

# **Science and the First-person**

**Lynne Rudder Baker**

**University of Massachusetts Amherst**

I want to raise a question for which I have no definitive answer. The question is how to understand first-personal phenomena—phenomena that that can be discerned only from a first-personal point of view. The question stems from reflection on two claims: First, the claim of scientific naturalism that all phenomena can be described and explained by science; and second, the claim of science that everything within its purview is intersubjectively accessible, and hence that all science is constructed exclusively from the third-personal point of view. Using these two claims as premises, we can construct a simple valid argument, which I'll label 'The Master Argument:'

## The Master Argument

1. All phenomena can be described and explained by science.
2. All science is constructed exclusively from a third-personal point of view.
- ∴ 3. All phenomena can be described and explained from a third-personal point of view.

I shall argue that there are strong reasons to suppose

that there are irreducible first-personal phenomena—phenomena that can be discerned only from a first-personal point of view. If there are any irreducible first-personal phenomena, then the conclusion of the Master Argument is false. But the Master Argument is valid; so, if its conclusion is false, then one of the premises is false. Although I'll make some tentative remarks about the premises of the Master Argument, my main goal is to show that third-personal sciences cannot explain everything there is to be explained.

By the phrase 'irreducible first-personal phenomena', I do not mean anything Cartesian or "ghostly." Rather, by calling something an 'irreducible first-personal phenomenon,' I mean only that it cannot be understood in third-personal, scientific terms. It cannot be understood without using first-personal terms. Broadly speaking, there are at least two kinds of apparent first-personal phenomena that (surprisingly) are not usually discussed together.

I'll use the terms 'subjective' and 'nonsubjective' to distinguish the two kinds of first-personal phenomena. Subjective first-personal phenomena are our conscious experiences. These include sensations characterized by "what it's like"—for example, the peculiar smell of garlic or the softness of velvet. Subjective first-personal phenomena also include our awareness of our intentional states—our

thoughts, hopes, fears, wishes, beliefs, desires, intentions. In short, subjective first-personal phenomena are states of awareness of the contents of one's own mind.

Nonsubjective first-personal phenomena, by contrast, are states of awareness of oneself from the first-person as an entity in the world. These states are expressed in first-personal language; e.g., "I'm the one in the red dress." Being aware of who one is or where one is are examples of nonsubjective first-personal phenomena.

My aim is to show that there are irreducible first-personal phenomena of both subjective and nonsubjective varieties, and hence that the conclusion of the Master Argument is false. In the first section, I'll discuss Daniel Dennett's attempt to eliminate first-person phenomena by a method that he calls 'heterophenomenology.' Then, I'll give two examples to show the inadequacy of heterophenomenology.<sup>1</sup> Next, I'll turn to nonsubjective first-personal phenomena, and present three examples that also resist third-personal treatment. Assuming that I have shown that there are irreducible first-personal phenomena,

---

<sup>1</sup> I focus on Dennett because his third-personal account of apparent first-personal phenomena has been most influential, and because his approach is more likely to be able to handle complex subjective first-personal phenomena than are exclusively "bottom-up" approaches like Thomas Metzinger's tour de force *Being No One: The Self-Model Theory of Subjectivity* (Cambridge MA: MIT Press, 2004). See my "Naturalism and the First-personal Perspective," in *How Successful is Naturalism?*, Georg Gasser, ed. (Frankfurt: Ontos Verlag, 2007): 203-226.

I'll briefly consider which of the premises of the Master Argument should be rejected.

### **Dennett's Heterophenomenology**

The aim of heterophenomenology is to give third-personal explanations of everything that needs explaining and thus to treat first-personal phenomena within the strictures of third-personal science.<sup>2</sup> The term 'heterophenomenology' is well-chosen; it suggests a kind of third-personal phenomenology. In contrast to a traditional phenomenologist, who describes her own mind in the first person, a heterophenomenologist uses third-personal language to describe a subject's mind from the subject's point of view. Dennett requires that anything that counts as a phenomenon be publicly observable. Thus, he holds that a theory of the conscious mind "will have to be constructed from the third-personal point of view since all science is constructed from that perspective."<sup>3</sup> Nevertheless, he claims that heterophenomenology can "do justice to the most private and ineffable subjective experiences, while never abandoning the methodological principles of science."<sup>4</sup>

---

<sup>2</sup> Dennett, "The Fantasy of First-person Science" online at <http://ase.tufts.edu/cogstud/papers/chalmersdeb3dft.htm> (accessed 1/21/08), p. 8.

<sup>3</sup> Daniel Dennett, *Consciousness Explained* (Boston: Little, Brown and Co., 1991): 71.

<sup>4</sup> Dennett, *Consciousness Explained*, p.72.

Here is how heterophenomenology is to work. Begin with a theorist and a subject who is an adult human being. The subject reports how things seem to her, from her point of view. The theorist makes a sound tape, perhaps accompanied by a video tape and an electroencephelogram. These devices record observable features of the subject: the sounds the subject emits, her bodily motions, her brain states. The recordings provide the raw data for the theorist, who puts the taped sounds through several processes of interpretation.

These processes of interpretations are to be understood in terms of Dennett's intentional-stance theory of content. (Intentional-stance theory provides a way to use intentional language like 'Jill believes that the door is closed' without commitment to any particular state of affairs in reality; the intentional language is only a tool for prediction.) The details of intentional-stance theory don't matter here. What matters is that it legitimates heterophenomenological interpretations in which "we treat the noise-emitter as an agent, indeed, a rational agent, who harbors beliefs and desires and other mental states that exhibit intentionality or 'aboutness', and whose actions can be explained (or predicted) on the basis of the content of these states."<sup>5</sup>

---

The theorist begins by interpreting the noises from the

<sup>5</sup> Dennett, *Consciousness Explained*, p. 76.

tape as sentences—for example, if the subject says what sounds like ‘the spot moved from left to right’, the theorist interprets the subject’s utterance as ‘the spot moved from left to right’.<sup>6</sup> Then, there is a second level of interpretation—from sentences to speech acts. For example, if the the first level of interpretation attributed to the subject the sentence, “The spot move from left to right, the second level of interpretation may be, “The subject reported that the spot moved from left to right.” At that point, the tape recordings have been turned into interpreted texts, which reveal the world according to the subject—like the world according to Garp, or Sherlock Holmes’s London. The theorist now has a text of the world from the subject’s own point of view—all without giving up science.<sup>7</sup>

The data to be explained are taken from the interpreted text of the subject’s verbal reports about how things seem to her. The theorist grants that these verbal reports are expressions of belief.<sup>8</sup> What heterophenomenology is to explain are the expressions of belief or reports that things seem a certain way to the subject. Heterophenomenology makes no assumptions about the existence of any actual conscious phenomena—

---

<sup>6</sup> Dennett, *Consciousness Explained*, p. 75.

<sup>7</sup> Dennett, *Consciousness Explained*, p.76.

<sup>8</sup> Recent brain research has shown that certain parts of the brain are more active when a person is lying or trying to deceive. So, brain imaging techniques can be used to show that subjects are sincere in their reports, and hence that they really had the beliefs in question.

such as its occurring to me that I left my keys in my office or my smelling burning rubber. All that needs explaining is the expression or report of conscious episodes; the conscious episodes themselves are invisible to heterophenomenology. As Dennett puts it, “what has to be explained by theory is not the conscious experience, but your belief in it (or your sincere verbal judgment, etc.).”<sup>9</sup>

From the theorists’ point of view, the subjects may be *zombies*. The word ‘Zombie’ is a philosophers’ term for hypothetical entities that are behaviorally indistinguishable from human beings, but have no consciousness of anything. They have no subjective experience at all; they are “all dark inside.” In Dennett’s words: “a zombie is or would be a human being who exhibits perfectly natural, alert, loquacious, vivacious behavior but is in fact not conscious at all, but rather is some sort of automaton.”<sup>10</sup> The neutrality between conscious beings and zombies is crucial to heterophenomenology. It insures that heterophenomenology “makes no assumption about the actual consciousness of any apparently normal adult human beings.”<sup>11</sup> As Dennett puts it, heterophenomenology “maintains a nice neutrality: it characterizes [subjects’] beliefs, their heterophenomenological worlds, without passing judgment, and then investigates to see what could

---

<sup>9</sup> Dennett, “The Fantasy of First-person Science,” p. 4.

<sup>10</sup> Dennett, *Consciousness Explained*, p. 73.

<sup>11</sup> Dennett, *Consciousness Explained*, p. 73.

explain *the existence of those beliefs.*"<sup>12</sup> So, heterophenomenology is

“a neutral path leading from objective physical science and its insistence on the third-personal point of view to a method of phenomenology that can (in principle) do justice to the most private and ineffable subjective experiences, while never abandoning the methodological scruples of science.”<sup>13</sup>

The methodological scruples of science ban any subjective experience that is not publicly accessible. And what guarantees that these scruples are observed is the neutrality between human beings (presumably with consciousness or “consciousness”) and zombies (beings without consciousness). So, we can formulate a two-part Neutrality Assumption:

Heterophenomenology (i) makes no distinction between human beings and zombies (who, by stipulation, have no conscious experience) and hence (ii) explains, not conscious experience itself, but only entities’ reports and beliefs about conscious experience.

By focusing on the interpreted tapes of the subject’s reports

---

<sup>12</sup> Dennett, “The Fantasy of First-person Science,” p. 3.

<sup>13</sup> Dennett, *Consciousness Explained*, 72; “The Fantasy of a First-person Science,” p. 2



and beliefs, the heterophenomenologist has the intersubjectively accessible data needed by third-personal science. No actual conscious or subjective experience can be recognized by heterophenomenology. Only reports and expressions of beliefs about conscious experience need to be explained. So, heterophenomenology explains how a subject came to have beliefs (which a zombie counterpart has too) about conscious experience.

Dennett describes heterophenomenology as “nothing but good old third-personal scientific method applied to the particular phenomena of human (and animal) consciousness.”<sup>14</sup> It allows us to “trade in the first-personal perspective of Descartes and Kant for the third-personal perspective of the natural sciences and answer all the questions—without philosophically significant residue.”<sup>15</sup>

In sum, Dennett’s view is that there is nothing essentially first-personal that science needs to recognize in order to explain all phenomena. All that needs to be explained about the way the world seems to a subject are the subject’s dispositions to react and to report. And these phenomena, it is assumed, can be understood in wholly third-personal terms. Indeed, Dennett has described his view as “third-personal absolutism.”<sup>16</sup>

---

<sup>14</sup> Dennett, “The Fantasy of First-person Science,” p. 4.

<sup>15</sup> Dennett, “The Fantasy of First-person Science,” p. 1.

<sup>16</sup> This description was reported by David J. Chalmers in “Moving Forward on the Problem of Consciousness”

## Subjective First-personal Phenomena Reclaimed

Now I want to consider two examples that suggest that heterophenomenology misses genuine phenomena that are first-personal.

1. Suppose that a heterophenomenologist has a subject, Jack, who tells the theorist this story: "I was standing on a tall ladder painting the ceiling; I lost my balance, and as I was falling, I had the thought, 'I'm about to die.' When I hit the floor, I lost consciousness. Some time later, I gradually became aware of the smell of fresh paint. That was when I realized that I was still alive." The theorist makes a transcript and interprets it, in part, to say that Jack reported that he lost consciousness, and that Jack reported becoming aware of the smell of fresh paint.<sup>17</sup> The theorist takes both these reports at face value and interprets them as expressing Jack's beliefs. This interpreted transcript yields the intersubjective data that are the stopping point for Dennett's heterophenomenology.

First, note that heterophenomenology allows us to say that it's *true* that Jack had the beliefs that he lost consciousness and later became aware of the smell of fresh paint; but we cannot say that his *beliefs* are *true*. If

---

(<http://consc.net/papers/moving.html>) (1995). Accessed 1/23/2008.

<sup>17</sup> In "Get Real," p. 527, Dennett takes me to task for not distinguishing between expressing and reporting a belief, but here I am talking about verbal reports of what he was aware of.

heterophenomenology affirmed that Jack's beliefs that he became aware of the smell of fresh paint, for example, was true, it would have to affirm that the subjective experience of becoming aware of the smell of fresh paint was a genuine phenomenon. And if it affirmed that the subjective experiences was a genuine phenomenon, then it would violate the Neutrality Assumption and lose the neutrality between Jack and a zombie. The reason that we imagined zombies in the first place was to insure that no irreducible first-personal phenomena would slip into our science. (A zombie, recall, is stipulated to have no subjective experience, to be like us—to express the same beliefs that we express—but not be conscious.) So, a heterophenomenologist cannot allow that Jack's beliefs about becoming aware of the smell of fresh paint and so on are true—on pain of contradicting the neutrality between Jack and zombies. To contradict the neutrality between Jack and zombies would undermine the whole project of heterophenomenology. Heterophenomenology would no longer have a claim to be an exclusively third-personal science.

Someone may object: "The recent research on neural correlates gives us at least some hope that we may yet get neural evidence of consciousness. So, it may be at least possible that we could get third-personal evidence for the truth of Jack's belief that he lost consciousness." But this is

no help to the heterophenomenologist. The only kind of link that a heterophenomenologist could recognize would be a link between brain states and reports or expressions of conscious experiences (both of which are intersubjectively accessible), not a link between brains states and conscious experiences that are not intersubjectively accessible. Again, conscious experiences themselves cannot be recognized on pain of violating the Neutrality Assumption.

The upshot of the story of Jack and his mishap on the ladder is that heterophenomenology cannot consider Jack's sincere reports of what happened to him to be true. Otherwise, the Neutrality Assumption is violated. (If his belief that he smelled fresh paint were true, then the heterophenomenologist would have to recognize a phenomenon that is not publicly accessible—viz., Jack's smelling fresh paint.) But by not allowing Jack's sincere expressions and reports of his conscious experience to be true, heterophenomenology rejects the natural explanation of his beliefs. The natural explanation is that Jack believed that he lost consciousness and then smelled fresh paint because he did lose consciousness and then smelled fresh paint.

Traditionally, people are accorded presumptive authority over their current apparently conscious states. If someone says, "I smell fresh paint," we suppose that they

are having some olfactory experience. (Maybe it's not fresh paint but fresh coffee, but the person is having the experience of smelling something.) If heterophenomenology is going to take away the natural explanation of why people have beliefs that they have subjective experiences, then it will have to show how they could ever had such beliefs at all. How could anybody have beliefs like Jack's that he lost consciousness and later became aware of the smell of fresh paint if they were not true?

Dennett may respond that all the Neutrality Assumption requires is that the theory not affirm that the beliefs about subjective experience are true. They still might be true, for all we know. But if they are true, then there are irreducible subjective first-personal phenomena and the conclusion of the Master Argument is false. In any case, heterophenomenology makes it a mystery how people could ever believe that they had any subjective experiences whatever.

Now turn to the second example that shows that conscious episodes are not just data to be explained (as they were in Jack's case), but also part of the explanatory apparatus for explaining overt behavior.

2. Suppose a subject, Jill, tells the heterophenomenologist this story: "I'm a high-school

student and I was assigned to stand up in front of my English literature class and recite Macbeth's soliloquy that begins, "Tomorrow and tomorrow and tomorrow, creeps in this petty pace from day to day."<sup>18</sup> Since I had not bothered to practice reciting the soliloquy the night before, I practiced on the bus to school by reciting the famous soliloquy to myself. I practiced silently because I would have felt silly reciting out loud in front of all these people on the bus. If I hadn't practiced, I would have botched the recitation. My silently practicing the soliloquy on the bus explains my perfect recitation in front of the class."

The theorist would take as data Jill's report that she practiced the soliloquy silently on the bus, but (as in the case of Jack) not that she actually practiced silently. The Neutrality Assumption rules out recognition of phenomena like silently practicing. Silently practicing is no more intersubjectively accessible than is becoming aware of the smell of fresh paint.

To practice is not just for a brain to undergo a

---

<sup>18</sup> To-morrow, and to-morrow, and to-morrow,  
Creeps in this petty pace from day to day,  
To the last syllable of recorded time;  
And all our yesterdays have lighted fools  
The way to dusty death. Out, out, brief candle!  
Life's but a walking shadow, a poor player,  
That struts and frets his hour upon the stage,  
And then is heard no more. It is a tale  
Told by an idiot, full of sound and fury,  
Signifying nothing.

sequence of states. To practice a soliloquy, one must be conscious of what one is doing. Since the zombie is *stipulated* not to have any conscious experience, the zombie cannot practice anything (silently or aloud; it can only make noises that sound like practicing). So, in order to maintain the neutrality between Jill and a zombie, the heterophenomenologist could not judge Jill's report that she was practicing silently on the bus to be true.

Since the theorist cannot recognize Jill's silently practicing as a real phenomenon, the theorist would also have to take Jill to be in error when she cited practice as the explanation of her flawless performance. But then we are left without the obvious explanation for Jill's flawless performance in front of the class. Confining attention to Jill's report, as heterophenomenology requires, misses the explanatory phenomenon itself. It is her *practice* of the soliloquy, whether silently or aloud, that accounts for her flawless performance in front of the class, not her *report* or *belief* that she was reciting the soliloquy.

It is simply implausible to deny that Jill's silently practicing on the bus explains her flawless performance in front of the class. If Jill had practiced out loud in her room, there would be no question that the practice explained the successful performance; the fact that she practiced silently could not change the explanatory value of practicing.

Moreover, practicing Macbeth’s soliloquy to oneself—in her head, so to speak—is clearly an action. It is done for a reason: she wants to perform well in front of the class, and believes that practice will enable her to do it. Practicing the soliloquy also has effects: Jill certainly performs better than she would have if she had not practiced the soliloquy. The expression ‘reciting Macbeth’s soliloquy’ is clearly an intentional action-description. Reciting Macbeth’s soliloquy out loud is obviously an action. If reciting Macbeth’s soliloquy out loud is an action, then so is reciting it to oneself. Out loud or to oneself, it is the same kind of action—done for the same reason, and with the same effects. It seems to me that there’s no doubt that reciting Macbeth’s soliloquy is an action—whether out loud or to oneself. The action is reciting Macbeth’s soliloquy; silently is the way that Jill did it.<sup>19</sup>

[Someone may object that the zombie and Jill would exhibit the same behavior: they both recited the soliloquy flawlessly. So, their behavior should have the same explanation. Since, by stipulation, the zombie’s behavior is not explained by silently practicing, neither is Jill’s behavior explained by silently practicing. Both the zombie and Jill had the same (functionally individuated) psychological

---

<sup>19</sup> This is an expansion of an example from one of my Gifford lectures, “First-personal Knowledge,” in *The Nature and Limits of Human Understanding: The 2001 Gifford Lectures at the University of Glasgow*, Anthony J. Sanford, ed. (London: T&T Clark, 2003): 165-185.



states, and the sequence of those functional states explained both their behavior. So, in neither case need we advert to any putative subjective experience to explain the flawless performances of the soliloquy.<sup>20</sup>

This objection raises deep issues about the mind. Let me just make a few remarks. To begin with, I take this objection to challenge the zombie-assumption. Jill and the zombie do not actually exhibit the same behavior—any more than Descartes' dog, who feels no pain, exhibits the same behavior when kicked that Fido, who is sensitive to pain, exhibits when he is kicked. Behavior that we want to explain is not and should not be identified physically. Fido's painful wince is a different behavior from Descartes' dog's (indiscernible) audible emission. Similarly, the zombie's audible emission that sounds like a recitation of the soliloquy is not really a recitation of the soliloquy. If this is right, then the zombie assumption should not be construed to imply that zombies and people are alike "on the inside". (Indeed, if the attempts to locate a neural correlate of consciousness succeed, we will have good reason to suppose that hypothetical zombies do not have brains that function just like ours.)

The conclusion is that if zombies exhibit behavior physically indiscernible from ours, then their brains must be organized rather differently from ours. We people are

---

<sup>20</sup> Louise Antony raised an objection similar to this.

motivated by feelings of pleasure and pain; zombies, by stipulation, are not. Necessarily, the pain of humiliation is conscious. There may be many disparate brain states that underlie the pain of humiliation; our psychology will group these together. Since, for zombies, there is no pain of humiliation, zombie psychology would not place these disparate brain states in the same category. Granted, if Jill tries to avoid the pain of humiliation, zombie-Jill exhibits similar behavior (typed physically): but we do not type behavior physically. Otherwise, we could never explain much of what we want to explain—e.g, “Why did she say that?” The correct answer cannot be given in physical terms. So, zombie-Jill and Jill do not have the same internal organization of brains; nor do they exhibit the same behavior.]

In sum, the obvious (and in my opinion, correct) explanation of Jill’s flawless performance is in terms of her practicing. But the explanation of Jill’s performance in terms of her silent practice cannot be allowed by heterophenomenology, because it entails that Jill has conscious experiences—thereby violating the Neutrality Assumption. Despite the fact that the silent practice was intersubjectively inaccessible, it was the silent practice itself, and not Jill’s report of it, that explained Jill’s flawless performance. I take the case of the silent practicing to be a decisive counterexample to Dennett’s claim that

heterophenomenology can “answer all the questions—without philosophically significant residue.”

Someone may object that the silent practice would be intersubjectively available via a “cerebroscope” that “reads” brains. But a cerebroscope could not confirm the silent practice. According to the Neutrality Assumption, a cerebroscope would yield the same results for a nonconscious zombie as for Jill. But since practice requires conscious effort and a zombie has no conscious experience at all, the zombie does not silently practice. So, given the Neutrality Assumption, whatever results a cerebroscope yields, they could not be interpreted to be silently practicing the soliloquy—since practice (silent or otherwise) requires conscious effort.<sup>21</sup> So, the objection misfires.

The objector may persist: “Zombie-Jill would have recited the soliloquy flawlessly too. And her perfect recitation would have had an explanation in terms of, say, functional states. If the functional-states explanation explains Zombie-Jill’s performance, they explain Jill’s performance; in that case, we do not need to appeal to Jill’s conscious silent practice at all.” I have several replies: In the first place, Zombie-Jill is an imposter. She looks and acts like us, but is not one of us. We are aware of things;

---

<sup>21</sup> Moreover, although I cannot argue for the point here, I doubt that it would ever be reasonable to interpret any particular neural events as the soliloquy that begins “Tomorrow and tomorrow and tomorrow.”

she is not. She only behaves as if she has feelings. She's like Descartes' dog, who does not feel pain, but behaves as if he did. If we had all been zombies, I believe, the background conditions for their to be soliloquys or recitations at all could never have been fulfilled. That is, there never could have been anything describable as 'a flawless performance' of anything.

In the second place, we people are motivated by pleasure and pain, feelings that we seek or avoid. Pleasure and pain cannot motivate Zombie-Jill, because she feels no pleasure or pain. She just behaves as if she does feel pleasure and pain. But Zombie-Jill feels no pleasure. So, Zombie-Jill is is not just like Jill psychologically: Zombie-Jill cannot be motivated by pleasure and pain, or by depression or by a sense of duty. So, Zombie-Jill will have no conscious motivation at all. Hence, Zombie-Jill's behavior in general will be explained differently from Jill's behavior, which will often be explained by Jill's conscious motivations. In that case, it is invalid to take explanations of Zombie-Jill's behavior in terms of her inner workings and apply them to Jill, whose inner workings support her conscious practice of the soliloquy. (The moral is that, assuming materialism, if Zombies are behaviorally indistinguishable from us, their inner workings must be different from ours. If materialism is true, and Zombies were physical duplicates of us, then they would have conscious states.)

So, the point at issue stands: Jill's conscious silent practice explains her flawless recitation of Macbeth's soliloquy. Hence, we must admit subjective first-personal experiences to be genuine phenomena.

### **Nonsubjective First-personal Phenomena Reclaimed**

Now I want to turn to nonsubjective first-personal phenomena—first-personal phenomena that are not described in terms of consciousness of our own minds. I'll give three more examples of different kinds of first-personal phenomena that resist third-personal treatment. These examples are not related to the controversies about theory of consciousness. They aim to show that—however we understand subjective experience—a complete third-personal description of the world would not be exhaustive. In each example, a person who was scientifically omniscient is still missing some knowledge. These examples are directed at any kind of third-personal absolutism, not just at heterophenomenology. They indicate that the range of irreducibly first-personal phenomena is much greater than just subjective awareness of our own minds.

1. The first example of nonsubjective first-personal phenomena that I want to consider is John Perry's famous case of the messy shopper.<sup>22</sup> Suppose that JP is shopping in

---

<sup>22</sup> John Perry, "The Problem of the Essential Indexical," *Noûs* 13 (1979): 3-21.

a supermarket, and he notices that somebody has a torn bag of sugar that's leaking. He follows the trail to inform the messy shopper that he or she is making a mess. He forms various beliefs—e.g., that the messy shopper isn't paying attention to his cart. He catches a glimpse of the messy shopper in the security mirror and forms the belief that the person he sees in the security mirror is the messy shopper. As he keeps going around and the trail gets thicker, it finally it dawns on him that he (himself) is the messy shopper. Until then, he had beliefs about the messy shopper (that is, about himself), but he didn't realize that he himself was the one that they were about. When he came to believe that he (himself) was the messy shopper, his behavior changed: He stopped and checked his cart and, sure enough, there was the leaking sugar. When he had the beliefs about himself, without realizing that they were about himself, he kept pushing the cart. When he realized that he himself was the messy shopper, he stopped—his change in behavior was explained by his change of belief.

Let us define the locution 'one knowingly refers to oneself' as 'one refers to oneself in the first-personal and realizes that it is oneself that one is referring to.' So, although Perry referred to himself by means of the description 'the messy shopper,' he didn't realize that it was himself that he was referring to. He referred to himself unknowingly. It is only when he could say from a first-

personal point of view, “I am the messy shopper” that he knowingly referred to himself. And the change in Perry’s behavior is explained by the change in his belief from unknowingly referring to himself (in the third-person) to knowingly referring to himself (in the first-person).

No acquisition of any belief in which Perry didn’t know that he was referring to himself would have led to a change in his behavior. It was his realization, “Oh! It’s me! I’m the one making the mess!” that led him to look in his cart. Since Perry’s first-personal belief that included his knowingly referring to himself led to his stopping and looking in his cart—and no third-personal belief about himself could have led to his change of behavior—it follows that the first-personal belief that included knowingly referring to himself (i) is explanatory, and (ii) is irreducible to any third-personal belief about himself. Hence, Perry’s coming to believe that he (himself) was the messy shopper is a nonsubjective irreducible first-personal phenomenon.<sup>23</sup>

[Consider for a moment how a heterophenomenologist might treat this case. The theorist would interpret the audio tape and have as an intersubjective datum, “Perry came to believe that he (himself) was making a mess.” For

---

<sup>23</sup> Another way to see this is to notice that the truth-condition for the sentence ‘Perry believed that he (himself) was the messy shopper’ entails what Perry would express by saying “I was the messy shopper.” The truth-condition for ‘Perry believed that he (himself) was the messy shopper’ thus entails that there be a first-personal reference.

this to be true, Perry had to knowingly refer to himself. Since the heterophenomenological *theorist's* sentence (not just the subject's sentence) would be false if the subject had not made a first-personal reference, there had to be a knowing first-personal reference. So, even if heterophenomenology were otherwise acceptable, it could not provide a completely third-personal account. The sentences of the interpreted text would not be true unless there really were first-personal references in the world. Therefore, earlier counterexamples aside, heterophenomenology does not rid the world of first-personal phenomena—unless (per impossibile) it can also construe first-personal reference in third-personal terms. And the point of Perry's example of the messy shopper is precisely to show that first-personal reference is essential to our explanations of behavior. No third-personal terms can explain the messy shopper's change of behavior. So, Dennett's heterophenomenology itself must countenance first-personal reference as a genuine first-personal phenomenon in the world.]

If the first-personal way of thinking of myself is essential to the belief that explains my change of behavior, and if science is, as we are assuming, exclusively third-personal, then the change of belief that explains my change of behavior is invisible to science.



2. The second kind of nonsubjective first-personal phenomenon is that I know who I am. I am LB. You know who you are too. Suppose that you are Sam. (Suppose that 'Sam' is your name. You might have been named something else and still have been the same person; let us suppose that you, in fact, were named 'Sam'.) Now suppose, Sam, that you were scientifically omniscient and had a complete third-personal scientific description of the world; your description includes all the people, one of whom is you. But from your scientific description, you would not know which person you were. You would know everything about Sam, even that he was named 'Sam'—as you would know about Jack and Jill and everyone else. You would know every detail about Sam's health and physical condition, whether Sam was married, what job Sam had, whether he was in debt, whether he enjoyed drinking beer—everything that could be known about yourself in the third-personal. But you would not know that *you* were Sam. From all your knowledge about Sam, you would be in no position to say: "I am Sam, and I am in good health; I am not married but I have a good job and I'm not in debt. I do enjoy drinking beer." It might even occur to you that all the scientific knowledge in the world would not compensate for not knowing which person you were.<sup>24</sup>

---

<sup>24</sup> Baker, "First-personal Knowledge," 184-5; cf. Thomas Nagel, *The View from Nowhere*, (New York: Oxford University Press, 1986): 54-5.

So, from a third-personal point of view, you would not know that you were Sam. If knowing who you are is something to be known, then reality is not completely describable in the third-person. This example raises more than an epistemological question of your knowing which person you are. It also raises the question of whether there is even a significant fact of the matter that you are Sam or that I am LB.

3. This brings us to the third example. For each of us, there seems to be a nontrivial fact of the matter about which person one is. But what kind of fact could it be that I am LB or that you are Sam? Consider the semantics of utterances of 'I am Sam.' 'I am Sam' is true iff it is uttered by Sam.

Suppose that the police are looking for Sam, and that you, Sam, are playing cards with your friends. The police come in and say, "We're looking for Sam; which one is Sam? Come on, confess: which one is Sam?" And you say meekly, "I am Sam." The police then take you to the police station for questioning; they leave your card-playing friends with their cards. In this case, your saying, 'I am Sam' expressed a proposition that had a certain effect on the police: they took you and left your friends. So, your saying, 'I am Sam' conveyed information to the police.

On standard semantic theories that hold that all

propositions are expressible in the third-personal, the proposition expressed by ‘I am Sam’ when uttered by Sam is ‘Sam is Sam’—a tautology; and the proposition expressed by ‘I am Sam’ when uttered by me is also expressed by ‘LB is Sam’—a contradiction.<sup>25</sup> So, standard semantic theories get rid of the first-personal reference, but the problem remains: ‘I am Sam’ when uttered by Sam seems to convey information—it tells the police which person to take to the station; but a tautology doesn’t convey information. Indeed, the suspect confessed to being Sam. How could a confession be a tautology? So, if there is a fact that you are Sam or that I am LB—and there surely seems to be—it is not well understood.

To summarize the argument for irreducible first-personal phenomena: Heterophenomenology is a valiant effort to show that a scientific account of reality can forego reference to conscious experiences in favor of reference to beliefs about conscious experiences. We saw that that didn’t work; but even if it had worked, there are irreducible first-personal phenomena beyond subjective consciousness. I presented three examples of first-personal phenomena that would be invisible to a completed third-personal science: first, beliefs about oneself as oneself that explain behavior (the messy shopper case); second, knowledge of

---

<sup>25</sup> This is so both on Millian theories that take the semantic role of ‘I’ to be exhausted by picking out a person; and on more complex indexical theories like Kaplan’s that add “character” to “content.”

which person one is (I know that I'm not Sam); third, the fact that one is a particular person (I am LB). None of these first-personal phenomena is expressible in third-personal terms and all of them have real consequences. Anything that has real consequences is real. So, if all science is third-personal, and if science explains all phenomena, then it seems that science must be incomplete as an account of all phenomena. It leaves out the first-personal parts of reality.

### **Return to the Master Argument**

I have just argued that the conclusion of the Master Argument is false: Not all phenomena can be described and explained from a third-personal perspective. If this is right, one of the premises must be false. Which premise should be abandoned? The premise that all phenomena can be described and explained by science, or the premise that all science must be constructed from a third-personal perspective? As I said at the beginning, I don't have a definitive answer. But both premises can be questioned. I'll consider them in reverse order.

Reject Second Premise: All science is constructed from a third-personal perspective?

David Chalmers has been in the forefront of suggesting that there can be a first-person science—a science that takes consciousness or experience as a fundamental feature

of the world, like electromagnetic force, and hence rejects the second premise that all science is from a third-personal point of view. He notes that it “has become possible to think that we are moving toward a genuine scientific understanding of conscious experience.”<sup>26</sup>

Chalmers would agree with me that the third-personal data cannot explain the first-personal data. Indeed, he recognizes two classes of data: third-personal data about brains and behavior, and first-personal data about subjective experience.<sup>27</sup> First-personal data include: occurrent thought, emotional experience, mental imagery, bodily experiences, visual experience, other perceptual experiences. He acknowledges that first-personal data, not being intersubjectively accessible, pose special problems, but he does not believe that they are insurmountable. We can assume that the subjects really have conscious experience and that their verbal reports reflect it. This is just the kind of background assumption made by physics—that the external world exists and that perception reflects the state of it.<sup>28</sup> Then, a theory of consciousness would

---

<sup>26</sup> David J. Chalmers, “How Can We Construct a Science of Consciousness?” <http://consc.net/papers/scicon.html> (2004), accessed 1/23/08, first paragraph.

<sup>27</sup> David J. Chalmers, “How Can We Construct a Science of Consciousness?” <http://consc.net/papers/scicon.html> (2004), accessed 1/23/08.

<sup>28</sup> “Just as a scientist can accumulate third-personal data by accepting reports of third-personal data gather by others (rather than simply treating those reports as noises), a scientist can also gather first-personal data by accepting reports of first-personal data

articulate systematic connections between first- and third-personal data.<sup>29</sup>

My own opinion, for what it's worth, is this: Whether there will be a science of consciousness of the sort that Chalmers envisages will depend on whether we require that all scientific data be equally intersubjectively available. It remains to be seen how far the idea of a first-personal science of consciousness can be worked out. However, even if Chalmer's program were to be fully successful, it would not handle the first-personal phenomena that are not described in terms of subjective experience. So, turn to the other alternative.

Reject First Premise: All phenomena can be described and explained by science.

The only argument that I know of in favor of this premise is an inductive argument from history. Crudely, it is

---

gathered by others....If there is positive reason to believe that a subject's report might be unreliable, then a researcher will suspend judgment about it. But in the absence of any such reason, researchers will take a subject's report of a conscious experience as good reason to believe that they are having a conscious experience of the sort that they are reporting" Chalmers, "How Can We Construct a Science of Consciousness?" p. 18. Other problems that Chalmers mentions are primitive methodology for gathering first-personal data and absence of formalisms to express first-personal data and construct a theory. But these latter problems do not seem insurmountable as work continues on a first-personal science.

<sup>29</sup> A theory of consciousness would ultimately be a theory of "systematic bridging principles that underlie and explain the covariation between third-personal data and first-personal data." Chalmers, "How Can We Construct a Science of Consciousness?"

this: “Over the centuries, the sciences have brought more and more phenomena into their domains. There is no stopping place. So, they will not stop until they have brought all phenomena into their domains.”

Well, how should we respond to such an argument? Perhaps, in some way that we cannot envisage now, all first-personal phenomena will be explained in some way that we will then count as scientific. Maybe, maybe not. But the prudent thing is to wait and see. To accept the claim that all phenomena can be described and explained by science is to accept a closure principle—“...and that’s all there is, folks!” The inductive argument from history seems to me awfully weak to support such a closure principle. I believe that whether we reject the second premise or not, we should reject the first premise. At best, we should just wait and see what phenomena resist integration into science.

### **Conclusion**

I have given five examples of different kinds of irreducible first-personal phenomena. Two are of the subjective variety—Jack’s losing and regaining consciousness; Jill’s silent practice of the soliloquy. Three are of the nonsubjective variety—the explanatory first-personal belief of the messy shopper; the knowledge of which person one is; the fact of the matter that I am LB.

My overall conclusion is that not all phenomena can be described and explained from a third-personal point of view. There are genuine irreducible first-personal phenomena. I leave open the question of how to understand these irreducible first-personal phenomena, and which of the premises of the Master Argument should be rejected. I claim only that a third-personal scientific view of the world must be an incomplete picture of reality.<sup>30</sup>

---

<sup>30</sup> I am grateful to Gareth B. Matthews and the Five-College Faculty Seminar on the Foundations of Cognitive Science for commenting on a draft of this paper.