maximally inclusive science of consciousness, because they are unknowable even to the 1st-person they are about!

¹Consider Option B for the simpler case raised earlier. Do you want to cling to a concept of visual consciousness according to which your conviction that your visual consciousness is detailed all the way out is *not* contradicted by the discovery that you cannot identify large objects in the peripheral field? You *could* hang tough: "Oh, all that you've shown is that we're not very good at identifying objects in our peripheral vision; *that* doesn't show that peripheral consciousness isn't as detailed as it seems to be! All you've shown is that a *mere behavioral capacity* that one might mistakenly have thought to coincide with consciousness doesn't, in fact, show us anything about consciousness!" Yes, if you are careful to *define* consciousness so that nothing "behavioral" can bear on it, you get to declare that consciousness transcends "behaviorism" without fear of contradiction. See "Are we Explaining Consciousness Yet?" for a more detailed account of this occasionally popular but hopeless move.

David Chalmers as a Heterophenomenological Subject

Of course it *still* seems to many people that heterophenomenology must be leaving something out. That's the ubiquitous Zombic Hunch. How does the A team respond to this? Very straightforwardly: by *including* the Zombic Hunch among the heartfelt convictions any good theory of consciousness must explain. One of the things that it falls to a theory of consciousness to explain is *why some people are visited by the Zombic Hunch*. Chalmers is one such, so let's look more closely at the speech acts Chalmers has offered as a subject of heterophenomenological investigation.

Here is Chalmers' definition of a zombie (his zombie twin):

Molecule for molecule identical to me, and identical in all the low-level properties postulated by a completed physics, but he lacks conscious experience entirely... he is embedded in an identical environment. He will certainly be identical to me *functionally*; he will be processing the same sort of information, reacting in a similar way to inputs, with his internal configurations being modified appropriately and with indistinguishable behavior resulting.... he will be awake, able to report the contents of his internal states, able to focus attention in various places and so on. It is just that none of this functioning will be accompanied by any real conscious experience. There will be no phenomenal feel. There is nothing it is like to be a Zombie... 1996, p95

Notice that Chalmers allows that zombies have internal states with contents, which the zombie can report (sincerely, one presumes, believing them to be the truth); these internal states have contents, but not conscious contents, only pseudo-conscious contents. The Zombic Hunch, then, is Chalmers' conviction that he has just described a real problem. It *seems to him* that there is a problem of how to explain the difference between him and his zombie twin.

The justification for my belief that I am conscious lies not just in my cognitive mechanisms but also in *my direct evidence* [emphasis added]; the zombie lacks that evidence, so his mistake does not threaten the grounds for our beliefs. (One can also note that the zombie doesn't have the same beliefs as us, because of the role that experience plays in constituting the contents of those beliefs.) (Reply to Searle)

This speech act is curious, and when we set out to interpret it, we have to cast about for a charitable interpretation. How does Chalmers' justification *lie in* his "direct evidence"? Although he says the zombie *lacks* that evidence, nevertheless the zombie *believes* he has the evidence, just as Chalmers does. Chalmers and his zombie twin are heterophenomenological twins: when we interpret all the data we have, we end up attributing to them exactly the same heterophenomenological worlds. Chalmers fervently believes he himself is not a zombie. The zombie fervently believes he himself is not a zombie. Chalmers *believes* he gets his justification from his "direct evidence" of his consciousness. So does the zombie, of course.

The zombie has the conviction that he has direct evidence of his own consciousness, and that this direct evidence is his justification for his belief that he is conscious. Chalmers must maintain that the zombie's conviction is false. He says that the zombie *doesn't* have the same beliefs as us "because of the role that

experience plays in constituting the contents of those beliefs,” but I don’t see how this can be so. Experience (in the special sense Chalmers has tried to introduce) plays no role in constituting the contents of those beliefs, since *exhypothesi*, if experience (in this sense) were eliminated—if Chalmers were to be suddenly zombified—he would go right on saying what he says, insisting on what he now insists on, and so forth.² Even if his “phenomenological beliefs” suddenly ceased to be phenomenological beliefs, he would be none the wiser. It would not *seem to him* that his beliefs were no longer phenomenological.

But wait, I am forgetting my own method and arguing with a subject! As a good heterophenomenologist, I must grant Chalmers full license to his deeply held, sincerely expressed convictions and the heterophenomenological world they constitute. And then I must undertake the task of explaining the etiology of his beliefs. Perhaps Chalmers’ beliefs about his experiences will turn out to be true, though how that prospect could emerge eludes me at this time. But I will remain neutral. Certainly we shouldn’t give them incorrigible status. (He’s not the Pope.) The fact that some subjects have the Zombic Hunch shouldn’t be considered grounds for revolutionizing the science of consciousness.³

Where’s the Program?

That leaves the B Team in a bit of a predicament. Chalmers would like to fulfil the Philosopher’s Dream:

To prove a priori, from one’s ivory tower, a metaphysical fact that forces a revolution in the sciences.

It is not an impossible dream. (That is, it is not logically impossible.) Einstein’s great insight into relativity comes tantalizingly close to having been a purely philosophical argument, something a philosopher might have come up with just from first principles. And Patrick Matthew could claim with some justice in 1860 to have scooped Darwin’s theory of natural selection in 1831 by an act of pure reason:

it was by a general glance at the scheme of Nature that I estimated this select production of species as an a priori recognizable fact—an axiom, requiring only to be pointed out to be admitted by unprejudiced minds of sufficient grasp. [see DDI, p49]

²“I simply say that invoking consciousness is not necessary to explain actions; there will always be a physical explanation that does not invoke or imply consciousness. A better phrase would have been ‘explanatorily superfluous’, rather than ‘explanatorily irrelevant.’” (Chalmers’ second reply to Searle, on his website).

³Chalmers seems to think that conducting surveys of his audiences, to see what proportion can be got to declare their allegiance to the Zombic Hunch, yields important data. Similar data-gathering would establish the falsehood of neo-Darwinian theory and the existence of an afterlife.

The Zombic Hunch is accompanied by arguments designed to show that it is *logically possible* (however physically impossible) for there to be a zombie. This logical possibility is declared by Chalmers to have momentous implications for the scientific study of consciousness, but as a candidate for the Philosopher's Dream it has one failing not shared with either Einstein's or Matthew's great ideas: *it pre-scribes no research program*. Suppose you are convinced that Chalmers is right. Now what? What experiments would you do (or do differently) that you are not already doing? What models would you discard or revise, and what would you replace them with? And why?

Chalmers has recently addressed this very issue in a talk entitled "First-Person Methods in the Science of Consciousness" (*Consciousness Bulletin*, Fall 1999, and on Chalmers' website), but I hunt through that essay in vain for any examples of research that are somehow off limits to, or that transcend, heterophenomenology:

I take it for granted that there are first-person data. It's a manifest fact about our minds that there is something it is like to be us - that we have subjective experiences - and that these subjective experiences are quite different at different times. Our direct knowledge of subjective experiences stems from our first-person access to them. And *subjective experiences are arguably the central data that we want a science of consciousness to explain*. [emphases added] I also take it that the first-person data can't be expressed wholly in terms of third-person data about brain processes and the like. There may be a deep connection between the two - a correlation or even an identity - but if there is, the connection will emerge through a lot of investigation, and *can't be stipulated at the beginning of the day* [emphasis added]. That's to say, no purely third-person description of brain processes and behavior will express precisely the data we want to explain, though they may play a central role in the explanation. So as data, the first-person data are irreducible to third-person data.

Notice how this passage blurs the distinctions of heterophenomenology. "Arguably?" I have argued, to the contrary, that *subjects' beliefs* about their subjective experiences are the central data. I've reviewed these arguments here today. So, is Chalmers rejecting my arguments? If so, what is wrong with them? I agree with him that a correlation or identity—or indeed, the veracity of a subject's beliefs—"can't be stipulated at the beginning of the day." That is the neutrality of heterophenomenology. It is Chalmers who is holding out for an opening stipulation in his insistence that the Zombic Hunch be granted privileged status. As he says, he "takes it for granted that there are first-person data." I don't. Not in Chalmers' charged sense of that term. I don't stipulate at the beginning of the day that our subjective *beliefs* about our first-person experiences are "phenomenological" beliefs in a sense that requires them somehow to depend on (but not causally depend on) experiences that zombies don't have! I just stipulate that the contents of those beliefs exhaustively constitute each person's (or zombie's) subjectivity.

In his paper on first-person methods, Chalmers sees some of the problems confronting a science of consciousness:

When it comes to first-person methodologies, there are well-known obstacles: the lack of incorrigible access to our experience; the idea that introspecting an experience changes the experience; the impossibility of accessing all of our experience at once, and the consequent possibility of "grand illusions"; and more. I don't have much that's new to say about these.

I think that could end up posing principled limitations, but none provide in-principle barriers to at least initial development of methods for investigating the first-person data in clear cases.

Right. Heterophenomenology has already made the obligatory moves, so he doesn't need to have anything new to say about these. I don't see anything in this beyond heterophenomenology. Do you? Chalmers goes on:

When it comes to first-person formalisms, there may be even greater obstacles: can the content of experience be wholly captured in language, or in any other formalism, at all? Many have argued that at least some experiences are "ineffable". And if one has not had a given experience, can any description be meaningful to one? Here again, I think at least some progress ought to be possible. We ought at least to be able to develop formalisms for capturing the structure of experience: similarities and differences between experiences of related sorts, for examples, and the detailed structure of something like a visual field.

What a good idea: we can let subjects speak for themselves, in the first-person, and then we can take what they say seriously and try to systematize it, to capture the structure of their experience! And we could call it heterophenomenology.

If Chalmers speaks of anything in this paper (remember, it is entitled "*First-person Methods in the Science of Consciousness*") that is actually distinct from 3rd-person heterophenomenology, I don't see what it is. Both there and in his contribution to our debate he mentioned various ongoing research topics that strike him as playing an important role in his anticipated 1st-person science of consciousness—work on blindsight and masking and inattentional blindness, for instance—but all this has long ago been fit snugly into 3rd-person science.

In the debate, Chalmers asserted that a heterophenomenological methodology would not be able to motivate questions about what was going on in consciousness in these phenomena. That is utterly false, of course; these very phenomena were, after all, parade cases for heterophenomenology in *Consciousness Explained*. It is important to remember that the burden of heterophenomenology is to explain, in the end, every pattern discoverable in the heterophenomenological worlds of subjects; it is precisely these patterns that make these phenomena striking, so heterophenomenology is clearly the best methodology for investigating these phenomena and testing theories of them.

I find it ironic that while Chalmers has made something of a mission of trying to convince scientists that they must abandon 3rd-person science for 1st-person science, when asked to recommend some avenues to explore, he falls back on the very work that I showcased in my account of how to study human consciousness empirically from the 3rd-person point of view. Moreover, it is telling that none of the work on consciousness that he has mentioned favorably addresses his so-called Hard Problem in any fashion; it is all concerned, quite appropriately, with what he insists on calling the easy problems. First-person science of consciousness is a discipline with no methods, no data, no results, no future, no promise. It will remain a fantasy.

Appendix: Goldman on Heterophenomenology

Alvin Goldman, responding to the paragraph above about Goldman 1997 (see page 5), entered into an email debate with me, lightly edited by me to avoid repetition and remove material not germane to the topics:

Goldman: First, a brief substantive reply to your points [see above, p5]. When cognitive scientists rely on subjects' reports about visual illusions, I take them to be relying on the veracity of the Ss' judgments (beliefs) about how the stimuli *look* (etc.). That is, after all, what the Ss presumably say, or can be interpreted as saying: "It looks as if such-and-such". And the cognitive scientist takes that to be true, i.e., that it does *look* that way to the S (roughly at the time of report). Similarly, the cognitive scientist obviously does not conclude that Ss who report a *déjà vu* experience really did have the same type of experience in his/her past. That could not be ascertained by the subject by introspection, which is restricted to present events. So even if the S's *déjà vu* report implies that he/she believes that a certain event or experience occurred in the past (I am not sure it does imply this), the cognitive scientist does not rely on the accuracy of this belief. However, the cognitive scientist (also) takes the S to report, and to believe, that he/she is currently having a "seems-like-this-happened-to-me-in-the past" experience. And the cognitive scientist *does* trust the S's report of *that*. In other words, the scientist concludes that the S does have (roughly at the time of report) an experience of the type "seems-like-this- happened- to-me-in-the-past".

In the context of the treatment of illusions, I do have to talk more about "looks" or "seems". As your discussion below indicates (and you have frequently said in print), you take "seems" only to express something about a S's *belief*. There is no further fact about S (beyond a belief fact) that is expressed by "It seems to S to be F". I, on the contrary, think that a seeming-state is not merely a belief, but a visual state, an auditory state, or other "perceptual-phenomenal" state. You think (see your discussion [above, p. 5]) that such an alleged state would have to involve "images" of certain sizes in the brain. But that is a totally unwarranted interpretation. Undergoing a perceptual-seeming episode need involve nothing like "sense-data" of the sort you conjure up. Cognitive scientists do not have to commit themselves to anything like that when they say that a S really is undergoing a certain type of perceptual-seeming episode (when the S reports that he is).

DENNETT REPLY interjected: EXACTLY! They don't have to commit themselves to anything like that. They can remain neutral. My example of mental images in the brain was just a for instance. My point was that to go beyond heterophenomenological agnosticism, they'd have to suppose something was implied by their S's judgments (beyond the bare fact that these were their judgments, which is what heterophenomenology happily allows). Now it MAY be that your point about "perceptual-phenomenal" states that go beyond "mere" belief—states will someday be supported somehow. But in the meantime, cognitive science proceeds along merrily, leaving itself strictly neutral about that. And in at least some instances (for instance, sudden hunches of *déjà vu*) the claim that there is

anything “perceptual-phenomenal” about the presentiment over and above the inclination so to judge seems particularly dubious. (Ask yourself what *deja vu* would be like if it didn’t have any so-called “phenomenal” stuffing. Isn’t that in fact what it is like?) But in any case, cognitive science can and should (and does!) remain strictly neutral about such questions of phenomenality until the case is clearly made. My point for years is that it never has been made, so it counts, so far, as just a set of tempting hunches (versions of the Zombic Hunch) that cognitive science should also be agnostic about. And I know of no research in cognitive science that has violated that neutrality except by accident.

You say that my view is that “There is no further fact about S (beyond a belief fact) that is expressed by “It seems to S to be F”.” Not quite. I have challenged people to show any way in which there is such a further fact. My view is that it has not been shown that there is any such further fact (beyond the obvious other “behavioral” facts that accompany such belief facts, typically) and in the meanwhile cognitive science can proceed quite happily in strict neutrality about this. In fact, it had better be neutral about this from the outset, so that it can actually have a standpoint from which it might confirm (or disconfirm) your belief.

GOLDMAN, continued: So what is going on when people have a perceptual-seeming episode (whether during actual perception or during imagery)? You point out, in connection with the Shepard, Kosslyn, and Pylyshyn debate, that cogscientists would never rely on Ss’ reports to try to settle that. I reply: That is certainly true! But I would never claim, and have never claimed, that scientists rely on all aspects or all details of what their Ss might say. This is explicitly addressed in my “Science, Publicity, and Consc” (SPC) paper on p. 544, the last page of the article. “Everyone nowadays agrees that introspection is an unreliable method for answering questions about the micro-structure of cognition. For example, nobody expects subjects to report reliably whether their thinking involves the manipulation of sentences in a language of thought. But this leaves many other types of questions about which introspection could be reliable”. This point is made again in my JCS paper, “Can Science Know When You’re Conscious?” (*Journal of Consciousness Studies* 2000) Here is what I say on p. 4 of that article: “Cognitive psychologists and neuropsychologists would not rely, after all, on their subjects’ reports about all psychological states or processes. When it comes to the nonconscious sphere of mental processing--the great bulk of what transpires in the mind-brain--scientists would not dream of asking subjects for their opinions. Moreover, if subjects were to offer their views about what happens (at the micro-level) when they parse a sentence or retrieve an episode from memory or reach for a cup, scientists would give no special credence to these views.”

So I fully acknowledge that for a wide range of questions, scientists do not allow their Ss’ introspections to settle anything. (Of course, usually the Ss have nothing to offer about what happens at the micro-level.) But for another large range of questions, I claim, they *do* trust their Ss’ introspections. (A more precise specification of which questions are which I have not yet tried to give. Nor do I know of anybody who has tried to be precise on this matter.)

DENNETT REPLY interjected: Try me. I have. I have pointed out that they trust their S's introspective reports to be fine accounts of how it seems to them—with regard to every phenomenon in all modalities. And that this exhausts the utility of their S's protocols, which they can then investigate by devising experiments that probe the underlying mechanisms. They “trust” their Ss only after they've discovered, independently, that their statements, interpreted as assertions about objective, 3rd-person—accessible processes going on in their brains, are reliable. In other words, they only “rely on” S's statements when they have confirmed that they can be usefully interpreted as ordinary reliable reports of objective properties.

Ask yourself how things would stand if Pylyshyn's most extreme line of mental imagery had turned out to be true (more than the barest logical possibility, I'm sure you would agree—he was not insane or incoherent to put forward his criticisms). In that case, I submit, everyone would agree that the agnosticism of heterophenomenology had paid off big time; people turn out to be deeply wrong about what they are doing. They think they are manipulating mental images with such and such features when in fact all that is happening in them is X. The fact that it sure seems to them that they are manipulating mental images would then have to be explained by showing how they are caused to have these heartfelt convictions in spite of their now demonstrated falsehood. Now if that was never a possible outcome of the research, what on earth could Pylyshyn have thought he was doing? For that matter, what could Kosslyn have thought he was doing?

GOLDMAN continued: In any case, the main point is that I of course agree that not everything a subject might say, in an introspective spirit, would be regarded as scientific gospel. So some of the things you say about conflicts between scientific practice and my reconstruction of it don't work.

DENNETT REPLY: I didn't say you did claim that they held that everything is regarded as scientific gospel. I said that you claimed that cognitive scientists aren't systematically agnostic. But they are, systematically, so systematically that they don't even both mention it, in all the cases I cite in this passage where I discuss your claim.

The proper way to criticize my view is to develop an independent case for “real seeming.” A number of people have tried. Nobody has yet succeeded. See, e.g., the essays in the *Phil Topics* issue of 1994, and my response, “Get Real”. But beyond establishing this as a philosophical point, there is the obligation to show that cognitive science has been (or should be) honoring it. When you can show experiments that get misinterpreted, or can't be analyzed, or would never be dreamt up, by people committed to heterophenomenology, then you can claim that I am mistaken in claiming that heterophenomenology both is, and should be, agnostic.

GOLDMAN, next response: I agree that *one* of the key issues is whether there is anything more to visual seeming (e.g.) than belief. At the risk of repeating what others have said (possibly ad nauseum, from your point of view), this just seems like the obvious, straightforward interpretation of what goes on in, e.g., the blindsight patient. The patient doesn't tell his physician that he doesn't *believe* that there are any objects of such-and-such type in the vicinity (in the area of his scotoma). He says that he doesn't see anything in that vicinity [*expressing*, not *reporting* his

belief that he doesn't see anything in that vicinity; see CE, pp. 305–6–DCD]. We might even arrange for there to be a case where he does have beliefs about the target properties—as a result of somebody else *telling* him about such properties. But he'll still say that he doesn't *see* anything there. And the standard, default, entry-level reaction of the cognitive scientist is to trust that report, to conclude that S really doesn't see anything there. Of course, the scientist might be a little more cautious, since, among other things, the S might be confabulating, or neglecting. But the reason blindsight is an interesting and challenging phenomenon, a phenomenon related to *vision*, is because it's an absence of *seeing*. How do we know about this absence? From the S. From the subjects' reports. So we are basing our conclusions on a trust of the subjects' reports.

DENNETT REPLY interjected: Not so. Anticipating this sort of response in my own discussion of blindsight in CE, I pointed out the problem of trust. See p. 326, where I show why “the phenomena of blindsight appear only when we treat subjects from the standpoint of heterophenomenology” and particularly point to how the phenomenon would evaporate if we concluded that subjects were malingering, or suffering from hysterical blindness. Heterophenomenology is tailor made for dealing with blindsight.

Again, in the *deja vu* case, it doesn't capture the phenomenon well to describe it as a *belief* that one experienced a similar thing in the past. Rather, it's a phenomenon in which it *feels* like one experienced such a thing in the past; or one has a seeming memory of such a thing. One might not *believe* that it happened at all, but one still *feels* as if it did. Again it's a reliance on the S's report of this phenomenon that makes the observer think that the S has really undergone this phenomenon at the time of report.

DENNETT REPLY interjected: To “feel as if it did” is to be strongly tempted to judge that it did. Of course the temptation can be overridden once one is no longer naive. And what is the feeling of temptation? Just noticing that one is so tempted to judge!

GOLDMAN next reply: I realize that a “doxological” (or representational) reductionist like yourself will want to reduce feeling states to dispositions-to-believe. A resistor like myself need not deny, of course, that feeling states do have a tendency to produce beliefs. The question is whether there are “categorical” features of feeling states in virtue of which they have that tendency, or whether they are just pure doxological tendency and nothing else. I find the former view more compelling, and don't think that representational reductionism will work across the board. But this is another big issue (admittedly one that is intimately tied to the issue at hand).

DENNETT REPLY: Fine. And isn't it nice that heterophenomenology can proceed with all of its research agenda without our having to settle anything about this “big issue” first! If you're right, the “categorical” features will eventually be confirmed to be important by some as yet unimagined test. (Or if, as I gather your colleague David Chalmers holds, no empirical or “behavioral” test could shed any light on this important but elusive sort of feature, I guess it will have to be some philosophical argument alone that settles the issue. Seems unlikely in the extreme to

me.) In the meantime, a 3rd-person science of consciousness can proceed apace. That's what is so good about its neutrality.

GOLDMAN: One last question about "neutrality". In your discussion of blindsight, do you agree that scientists give *prima facie* credence to a subject who claims to have no sight in a certain area? You stress that they do not uncritically trust these subjects. They want to check to see if there is neurological damage, and they want to rule out the possibility of "hysterical blindness". But don't they give some *prima facie* credence to the subject's report? Or do you deny this? If you agree that they do this, the question arises as to whether this is "neutrality", or agnosticism. I think not. Most epistemologists would agree that all of our sources of belief or justification are subject to correction from other sources. We don't trust vision uncritically, or memory, etc. But to say this is not to say that we are "agnostic" toward vision or memory. By giving *prima facie* credibility to each of these sources, we are doing the most that we ever do to any one source (or any one deliverance of a particular source). I would argue that the same holds here. Although the scientist does not uncritically trust a S's introspection (and there's an additional factor here—the S's report might not stem from introspection at all), he does give it *prima facie* trust. And that is very far from agnosticism. So if heterophenomenology ascribes true agnosticism to scientists, as you claim it does, then it doesn't get matters right.

DENNETT REPLY: As I try to make clear in CE, in the section entitled "The Discreet Charm of the Anthropologist," (pp. 82–3, on "Feenoman") heterophenomenology is NOT the NORMAL interpersonal relationship with which we treat others' beliefs—with its presumption of truth (marked by the willingness of the interlocutor to argue against it, to present any evidence believed to run counter, etc.). That is also true of anthropologists' relationships with their subjects when investigating such things as their religion. Actually, it extends quite far—when the native informants are telling the anthropologists about, say, their knowledge of the healing powers of the local plants, the anthropologists' first concern is to get the lore, true or false—something to be investigated further later. Ditto for heterophenomenology: get the lore, as neutrally and sympathetically as possible. That is a kind of agnosticism, differing in the ways I detail on pp. 82–3 from the normal interpersonal stance, but it is the normal researcher/subject relationship when studying consciousness with the help of S's protocols. If it doesn't fit your (or a dictionary's, or the majority of epistemologists') definition of agnosticism perfectly, I have at least made clear just what kind of agnosticism it is, and why it is the way it is.

As for blindsight, do the researchers give some *prima facie* credence to the reports? Of course—otherwise they wouldn't even consider investigating them. As I say, their attitude is to take what subjects say as seriously as possible—a policy that is entirely consistent with a kind of agnosticism, of course. The old introspectionism failed precisely because it attempted, unwisely, to give subjects more authority than they can handle; as the years rolled on, more cautious and savvy researchers developed the methodology I have dubbed heterophenomenology. They crafted a maximally objective, controlled way to turn verbal reports (and interpreted

button-pushes, etc., etc.) into legitimate data for science. All I have done is to get persnickety about the rationale of this entirely uncontroversial and ubiquitous methodology, and point out how and why it is what it is—and then I've given it an unwieldy name. So when, in my forthcoming Cognition essay, in the special issue on the cognitive neuroscience of consciousness, I point out that the hundreds of experiments discussed in the various pieces in that issue all conform to heterophenomenology, the editors and referees nod in agreement. Of course. It's just science, after all. And it does study consciousness. Obviously—unless you believe that the “easy” problems of consciousness are not about consciousness at all.

Now I have challenged David Chalmers to name a single experiment (in good repute) which in any way violates or transcends the heterophenomenological method. So far, he has not responded to my challenge. My challenge to you is somewhat different: to show that I misdescribe the standard methodology of cognitive science when I say it adopts the neutrality of heterophenomenology.

References

- G.H. Bower, J.P. Clapper, Experimental methods in cognitive science, in *Foundations of Cognitive Science*, ed. by M. Posner (MIT Press, 1989)
- Chalmers, *The Conscious Mind* (1996)
- Dennett, 'Quining Qualia' in *Consciousness in Contemporary Science*, ed. by A.J. Marcel, E.E. Bisiach (CUP, 1988)
- Dennett, Intrinsic changes in experience: swift and enormous commentary on palmer. *BBS* **22**(6), December 1999
- Dennett, The zombic hunch: extinction of an intuition?, in *Philosophy at the New Millenium*, ed. by A. O'Hear, vol. 48 (Cambridge Univ. Press, Royal Institute of Philosophy Supplement, 2001), pp. 27–43
- A. Goldman, *Philos. Sci.* **64**, 525–545 (1997)
- A. Goldman, Can science know when you're conscious? *J. Conscious. Stud.* (2000)
- K.M. O'Craven, B.R. Rosen, K.K. Kwong, A. Treisman, R.L. Savoy, Voluntary attention modulates fMRI activity in human MT/MST. *Neuron*, **18**, XXX (1997)
- S. Palmer, *Behav. Brain Sci.* **22**(6), December 1999

Shyam Wuppuluri · Francisco Antonio Doria
Editors

The Map and the Territory

Exploring the Foundations of Science,
Thought and Reality

Foreword by Sir Roger Penrose and
Afterword by Dagfinn Føllesdal



Springer

Editors

Shyam Wuppuluri
R. N. Podar School
Mumbai
India

Francisco Antonio Doria
Advanced Studies Research Group
Universidade Federal do Rio de Janeiro
Rio de Janeiro
Brazil

Additional material to this book can be downloaded from <http://extras.springer.com>.

ISSN 1612-3018

ISSN 2197-6619 (electronic)

THE FRONTIERS COLLECTION

ISBN 978-3-319-72477-5

ISBN 978-3-319-72478-2 (eBook)

<https://doi.org/10.1007/978-3-319-72478-2>

Library of Congress Control Number: 2017961750

© Springer International Publishing AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland