Gualtiero Piccinini

How to Improve on Heterophenomenology

The Self-Measurement Methodology of First-Person Data[1]

Abstract: *Heterophenomenology is a third-person methodology proposed by Daniel Dennett for using first-person reports as scientific evidence. I argue that heterophenomenology can be improved by making six changes: (i) setting aside consciousness, (ii) including other sources of first-person data besides first-person reports, (iii) abandoning agnosticism as to the truth value of the reports in favor of the most plausible assumptions we can make about what can be learned from the data, (iv) interpreting first-person reports (and other first-person behaviours) directly in terms of target mental states rather than in terms of beliefs about them, (v) dropping any residual commitment to incorrigibility of first-person reports, and (vi) recognizing that third-person methodology does have positive effects on scientific practices. When these changes are made, heterophenomenology turns into the self-measurement methodology of first-person data that I have defended in previous papers.*

Correspondence:
Gualtiero Piccinini, Philosophy Department, University of Missouri, 599 Lucas Hall (MC 73), 1 University Blvd., St. Louis, MO 63121-4400, USA.
Email: piccininig@umsl.edu

[1] Special thanks to David Chalmers, Daniel Dennett, and Marcin Milkowski for prompting me to be more explicit on the contrast between Dennett's heterophenomenology and my view on the methodology of first-person data. Thanks to David Chalmers, Pete Mandik, Marcin Milkowski, Anna-Mari Rusanen, and Eric Schwitzgebel for discussions on heterophenomenology and related topics; to Justin Sytsma, Marcin Milkowski, and Eric Schwitzgebel for helpful comments on a previous draft; and to Jim Virtel for editorial assistance.

1. Reinventing Heterophenomenology?

In a previous paper (Piccinini, 2003a), I argued that when psychologists and neuroscientists rely on first-person reports as sources of data, they follow — or at any rate they *should* follow — public (i.e., third-person) methods. By 'public method' I mean, roughly, a method that different investigators can follow to the same effect (Piccinini 2003b). All sciences are based on public methods. Therefore, it is incorrect to maintain, as some have (e.g., Chalmers, 2004; Gertler, 2009; Goldman, 1997; Varela and Shear, 1999), that reliance on first-person reports as scientific evidence constitutes a 'first-person' method, different in kind from public scientific methods.

The debate between first-person and third-person methodologies may also be framed in terms of *first-person data*, such as data derived from first-person reports. *First-person methodologists* maintain that first-person data are *private*, that is, directly accessible only to the subject issuing the reports. Since first-person data are private, they are different in kind from ordinary (public) scientific data. Some first-person methodologists define first-person data so that they are private *by definition*. I am using the term in a broader, more neutral sense. As I use the term, first-person data are a class of data associated with first-person reports (or other 'first-person behaviours'; more on this below). The question of whether first-person data in the present sense are private is not settled by definition — it is a further, substantive question. First-person methodologists answer in the positive. By contrast, *third-person methodologists* (myself included) argue that first-person data are *public* — that is, directly accessible to any competent external observer.

So far, the view I defend is the same as Daniel Dennett's. Dennett has been a critic of first-person methodologies for decades (Dennett, 1982; 1991) and has proposed an alternative called 'heterophenomenology'. Heterophenomenologists record first-person reports as raw data and then process the raw data to obtain final data about a subject's mind. Heterophenomenological data are no less public than any other scientific data, and the heterophenomenological science of mind is as much a third-person science as *any* other science. Heterophenomenology has become the standard-bearer for third-person methodologies — so much so that it is tempting to identify any third-person methodology of first-person data, such as the one I defend, with heterophenomenology. In fact, Dennett himself has charged me (Piccinini, 2003a) with an 'unwitting re-invention of heterophenomenology' (Dennett, 2007, p. 248, fn. 1).

I plead guilty, up to a point. The methodology I proposed and later articulated further (Piccinini 2009) overlaps with heterophenomenology, but it is not the same. Heterophenomenology is on the right track — first and foremost, in its commitment to third-person methods and the fallibility of first-person reports. Dennett's writings on heterophenomenology contain many additional insights, some of which I unwittingly reinvented in my 2003a. But heterophenomenology — as Dennett presents it — can be improved. In this paper, I will propose some friendly amendments to heterophenomenology, leading towards a more adequate methodology.

I will object to and propose replacements for each of the following Dennettian assumptions:

(1) Heterophenomenology is a methodology for the study of consciousness.
(2) The source of heterophenomenological data is (solely) first-person reports.
(3) Heterophenomenologists are agnostic about the truth value of first-person reports.
(4) Heterophenomenologists interpret reports as expressions of subjects' beliefs (about their conscious experience).
(5) Heterophenomenologists treat people as (near) incorrigible about what it's like to be them.
(6) Heterophenomenology licenses the same experiments as first-person methodologies, with the exception of 'lone-wolf autophenomenology' (Dennett, 2003, p. 23).[2]

I will devote one section of this paper to each of these assumptions. Revising them leads to a different third-person methodology of first-person data, which I call *self-measurement methodology*. The self-measurement methodology improves upon heterophenomenology in important ways while preserving the third-person spirit of the enterprise.

2. Setting Aside Consciousness

Dennett introduces heterophenomenology as a methodology for the study of consciousness — indeed, as '*the* bridge between the subjectivity of human consciousness and the natural sciences' (Dennett, 2007, p. 249, emphasis original). Dennett's concern with consciousness pervades his discussion of heterophenomenology. His writings

[2] In fairness to Dennett (2007), only *some* of the presently proposed amendments to heterophenomenology can be extracted from my 2003a. Specifically, the amendments to (1), (3), and (4) are already implicit in my 2003a; the rest come from later work.

on heterophenomenology contain references to zombies, the subjectivity of consciousness, and related topics. By contrast, one of my opening statements was that 'the present topic is not consciousness or qualia or any particular mental state … My topic is the use of introspective reports to generate evidence. What that evidence is *about* is a separate question, on which I will remain as neutral as possible' (Piccinini, 2003a, pp. 141–2). Let me explain why I recommend separating the methodology of first-person reports from the study of consciousness.

As Dennett surely knows, being a first-person report is neither necessary nor sufficient for being *about* consciousness. On one hand, consciousness can be studied using many kinds of data, including neurophysiological data, imaging data, and behavioural (not necessarily first-person) data. On the other hand, first-person reports are used as evidence by scores of scientists who have no particular interest in consciousness — including people studying problem solving, subjective wellbeing, and mental illness. In some cases, first-person reports may even be used as a window into the *un*conscious mind. For example, some psychotherapists ask subjects to report their dreams, fantasies, and free associations, all in an effort to extract information about states and processes that are inaccessible to the subjects themselves.

A heterophenomenologist might object that subjects who are reporting to their therapists about their dreams, fantasies, and free associations are still reporting on their *conscious* dreams, fantasies, and free associations — hence, their first-person reports are still about their conscious mind. In reply, I have three points to make. First, in the present context, what the reports are really 'about' is for the external observers to determine (more on this in Section 5). Nothing prevents external observers from determining that certain reports, suitably interpreted and understood, are about unconscious states and processes. Second, there may be conditions under which subjects issue reports about states and events that are unconscious in virtually every sense of the term. Consider reports issued under hypnosis: it has been argued that hypnotized subjects may report on mental states that they cannot access otherwise. I don't know whether this is true, but its possibility should not be ruled out a priori by the methodology of first-person reports.[3] Finally, the thesis under discussion is that the methodology of first-person reports coincides with a methodology for

[3] Dennett himself once argued that contrary to the received view that dreams are conscious experiences, '[i]t is an *open*, and *theoretical* question whether dreams fall inside or outside the boundary of experience' (1976, pp. 169–70).

the study of consciousness.[4] Patently, any use of first-person reports to study unconscious states and events falls outside this characterization. But there is nothing wrong with such a use. Thus, the methodology of first-person reports as scientific evidence is orthogonal to the study of consciousness.

This is fortunate, because the study of consciousness is fraught with controversies. Different authors think of consciousness in different ways — and they use the term 'consciousness' for different things. There is creature consciousness, state consciousness and transitive consciousness (Rosenthal, 1993); access consciousness and phenomenal consciousness (Block, 1995). It is likely that most consciousness-related phenomena have some connection or other to some kind of first-person reports. But these are not connections that can be established prior to the investigation of consciousness. Establishing the relations between first-person reports and the various phenomena called 'consciousness' is a difficult and controversial enterprise, which must be based on empirical investigation. It is best left to a theory of consciousness. The present topic is neither consciousness nor the connection between consciousness and first-person reports. The present topic is the methodology of first-person reports (or rather, as I will soon argue, first-person *behaviours* more generally). To avoid getting entangled with the difficult questions pertaining to consciousness, and in order to focus on our proper subject, we should separate the methodology of first-person reports from the theory of consciousness.

To sum up, if heterophenomenology is the methodology for studying consciousness, it is not quite what I'm after. I'm after the methodology of first-person reports — or more precisely, of first-person behaviours.

3. More than First-Person Reports: First-Person Behaviours

When Dennett discusses heterophenomenology, the only sources of data he considers are first-person reports — a kind of speech act. But first-person data need not be derived from first-person reports. For instance, subjects may be asked to express their level of happiness (or pain) by pointing at images of smiling vs. frowning faces. Or a subject may press one of two buttons depending on which of two ways she

[4] I do not mean to suggest that Dennett explicitly endorses this view; but the way he presents heterophenomenology suggests it.

perceives a Necker cube at a time. Such button pressing is a source of first-person data.

In some cases, first-person behaviours such as button presses and circles around smiley faces may be assimilated to first-person reports. As Dennett puts it, 'button-presses can be set up to be interpreted as speech acts as well, with highly specific meanings and fine temporal resolution' (Dennett, 2003, p. 20). This assimilation is somewhat plausible for button-presses by linguistically competent human beings. It is implausible, however, for button-presses by non-human animals.

Some researchers train monkeys to press one of two buttons depending on which of two ways they see a Necker cube, or closely analogous conditions (Leopold *et al.*, 2003). Such researchers treat data from monkey button-presses as first-person data, and with good reason: the button presses are the only reliable discriminating factor between different mental states of the subject under experimental conditions that are quite difficult to set up. But to call monkey button-presses 'first-person reports' is worse than a stretch — it's a misleading anthropomorphizing of the monkey, and it obscures the important differences between the ways human versus monkey data are obtained. After all, monkeys — unlike linguistically competent human beings — cannot be given linguistic instructions and cannot be debriefed about what they mean by their button presses. Thus, 'first-person report' is too restrictive to capture all sources of first-person data. Instead, I use 'first-person behaviour' to denote any behaviour that is a source of first-person data.[5]

4. Against Agnosticism

Dennett says that heterophenomenologists are agnostic about the truth value of first-person reports. As he puts it, heterophenomenology maintains '(contrary to everyday interpersonal communicative practice) a deliberate bracketing of the issue of whether what [subjects] are saying is literally true, metaphorically true, true under-an-imposed-interpretation, or systematically-false-in-a-way-we-must-explain' (Dennett, 2007, p. 252). I think this 'deliberate bracketing' is neither how most scientists generate first-person data nor how they should.

[5] Since Dennett focuses only on first-person reports about the contents of conscious experience at the expense of other first-person behaviours and other target mental states, I will often restrict my attention to first-person reports about the contents of conscious experience.

To begin with, I just argued in the previous section that not all first-person behaviours are verbal reports — first-person behaviours include such things as (some) monkey button-presses, which are neither true nor false. So at least in this case, the question of truth versus falsehood is beside the point. Nevertheless, monkey button-presses are informative. Depending on which button a monkey presses, and assuming that the monkey is well trained, a button-press informs external observers about something in the monkey's mind.

It is important to realize that the inference from specific monkey button-presses to specific mental events is supported by public evidence, such as evidence about the relevant neural systems and the way the monkey was trained. This point generalizes to all first-person behaviours including first-person verbal reports. Scientists need not be (and, I submit, generally are not) agnostic about the truth value of first-person reports. This is not to say that they accept the reports uncritically. On the contrary, scientists should (and for the most part do) use the best evidence they have, plus some educated guesses, to extract information from first-person reports.

By and large, this is also what ordinary folk do in their everyday life. People do not always trust what others tell them about their mind (nor do they remain agnostic about first-person reports; Dennett would agree with that). Sometimes people explicitly ask themselves whether their interlocutor — who is issuing first-person reports — is sincere, and if she is, whether she knows what she is talking about. Most of the time, people simply rely on their implicit folk psychological skills, which allow them to (fallibly) detect faults in other people's reports.

For example, virtually all parents learn to take their children's first-person reports with a large dose of salt. When a child is tired, hungry, or crabby, among other conditions, much of what she says about her mind is inaccurate. The way children's reports are delivered — the tone of voice, the amount of whining accompanying the reports, etc. — may be more informative than the literal meaning of the utterances. (The same is often true of adults, come to think of it.) A report may say something like, 'I need my special toy.' The message received may be something like, 'she is hungry.' Scientists do the same thing. The main difference is that scientists ought to be more careful and sophisticated about how they extract information from first-person behaviours — relying on controlled conditions, independent empirical studies, and scientific theories that are not generally available to the average folk.

Thus, contrary to Dennett's suggestion that scientific reliance on first-person behaviours marks a discontinuity with 'everyday interpersonal communicative practice', there is in fact continuity between ordinary and scientific reliance on first-person reports. Neither scientists nor ordinary folk take an agnostic stance; both make assumptions, based on public evidence, about the validity and reliability of first-person data.

One may object that all that Dennett suggested is that scientists ought to *begin* their experiments from an agnostic stance (about what may be inferred from first-person behaviours). Later, they surely can use other evidence to constrain their interpretation of the data.

This objection is unhelpful and inadequate. It is unhelpful because it is too vague about when 'later' is. When are scientists allowed to use evidence and make assumptions to interpret first-person behaviours? After they develop a theory of the phenomena? After they run an experiment? After they record the behaviours? After they design the experiment? I reject the presupposition of these questions: scientist need not be (and generally, are not) agnostic at any time. And this is why the objection is inadequate.

It is inadequate because if scientists began their investigation with an agnostic stance towards what first-person behaviours indicate, it's difficult to see how they could ever figure out what to learn from them. Consider an analogy. Suppose that Galileo had remained agnostic about the deliverances of his telescope. Not only would he not have revolutionized astronomy. Lacking the means to travel through outer space and check what is out there, he would have lacked any means to validate his data. But remain agnostic is not what Galileo did. He assumed that his telescope showed, by and large, the way heavenly bodies are — approximately in the same way that pointing the telescope towards earthly objects magnified them. This is not to say that everything a telescope shows is easy to interpret or even that it correlates with objective facts. Telescopes, like other measurement instruments, produce aberrations and artifacts. But the way to get around aberrations and artifacts is not to remain agnostic about the deliverances of the instruments — it is to collect as much evidence as possible, by whatever means are available, and use it to investigate which aspects of the data are valid and reliable and which are not.

In a separate paper (Piccinini, 2009), I argue that using first-person behaviours as sources of scientific data is an example of using subjects as measurement instruments — more specifically, as self-measuring instruments. As with telescopes and other measurement instruments, scientists need not be agnostic about instruments'

outputs. They should (and do) use the best evidence and best assumptions they can make to infer what their data show about the things they are measuring.

Scientists — like ordinary folk — bring their evidence and their assumptions with them from the very beginning of the scientific process. Evidence and assumptions inform the questions scientists ask, the tasks they set up, the experiments they design, the way they train the subjects (if applicable), the reports they collect, and the way they process first-person behaviours to extract information from them.

But there is a caveat. Scientists ought to (and often do) make their assumptions in interpreting first-person behaviours explicit as well as make their assumptions as few, weak, and uncontroversial as possible (cf. Ericsson and Simon, 1993). The fewer, weaker, and less controversial the assumptions, the harder — the more valid — the resulting data. Perhaps this is the kernel of truth in Dennett's claim of agnosticism: scientists need not aim at agnosticism, but they should aim at weak assumptions.

5. Which Mental States Can We Infer From First-Person Behaviours?

How should we interpret first-person behaviours — or at least first-person reports? According to Dennett, heterophenomenologists interpret first-person reports as expressing the subjects' *beliefs* about their (conscious) mental states. This interpretation is alleged to be theoretically unproblematic. More specifically, Dennett proposes the following:

> [A] nesting of proximal sources is presupposed as we work our way from raw data to heterophenomenological worlds:[6]
>
> (a) 'conscious experiences themselves';
> (b) beliefs about these experiences;
> (c) 'verbal judgments' expressing those beliefs;
> (d) utterances of one sort or another.

[6] A heterophenomenological world is a field of putative conscious experience 'constituted' by a subject's interpreted first-person reports (Dennett, 1991, p. 81). A heterophenomenological world may or may not correspond to what the subject actually experiences; it is analogous to a fictional world constituted by the statements made in a work of fiction (which may in fact describe reality correctly, at least sometimes). I will ignore Dennett's analogy between the interpretation of first-person reports and the interpretation of fiction, because I find the analogy misleading (following Schwitzgebel 2007). As I argue in the main text and elsewhere (Piccinini 2009), interpreting first-person reports (and more generally, first-person behaviours) is a special case of interpreting the output of a (self-) measuring instrument.

> What are the 'primary data'? For heterophenomenologists, the *primary* data are the sounds recorded when the subjects' mouths move, or (d) the utterances, the *raw* uninterpreted data. But before we get to theory, we can interpret these data, carrying us via (c) speech acts to (b) beliefs about experiences. These are the primary *interpreted* data, the pretheoretical data, the *quod erat explicatum* … Sticking to the heterophenomenological standard, and treating (b) as the maximal set of primary data, is a good way of avoiding a commitment to spurious data' (Dennett, 2005, pp. 44–5; cf. Dennett, 2003, p. 21).

I agree that we are aiming to learn about (a), conscious experiences (and more generally, mental states, as per Section 2). I also agree that the raw data are (d), (recordings of) utterances (more generally, recordings of first-person behaviours, as per Section 3). But I find the insertion of (b) and (c) unhelpful at best.

(b) and (c) refer to beliefs. Beliefs are not 'pretheoretical data', as Dennett suggests, but folk psychological theoretical constructs. There are different ways to explicate belief talk. Some authors think of beliefs as dispositions, others as monadic internal states, still others as structured mental representations. If we construe beliefs as any kind of objective state of the subject that acts as a 'proximal source' of the report in the sense of *causing* it, heterophenomenology is committed to two empirical assumptions about beliefs: (1) first-person reports are always mediated by beliefs; they cannot be produced directly from desires, emotions, memories, perceptual experiences, and other mental states without producing beliefs first; and (2) first-person reports always accurately represent the content of beliefs. Why should we accept these assumptions?

Suppose we ask a subject how she feels. Suppose she answers, 'I feel good.' A heterophenomenologist is not allowed to infer that the subject feels good; all she can infer is that the subject *believes* she feels good. Given a standard construal of belief as an objective state of the subject, this means that (i) the report is caused by a belief of the subject, and (ii) the report accurately represents the content of the belief. These assumptions are not always implausible, but there are two problems with them.

First problem: the assumptions may break down. It seems at least possible that, contra (i), the report was produced by verbalizing a positive feeling directly, rather than a belief about a positive feeling. If so, the subject may or may not have had *any* belief about her feeling (prior to issuing the report). It also seems possible that, contra (ii), if the subject did have a belief about her feeling prior to issuing the report, the

report failed to express her belief correctly. The subject might be lying to the interrogator or to herself, be confused, be confabulating, etc.

Dennett is aware that (ii) may be false. His way around it seems to be that the interviewer should make sure the subject expresses her beliefs correctly:

> When doubts arise about whether the subject has said what he meant, or understood the problem, or knows the meanings of the words being used, we can ask for clarifications. Usually we can resolve the doubts. Ideally, the effect of these measures is to remove all likely sources of ambiguity and uncertainty from the experimental situation, so that *one* intentional interpretation of the text … has no plausible rivals. It is taken to be *the* sincere, reliable expression by a *single, unified subject* of that very subject's beliefs and opinions [footnote omitted]. As we shall see, though, there are times when this presumption is problematic — especially when our subjects exhibit one pathology or another (Dennett, 1991, pp. 77–8, emphasis original; cf. also Dennett, 1982, pp. 162–3).

What remains unclear is why we should assume that reports always express beliefs about the target mental states rather than expressing the target mental states themselves. In light of the difficulties with making sure that subjects express their beliefs correctly, which Dennett points out, it is also unclear why heterophenomenologists should insist on interpreting all first-person reports as expressing beliefs. In other words, it is unclear why heterophenomenologists should first establish to the extent possible that a report expresses a belief, and then interpret the report as expressing the belief. A more prudent treatment of most reports is at hand: establish to the extent possible that a report expresses the target mental state (e.g., an emotion, desire, memory, or whathaveyou, including a belief if that's the target), then interpret the report as providing information about the target mental state. After all, that is what we are trying to investigate.

Here is another way to put the point. If interviewers can (try to) ensure that subjects correctly express their beliefs, as Dennett implicitly maintains, why not cut the middleman — the beliefs — and simply (try to) ensure that subjects correctly express the target mental states (feelings, desires, perceptual states, etc., including beliefs if those are the targets)? I submit that that's what interviewers should do — and for the most part, what they *do* do.

This suggestion helps us also solve the second problem with heterophenomenology's folk psychological assumptions. What we really wanted to know is whether the subject feels good — not whether she has a belief to that effect. By enforcing an interpretation

of reports in terms of beliefs, heterophenomenology is not helping us. When someone tells us she feels good, we may have reasons to doubt the report, but presumably such reasons should have to do with whether she feels good and is reporting correctly about it, not with what she believes about herself. A good methodology of first-person reports (mutatis mutandis, first-person behaviours) should interpret reports so as to extract information about the target mental states. It does not need to, and may be harmed by, interpreting reports as always expressing beliefs.

At this point, a careful reader of Dennett may object on the following grounds. According to Dennett, beliefs are not objective states of the person but 'theorists' fictions': 'abstractions that measure or describe the complex cognitive state of a subject' (Dennett, 2003, p. 20). This theory of belief, based on Dennett's 'intentional stance' (Dennett, 1987), may appear to avoid the above-mentioned concern that by accepting heterophenomenology, we are committing ourselves to problematic empirical assumptions. Ascribing beliefs on the grounds of first-person reports does not depend on empirical assumptions, the objection might go; it's just an explication of what it means to issue a report (cf. Dennett, 1991, pp. 303–9). In other words, 'X reports that P' entails that 'X believes that P' by definition. In fact, 'X believes that P' is little more than another way of saying that 'X reports that P'.[7]

But this objection only pushes the problem one step back. On one hand, if 'X believes that P' is just a theoretically unproblematic re-description of 'X reports that P,' then it would be simpler and less confusing to avoid this re-description altogether. On the other hand, it is not clear that Dennett's theory accurately construes belief as it's understood by scientists. Many people will think not; perhaps more importantly, many scientists will simply not follow Dennett's account when interpreting reports. What are they to do? If they take beliefs to be an objective state of the subject, they cannot accept heterophenomenology on pain of committing themselves to problematic empirical assumptions about beliefs.[8]

[7] For the sake of the argument, I'm still setting aside the observation, made in Section 3, that not all first-person data are derived from first-person reports.

[8] Most psychologists and neuroscientists who use first-person behaviours as evidence do quite well without ever using the term 'belief' or anything strictly equivalent to it. Some of them have developed fairly rigorous theories of the processes underlying (some kinds of) first-person behaviours, which theories support a sophisticated methodology for extracting information from them (Ericsson and Simon, 1993; Ericsson, 2003).

There is also another problem with this objection: it does not sit well with some other things that Dennett says about expressing beliefs. Dennett calls beliefs a 'proximal source' or 'proximal cause' of first-person reports: 'When a subject expresses a belief about anything, that belief is a (salient, crucial part of the) proximal cause of the expression' (Dennett, 2007, p. 253). In addition, Dennett says that belief expressions are *evidence* for beliefs: 'Expressing a mental state, deliberately or not, is just doing something that provides good evidence for, or makes manifest, that state to another observer' (Dennett, 1991, p. 306). But if reports are caused by beliefs and are evidence for them, belief ascriptions cannot be theoretically unproblematic re-descriptions of report ascriptions. If reports are caused by and evidence for beliefs, beliefs must be objective states of subjects that exist prior to the reports that express them. Once again, Dennett's 'nesting of proximal sources' seems to build empirical assumptions couched within folk psychological vocabulary into our methodology of first-person reports. That is unwarranted, unwise, and unnecessary.

It's preferable to avoid folk psychological notions and empirical assumptions within our methodology, if we can — and we can. What we have to do is to leave open as an empirical question what the proximal sources of first-person behaviours are, and leave up to empirical investigation what those sources are. This alternative methodology is both more descriptively and more normatively adequate.

Accordingly, I recommend the following replacement for Dennett's 'nesting of proximal sources':

(a′) mental states;
(b′) cognitive process that encode mental states into first-person behaviours;
(c′) first-person behaviours.

(a′) and (c′) are straightforward replacements for (a) and (d) respectively, in accordance with Sections 2 and 3. The crucial change is the replacement of both (b) and (c) with (b′). While (b) and (c) involve folk psychological assumptions of questionable empirical and theoretical standing, (b′) is simply a placeholder for whatever processes mediate the production of first-person behaviours. It is the process by which subjects produce first-person behaviours on the basis of the relevant mental states — whatever that process may be. Notice that the target mental states (a′) need not be conscious experiences, and even when they are, they certainly need not be understood under any particular theory of consciousness. By eliminating heterophenomenology's postulation of beliefs and by interpreting first-person behaviours

directly in terms of target mental states, the present proposal is maximally theory-neutral: what the target mental states are is an open empirical question, whose answer may vary from case to case.

One worry about this account is that it deems the subject always right about her mental states — or so it might seem. For it might seem that according to this account, first-person behaviours (c′) always correspond to the target mental states (a′). Thus, it might seem that the subject is always right about her mental states. But as Dennett (along with others) has convincingly argued, nothing could be further from the truth.

This worry is based on an equivocation. It's true that, according to the present account, first-person behaviours always correspond to the relevant mental states. But this is the correspondence of *causation* not that of *epistemic access*. In other words, (a′) via (b′) is here literally taken to cause (c′). From this, it doesn't follow that (c′) is an accurate description of (a′). In fact, it need not even be a description of anything, as in the case of monkey button-pressings.

Of course, in many circumstances first-person behaviours do describe mental states. If so, they may or may not be (partially or wholly) accurate. If an experiment is aimed at finding information about (a′) from (c′) and the experiment is set up well, then (c′) may well be a reliable guide to some features of (a′). But this is by no means guaranteed. There may be a gap between what a first-person report says and the mental states that are responsible for that report.

As an illustration, consider the experimental paradigm of 'choice blindness,' in which the outcome of a subject's choice is surreptitiously altered by an experimenter (Johansson *et al.*, 2005; 2006a,b; 2008). An experimenter shows a subject two cards. On each card is a picture of a face. The experimenter asks the subject to indicate which face he or she finds more attractive. After the subject points at one card, the experimenter drops the cards on the table, face down. She then gives the chosen card to the subject and asks the subject to explain why she made that choice. Sample answers include: 'The eyes are radiating there, and the mouth too it has that little… about to smile thing going on', 'Again, she was just more beautiful than she [pause] than the other', 'The other one looked a bit crazy, I guess this one had a better nose' (Johansson *et al.*, 2006a, p. 677).

During some trials, the experimenter employs a double-card magic trick: while pretending to present the subject with the card she chose, she actually presents the subject with a copy of the card the subject chose *against*. In at least 75% of these manipulated trials, subjects do not notice the switch. Yet they proceed to give reasons for their

alleged choice, as if it were the choice they actually made. Examples include: 'She's radiant. I would rather have approached her at a bar than the other one. I like earrings!', 'Because she was smiling' (Johansson *et al*., 2005, p. 118). Reports given in the manipulated condition are virtually indistinguishable from those given in the normal (non-manipulated) condition.

Several observations are worth making about how Johansson et al. interpret these reports.

First, contrary to heterophenomenology, Johansson et al. are not agnostic about the truth value of the reports. For reports issued under the normal condition, experimenters tentatively assume that the reports provide information about the reasons for the subjects' choices. By contrast, for reports issued under the manipulated condition, the experimenters tentatively assume that (i) any report that refers to specific features of the actually chosen face (e.g., being blond, smiling) probably provides accurate information about the reasons for the actual choice, whereas (ii) any report that refers to specific features of the *allegedly* chosen face (which was actually chosen *against*) is confabulatory. With respect to the latter class of reports, there is a wide chasm between the literal content of the reports and the mental states that cause them. The reports purport to describe the factors that motivated the choice of A over B, whereas in fact, they can't possibly do any such thing, because the subject actually chose B over A.

Second, contrary to the assumptions of first-person methodologists, Johansson et al. do *not* make any general assumptions about the reliability of first-person reports. Instead, they make specific assumptions about specific classes of reports, and justify their assumptions on public grounds, such as what choice the subject actually made and whether the features described in a report match the features of the depicted faces.

Finally, contrary to heterophenomenology, at no point do Johansson *et al*. interpret subjects' reports as expressing the subjects' beliefs. This is especially sensible in the manipulated condition, in which it is extremely unlikely that the reports are caused by prior beliefs. In the manipulated condition, subjects are trying to justify ex post facto an alleged choice that they did not actually make. It's hard to see why subjects should have any beliefs pertaining to such a non-existent choice prior to formulating their reports.

Johansson *et al.*'s treatment of their subjects' reports is a good example of what I am recommending. The bottom line is that what causes first-person behaviours is an open empirical question.

Scientists need not interpret all reports as expressions of beliefs (and, per Section 4, need not be agnostic about the truth value of the reports). Instead, they should use the best evidence they have, including knowledge of the stimuli, tasks, subject training, background conditions, and pertinent cognitive mechanisms, to establish — or at least make their best guess at — what can be inferred about the mind from first-person behaviours.

6. Dropping Incorrigibility

Although Dennett is commendably explicit and insistent that subjects can be wrong about what goes on in their conscious experience, he occasionally seems to say just the opposite: '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*' (1991, p. 96, emphasis original). This passage sounds contradictory, because the phrase 'what it's like' is generally used to mean the contents of phenomenal consciousness — the very same contents that the first statement in this passage says subjects lack authority about (cf. Schwitzgebel, 2007, who raises the present problem and suggests dropping incorrigibility).

If subjects can be wrong about the contents of experience, how can subjects also have 'total, dictatorial authority' over them? A charitable way to read Dennett, so as to avoid attributing him an inconsistency, is to reinterpret Dennett's use of 'what it's like' to refer not to the contents of conscious experience but to the beliefs about their experience that, according to heterophenomenology, subjects express by their first-person reports. In other words, using the scheme presented in the previous section, Dennett's 'what it's like' refers to (b) not (a).[9]

This interpretation generates a dilemma. Either beliefs are construed as *thin* beliefs (mere re-descriptions of first-person reports), which makes the claim of incorrigibility true but trivial, or beliefs are construed as *thick* beliefs (objective causes of the reports), which makes the claim of incorrigibility nontrivial but false.

First horn: beliefs are understood thinly, as something like attributions made from the intentional stance for the convenience of observers, attributions that do not necessarily correspond to any objective state of the subject, so that any report expresses a belief by definition.

[9] This interpretation is a slight reformulation of the one offered by Schwitzgebel at http://schwitzsplinters.blogspot.com/2007/07/making-sense-of-dennetts-views-on.html (as accessed on 10/28/2009).

If so, then the (near) incorrigibility that is part of heterophenomenology is trivial. It is equivalent to the claim that subjects get to issue their own reports — they get to say what they say. In other words, this incorrigibility claim is equivalent to saying that any corrigible first-person report to the effect that *P* can be reformulated as a (near) incorrigible report to the effect that *it seems to the subject that P*. The reason why this is trivial is that the same thing can be done with *any* report, whether or not it is about the mind of the subject. If someone says, 'It will rain tomorrow,' this is not an especially authoritative or incorrigible statement. But we can rephrase the assertion as saying, 'It seems to me that it will rain tomorrow.' This new report is authoritative and (near) incorrigible. But this has nothing to do with consciousness or the mind.

Second horn: beliefs are understood thickly, as objective states of the subject that cause the reports. If so, then the incorrigibility is nontrivial, but there is no reason to grant it. Indeed, there is no reason even to suppose that thick beliefs always exist as a source of reports. For one thing, subjects may well make mistakes in expressing their thick beliefs. (Dennett knows this, but he seems to finesse it by assuming that possible confounding factors have been taken care of before the reports are interpreted. But then the incorrigibility claim begins to slide towards triviality again.) For another thing, why couldn't subjects, at least in some cases, formulate reports directly on the basis of the target mental states (which need not be beliefs), rather than always formulate a (thick) belief first?

We can avoid this dilemma entirely by dropping any claim of (nontrivial) incorrigibility and endorsing the self-measurement methodology instead. According to the latter, first-person behaviours are the result of self-measurement on the part of the subject. Like other measurements, self-measurements are neither incorrigible nor always reliable. It is the job of the external observers to calibrate the instrument, set up the experiment, use the instrument, and interpret the data so as to rule out confounding factors and establish what phenomena can be inferred from the data.

7. How Methodology Affects Scientific Practice

Dennett maintains that heterophenomenology and 'first-person methodologies' allow the same range of experiments (Dennett, 2003, pp. 26, 29; for a concurring opinion, see van de Laar, 2008). More generally, Dennett seems to think that third-person and first-person methodologies are two different ways of describing the same experiments,

with no practical implications for scientific practice due to choosing one over the other. The only exception is 'lone-wolf autophenomenology, in which the subject and experimenter are one and the same person' (Dennett, 2003, p. 23). Lone-wolf autophenomenology is explicitly banned by heterophenomenology, whereas it is allowed by first-person methodologies.

I couldn't agree more that lone-wolf autophenomenology — regardless of its heuristic value, which may well be significant — is no scientific method. The chances for idiosyncrasy, bias, and self-serving conclusions are too obvious and too great.

Lone-wolf autophenomenology is the only method that both parties seem to agree is private. If so, then first-person methodologists should allow lone-wolf autophenomenology while third-person methodologists reject it. That would be a prescriptive difference between the two views, and third-person methodologists would give the right prescription. Curiously, however, some prominent first-person methodologists, such as David Chalmers and Alvin Goldman, shy away from defending lone-wolf autophenomenology, nor do they complain that scientific journals reject papers based solely on lone-wolf autophenomenology even though lone-wolf autophenomenology does not seem to violate any tenet of first-person methodology. It would be interesting to hear on what grounds first-person methodologists who reject lone-wolf autophenomenology do so.

A candid and sophisticated defense of lone-wolf autophenomenology is offered by Charles Siewert (2007) under the banner of 'plain phenomenology'. Plain phenomenology consists in theorizing about the mind based on 'a source of warrant special to' first-person evidence; in addition, such special warrant is not derived from third-person evidence (*ibid*., p. 202). Although Siewert does not say explicitly whether plain phenomenology belongs in the sciences of the mind, I assume it is intended to, since Siewert presents it as an alternative to heterophenomenology.

As far as science is concerned, first-person evidence has no special warrant. In fact, first-person evidence is scientifically illegitimate (Piccinini, 2003b). This is not to say that scientists' conscious experience is irrelevant to understanding scientific discourse. As Siewert points out, we wouldn't understand much of our discourse about minds if we couldn't apply such discourse within our own conscious experience. And, I add, the same is true of the rest of science: we couldn't understand much of our discourse about physical objects and properties if we didn't experience physical objects and properties and couldn't apply our physical discourse to what we experience.

Presumably, no science — or at least no science as human beings practice it — is even possible without conscious experience. This goes for physics no less than for the study of the mind. Nevertheless, scientific theorizing is constrained and tested by public not private (i.e., first-person) evidence.

Siewert objects to heterophenomenology's agnosticism about first-person reports (more generally, first-person behaviours). Instead of agnosticism, Siewert proposes that the alleged 'special warrant' of first-person evidence entitles us to presume that first-person evidence is correct (until proven wrong). I agree with Siewert that heterophenomenology's agnosticism is unwarranted and unnecessarily limiting. But this feature of heterophenomenology is not mandated by third-person methodology. Thus, the right corrective is not to abandon heterophenomenology for Siewert's first-person methodology. As I argued in Section 4, the right corrective is to use our best (public) evidence and most plausible assumptions to interpret our (public) data.

This critique of lone-wolf autophenomenology can be generalized to the practical implications of first-person methodology for scientific practice, in a way that goes beyond what Dennett says. It will help to separate two claims: first, that there are no prescriptive differences between first-person methodologies and third-person methodologies; second, that the same set of studies is permitted by first-person methodologies and third-person methodologies. The second claim is a special case of the first.

There are in fact important prescriptive differences between first-person and third-person methodologies, as I also argue elsewhere (Piccinini forthcoming). The questions we should ask are the following: Which view leads to more rigorous, less biased, and more reliable experimental practices? How does each methodology suggest taking precautions against experimental artifacts, searching for and ruling out confounding factors (such as delusion, confabulation, etc.), making explicit assumptions in generating the data, and articulating explicit procedures for encoding the data? Which methodology recommends pursuing research programs aimed at publicly validating the data (or invalidating them, as the case may be)?

The critical difference between first-person and third-person methodologies is where they place the epistemic burden of validating first-person data. According to first-person methodologies, the subjects are the true scientific observers: they are the ones who must avoid biases and confounding factors; they are responsible for telling the truth about their mind; no one else can look into the subjects' minds, so no one else can truly check. Besides noticing

inconsistencies, external observers cannot do anything to either validate or invalidate first-person data (Chalmers, 2004; Gertler, 2009; Goldman, 1997; Hatfield, 2005).

By contrast, according to third-person methodologies, the subjects are not scientific observers but part of the experiment's 'materials' (and more specifically, as far as the generation of first-person behaviours is concerned, they constitute part of the measurement apparatus): it's the external observers who must establish what first-person behaviours can tell us about the subjects' mind; external observers, not subjects, have the burden of avoiding biases and confounding factors.

Given this difference, the critique of lone-wolf autophenomenology generalizes to any experimental procedure that follows first-person methodological precepts. Experimental subjects should not be expected to avoid all biases and confounding factors while generating first-person behaviours. They are simply not in a position to do so. Thus, insofar as a method places such a burden on the experimental subjects, it is dubious.

One may object that I am taking the statements of first-person methodologists too literally. Even self-avowed first-person methodologists (e.g., Lutz *et al*., 2002) perform experiments like everyone else. It's the experimenters (rather than the subjects) who look for confounders, try to avoid experimental artifacts, and perform statistical tests in line with standard scientific methodology. Experimenters may even look for reliable correlations between first-person data and third-person data, thereby offering independent evidence that the first-person data are valid.

It's true that when methodological push comes to experimental shove, at least some first-person methodologists do not behave as such. Rather than applying their own methodological prescriptions, they follow standard, third-person methodology. This is all the more reason to reject first-person methodologies. But there are two caveats. First, there is no guarantee that all first-person methodologists will apply standard methodological checks in their studies — especially when their own methodology does not prescribe them because it deems them impossible. In fact, even lone-wolf autophenomenology is far from eradicated. Second, methodological rigor (and consequently, publicity of method) comes in degrees. Even if you are a first-person methodologist in name only — someone who in fact follows a third-person methodology — your background commitments may still incline you towards cavalier reliance on subjective reports as

opposed to checking data as much as you can and striving for public validation.

For example, Lutz *et al.*'s study (2002), which is often cited as a paradigmatic example of first-person method, is especially opaque on how the first-person behaviours were collected and clustered into categories. The authors do not say how often subjects described their experience one way or another, what assumptions were made in encoding the data, how many experimenters encoded the data, how much the encoders agreed in their clustering of the data, and whether the encoders were blind to the hypothesis being tested. Ideally, a third-person methodology recommends processing the data so as to maximize their publicity — that is, making explicit assumptions about how to cluster the data and having two encoders, blind to the hypothesis being tested, encode the data independently of one another. These procedures would contribute to maximizing data publicity and minimizing sources of bias during the phases of data collection and encoding. As we have seen, whether or not Lutz *et al.* proceeded in the way recommended by third-person methodology, they did not see the need to inform their readers. This may be a consequence of their self-avowed first-person methodology.

In conclusion, there are prescriptive differences — in addition to the rejection of 'lone-wolf autophenomenology' — between serious third-person methodologies and first-person methodologies, and the third-person prescriptions are the ones that maximize the validity of the data. This is because (1) first-person methodologies incorrectly maintain that the subjects are the scientific observers, even though the subjects are not in a position to eliminate their biases, (2) there is more to sound methodology than employing many subjects and separating the experimenter from the subjects, and (3) in practice, publicity of methods and data comes in degree — the more publicity, the merrier.

8. Conclusion

Dennett often maintains that heterophenomenology accurately describes the methods employed by scientists when they employ first-person reports as evidence (e.g., Dennett, 2003, p. 22; 2005, pp. 36, 50). If that were true, I would recommend that scientists refine their methodology along the lines suggested here. But in fact, most scientists who use first-person reports — and more generally, first-person behaviours — as sources of data do it in a way that is more accurately captured by the self-measurement methodology I advocate than by heterophenomenology as Dennett describes it.

I have argued that heterophenomenology can be improved by making six changes: (i) setting aside consciousness, (ii) including other sources of first-person data besides first-person reports, (iii) abandoning agnosticism as to the truth value of the reports in favor of the most plausible assumptions we can make about what can be learned from the data, (iv) interpreting first-person reports (and other first-person behaviours) directly in terms of target mental states rather than beliefs about them, (v) dropping any residual commitment to incorrigibility of first-person reports, and (vi) recognizing that third-person methodology does have (positive) effects on scientific practices. When these changes as made, heterophenomenology turns into the self-measurement methodology that I have defended in previous papers.

According to the self-measurement methodology of first-person data, scientists treat subjects issuing first-person behaviours as a self-measuring instrument. It is the scientist's responsibility to insure that the instrument is calibrated and set up properly and that the measurement is carried out correctly. It is also the scientist's responsibility to determine what can and cannot be reliably measured by the instrument under the relevant circumstances. That being said, heterophenomenology got two important points right: first-person behaviours are not always a reliable window into the mind, and their use as evidence ought to be a form of third-person science.

References

Block, Ned (1995), 'On a confusion about a function of consciousness', *Behavioral and Brain Sciences*, **18**, pp. 227–87.

Chalmers, David J. (2004), 'How can we construct a science of consciousness?', in *The Cognitive Neurosciences III*, ed. Michael S. Gazzaniga (Cambridge, MA: MIT Press), pp. 1111–19.

Dennett, Daniel C. (1976), 'Are dreams experiences?', *Philosophical Review*, **85**, pp. 151–71.

Dennett, Daniel C. (1982), 'How to study consciousness empirically, or nothing comes to mind', *Synthese*, **53**, pp. 159–80.

Dennett, Daniel C. (1987), *The Intentional Stance* (Cambridge, MA: MIT Press).

Dennett, Daniel C. (1991), *Consciousness Explained* (Boston: Little, Brown & Company).

Dennett, Daniel C. (2003), 'Who's on first? Heterophenomenology explained', *Journal of Consciousness Studies*, **10** (9–10), pp. 19–30.

Dennett, Daniel C. (2005), *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness* (Cambridge, MA: MIT Press).

Dennett, Daniel C. (2007), 'Heterophenomenology reconsidered', *Phenomenology and the Cognitive Sciences*, **6,** pp. 247–70.

Ericsson, K. Anders (2003), 'Valid and non-reactive verbalization of thoughts during performance of tasks: Toward a solution to the central problems of introspection as a source of scientific data', *Journal of Consciousness Studies*, **10** (9–10), pp. 1–18.

Ericsson, K. Anders and Herbert A. Simon (1993), *Protocol Analysis: Verbal Reports as Data* (Cambridge, MA: MIT Press).
Goldman, Alvin (1997), 'Science, publicity, and consciousness', *Philosophy of Science*, **64**, pp. 525–45.
Gertler, Brie (2009), 'Introspection', in *The Oxford Companion to Consciousness*, eds. Tim Bayne, Axel Cleeremans, and Patrick Wilken (Oxford: Oxford University Press).
Hatfield, Gary (2005), 'Instrospective evidence in psychology', in *Scientific Evidence: Philosophical Theories and Applications*, ed. Peter Achinstein (Baltimore: Johns Hopkins University Press)**,** pp. 259–86.
Johansson, Petter and Lars Hall (2008), 'From change blindness to choice blindness.' *Psychologia*, **51**, pp. 142–55.
Johansson, Petter, Lars Hall, Sverker Sikström and Andreas Olsson (2005), 'Failure to detect mismatches between intention and outcome in a simple decision task', *Science* **310**, pp. 116–19.
Johansson, Petter, Lars Hall, Sverker Sikström, Betty Tärning and Andreas Lind (2006a), 'How something can be said about telling more than we can know: On choice blindness and introspection', *Consciousness and Cognition*, **15**, pp. 673–92.
Johansson, Petter, Lars Hall, Sverker Sikström, Betty Tärning and Andreas Lind (2006b), 'Reply to commentary by Moore and Haggard', *Consciousness and Cognition*, **15**, pp. 697–99.
Leopold, David. A., Alexander Maier and Nikos K. Logothetis (2003), 'Measuring subjective visual perception in the nonhuman primate', *Journal of Consciousness Studies*, **10** (9–10), pp. 115–30.
Lutz, Antoine, Jean-Philippe Lachaux, Jacques Martinerie and Francisco J. Varela (2002), 'Guiding the study of brain dynamics by using first-person data: Synchrony patterns correlate with ongoing conscious states during a simple visual task', *Proceedings of the National Academy of Sciences USA*, **99**, pp. 1586–91.
Piccinini, Gualtiero (2003a), 'Data from introspective reports: Upgrading from commonsense to science', *Journal of Consciousness Studies*, **10** (9–10), pp. 141–56.
Piccinini, Gualtiero (2003b), 'Epistemic divergence and the publicity of scientific methods', *Studies in the History and Philosophy of Science*, **34**, pp. 597-612.
Piccinini, Gualtiero (2009), 'First-person data, publicity, and self-measurement', *Philosophers' Imprint*, **9**, pp. 1–16.
Piccinini, Gualtiero (forthcoming), 'Scientific methods ought to be public, and descriptive experience sampling is one of them', *Journal of Consciousness Studies*.
Rosenthal, David (1993), 'State consciousness and transitive consciousness', *Consciousness and Cognition*, **2**, pp. 355–63.
Schwitzgebel, Eric (2007), 'No unchallengeable epistemic authority, of any sort, regarding our own conscious experience — contra Dennett?', *Phenomenology and the Cognitive Sciences*, **6**, pp. 107–13.
Siewert, Charles (2007), 'In favor of (plain) phenomenology', *Phenomenology and the Cognitive Sciences*, **6**, pp. 201–220.
van de Laar, Tjeerd (2008), 'Mind the methodology: Comparing heterophenomenology and neurophenomenology as methodologies for the scientific study of consciousness', *Theory and Psychology*, **18**, pp. 365–79.
Varela, Francisco J. and Jonathan Shear (1999), 'First person methodologies: What, why, how?' *Journal of Consciousness Studies*, **6** (2–3), pp. 1–14.