

In *American Philosophical Quarterly* 48.2 (2011): 161-173. (University of Illinois Press)

Does Naturalism Rest on a Mistake?

Lynne Rudder Baker

University of Massachusetts Amherst

Naturalism has been challenged by philosophers of many persuasions—especially by those who doubt that intentionality or normativity can be understood in naturalistically-acceptable terms. Here I want to mount a different challenge: a challenge from first-personal phenomena. So, the answer to the question—Does naturalism rest on a mistake?—is, I shall argue, yes; and furthermore, it is a mistake that Wittgenstein would never have made.

I am going to use the term ‘naturalism’ in what I take to be the orthodox way endorsed by Quine. For purposes here, naturalism is the thesis that the only correct conception of reality is provided by science. Or as Sellars’ put it, “[I]n the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, of what is not that it is not.”¹

More precisely, I’ll take naturalism to be the ontological view that science can provide a supervenience base that accounts for all reality: that is, given the items in the scientific supervenience base—whatever they are—the existence of everything else that there is (objects, states of affairs, or whatever it takes for true beliefs to be true) is guaranteed. On the common assumption that science is expressible exclusively in third-personal terms, the supervenience base that science provides will likewise be expressed exclusively in third-personal terms. So, if naturalism is true, then first-personal phenomena will have to be reduced to third-personal phenomena or eliminated altogether.² To the contrary, I shall

suggest, a wholly third-personal account of reality either is incomplete or it covertly contains some first-personal phenomena.

Here is my plan: After briefly sketching my own view of first-personal phenomena, I'll give three very different examples to suggest the inadequacy of the third-personal point of view to capture all of reality—David Lewis on *de se* belief, Daniel Dennett on consciousness, and Thomas Metzinger on the self-model theory of subjectivity. Then, I'll draw some conclusions about the mistake on which naturalism rests, and say why I think that Wittgenstein never would have made it.

First, a caution: My concern is metaphysical—not semantic or conceptual. According to naturalism, *reality* can be exhaustively described in the third person. If so, then first-person phenomena are eliminable or reducible to phenomena described wholly in third-personal terms.

First-Personal Phenomena

On my view, human persons are essentially embodied, and embedded in environments; they are constituted by bodies to which they are not identical. Nevertheless, as I have argued elsewhere, human persons have first-personal persistence conditions.³

It is important to see the extent of first-personal phenomena. They are not just subjective phenomena like smelling an acrid odor or hearing a high-pitched whistle—not just phenomena that are allegedly logically private, or phenomena about which a single person is thought to be infallible. First-personal phenomena also include nonsubjective phenomena like my gratitude that I have close friends, or my wish that I were a movie star.

First-personal phenomena should not be thought of in a Cartesian way. They are

rather phenomena that are accessible only if there are first-person perspectives, and many such phenomena—like a wish that one were a movie star—are inextricably tied to what Descartes would call the “external world.” (One could not wish that one were a movie star unless one had the concept of being a movie star, and there would be no concept of being a movie star except in a society in which there are practices of making movies.) So, the first-person/third-person distinction is not a subjective/objective distinction.

First-personal phenomena are enjoyed by entities with first-person perspectives. Many nonhuman mammals and all human infants have rudimentary first-person perspectives. A rudimentary first-person perspective requires consciousness, intentionality, and the ability to imitate. Nonhuman primates and other higher animals are conscious, and they have psychological states like believing, fearing and desiring. They have points of view (e.g., “danger in that direction”), but there is no evidence that they can conceive of themselves as the subjects of such thoughts. They cannot *conceive of* themselves from the first-person.

By contrast, we adult human beings with language have *robust* first-person perspectives: we can conceive of ourselves *as* ourselves from the first-person. To conceive of oneself as oneself is to think of oneself, from the inside, as it were. A robust first-person perspective is not a qualitative property, but rather a formal property: the property of being able to think of oneself without any third-person referential device like a name, description or demonstrative.

Linguistic evidence of a robust first-person perspective comes from use of first-person pronouns embedded in sentences whose main verbs are linguistic or psychological verbs—e.g., “I protested that I has overcharged,” or “I wonder how I will die.” If I protest that I was overcharged or I wonder how I will die, then I am thinking of myself as myself; I am not

thinking of myself in any third-person way (e.g., not as Lynne Baker, nor as the person who is thinking a certain thought, nor as woman in the front of the room) at all.

If I say, “I am tall,” I *refer* to myself in the first-person. If I say, “I am glad that I am tall,” I *attribute* to myself first-person reference. Following Castañeda and Matthews, I’ll mark the second occurrence of ‘I’ with an asterisk; I’ll call any thought expressed by a first-person sentence with a psychological or linguistic main verb and an embedded clause with a first-person pronoun, an ‘I*-thought’. An I*-sentence expresses such an I*-thought.⁴ For example, ‘I’m glad that I* am tall’ and ‘I wish that I* were tall’ and ‘I told you that I* am tall’ are all ‘I*-sentences. In an ‘I*-sentence’, a speaker not only *refers* to herself (with the first occurrence of ‘I’), but also she *attributes* to herself a first-person reference (with the second occurrence of ‘I’). If you say, “I believe that I* am tall,” you attribute to yourself a first-person reference. If you say, “Obama believes that he* is tall, you attribute to Obama a first-person reference.

Note that neither the sentence ‘I believe that I* am tall’ (uttered by me) nor the sentence ‘Obama believes that he* is tall’ is an instance of ‘x believes that x is tall.’ ‘x believes that x is tall’ is made true by Obama’s believing that Obama is tall, whether Obama believes that he* is tall or not. ‘Obama believes that he* is tall’ is true only if Obama would express his belief in the first person: I am tall. Neither ‘he*’ nor ‘I*’ can be replaced by a variable. No sentence that contains an occurrence of ‘he*’ or ‘I*’ can be true in a wholly third-personal world lacking first-person reference.

There are several features of I*-thoughts worth noticing: First, as already mentioned, they are not limited to ‘Cartesian’ thoughts about what one is thinking; they include mundane thoughts like ‘I wish that I* were in the movies.’ Second, I*-thoughts need no recourse to

any peculiar object like a self, or a soul, or an ego. My I*-thoughts refer to me, a person--the same entity that you refer to by saying, 'Lynne Baker'. Third, I*-thoughts and I*-sentences cannot be replaced *salva veritate* by wholly third-personal thoughts or sentences.⁵ I may be glad that I* survived the storm, without being glad that Lynne Baker or the woman wearing a black jacket survived the storm—if I don't realize that I am Lynne Baker or wearing a black jacket. Again, no name, description or third-person referential device is substitutable *salva veritate* for 'I*' or, as the Obama example indicates, for 'he*'.

Let me suggest that this view of 'I' and 'I*' as referring expressions of a special kind need not conflict with Wittgenstein's views in the *Investigations*—pace Anscombe and Malcolm.⁶ In PI 405, Wittgenstein has his interlocutor say, "But at any rate when you say 'I am in pain', you want to draw the attention of others to a particular person," and Wittgenstein responds by saying, "The answer might be: No, I want to draw their attention to *myself*." (PI, 405) One way to draw attention to oneself is to refer to oneself from the unique first-person perspective. Moreover, simple sentences like 'I am in pain'—relevant in discussions of solipsism⁷—do not even hint at the vast territory of first-personal phenomena under discussion here.

My view of the robust first-person perspective has certain naturalistic elements. It does not appeal to immaterial souls, or selves, or other peculiar objects. It allows that robust first-person perspectives may well be a product of natural selection. But it has non-naturalistic elements as well: The I*-thoughts and I*-sentences that express a robust first-person perspective resist wholly third-personal treatment. Now to the cases:

David Lewis on De Se Belief

Although David Lewis had too much confidence in the a priori to be a mainstream naturalist, his analysis of de se knowledge and belief may seem to provide a wholly third-personal account of the objects of knowledge and belief. Lewis proposed a novel view of objects of attitudes that is to accommodate I*-beliefs and knowledge. On his view the objects of the attitudes are not propositions, but properties: To have an attitude is to self-ascribe a property. *All* knowledge and belief, on Lewis's view, is self-ascription of properties—and hence, is broadly de se.⁸

Lewis, who calls himself “a robust realist about beliefs and desires,”⁹ very neatly distinguishes self-belief expressed by I*-sentences from de re belief and de dicto belief. De dicto belief that p is self-ascription of the property of being in a world in which p. De re belief is ascription of a property X to an individual Y under a suitable description Z, where a suitable description either denotes an essence of Y or a relation of acquaintance with Y.¹⁰ Finally, I*-belief is a special case of de re belief, where the relation of acquaintance is identity. “Self-ascription of properties is ascription of properties to oneself under the relation of identity.”¹¹ Lewis says: de se belief “concerns not the world but oneself.” It is belief whereby I “self-ascribe the properties I think myself to possess.”¹²

So, if I say, “I believe that I* live in Amherst,” I self-ascribe the property of living in Amherst. (“The de se content of my belief that I have F is just the property F itself.”¹³) If I say, “I believe that you live in Amherst,” there is a relation of acquaintance (or a description that captures essence) that I uniquely bear to you, and I self-ascribe bearing that relation uniquely to someone who lives in Amherst. If I say, “I believe that many people live in Amherst,” I self-ascribe the property of inhabiting a world in which many people live in

Amherst. The first belief—the I*-belief is (narrowly) de se; the second, de re; and the third, de dicto.

My question is this: Does this account capture first-personal phenomena in non-first-personal terms? Not obviously so: All belief is self-ascription of properties, and self-ascription as Lewis is using the term is not just ascription to someone who happens to be oneself. To self-ascribe a property is to ascribe it to oneself as oneself, in the first-person. Self-ascription (of anything) requires that the self-ascriber realize that it is she* to whom she is ascribing the property. Oedipus may have self-ascribed the property of being in a world in which someone killed Laius at a crossroads. But Oedipus would not have self-ascribed the property of being the killer of Laius at a crossroads, if he did not realize that he* himself killed Laius at a crossroads. But self-ascription of *either* property—of being in a world in which someone killed Laius at a crossroads, and of being the killer of Laius—requires that the self-ascriber have a robust first-person perspective. Lewis’s whole account in terms of self-ascription presupposes that the believer has a first-person perspective.

This observation, though true, understates the role of the first-person perspective in Lewis’s account. The first-person perspective shows up, not only in all self-ascribing of properties, but also in some of the properties that are the *objects* of belief and knowledge. Suppose that Smith is trying to recall how she felt when she* won the booby prize. She says, “I believe that I* was embarrassed (and not amused) that I* won the booby prize.” How would Lewis construe Smith’s remark? Considering his other cases, I think that he should say this:

Smith self-ascribed the property of being embarrassed that she* won the booby prize.

So, in addition to the first-person perspective embedded in any self-ascription of any

property, the particular property that Smith self-ascribed—the property of being embarrassed that she* won the booby prize—is a first-personal property that is the object of Smith’s knowledge, a property that itself involves attribution of a first-person reference. That is, first-person reference is entailed by the *object* of knowledge attributed to Smith. And if the attribution is correct, then Smith made a first-person reference that we have no idea how to paraphrase in third-personal terms.

So, Lewis’s analysis has not eliminated the first-person perspective from the *objects* of belief and knowledge, and objects of knowledge belong in a full account of reality. In order to reduce the I*- property to make it naturalistically suitable, we would have to give a third-personal account of the distinction between Smith’s being embarrassed that she* was F (with its first-person implication) and Smith’s being embarrassed that Smith was F in wholly third-personal terms.¹⁴ I am dubious that this can be done. In any case, as it stands, Lewis’s analysis of de se belief gives no comfort to the naturalist.

Daniel Dennett on Consciousness

Daniel Dennett explicitly aims to construct a model of consciousness “from the third-personal point of view since all science is constructed from that perspective.”¹⁵ His strategy is to show how a theorist can describe a subject’s mind from the subject’s own point of view.

Dennett’s account has two phases. The first phase is what he calls ‘heterophenomenological’—a theorist’s third-personal view of the way a subject sees herself. The second phase is a neuroscientist’s account of what is really going on in the subject’s brain. The result, interpreted intentionally, is Dennett’s ‘Multiple Drafts model’ of consciousness.

The first-phase theorist, the heterophenomenologist, prepares data given by a subject who reports her subjective experiences. The raw data are the uninterpreted output of 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. I'll focus only on the sounds the subject emits, as Dennett does. The heterophenomenologist puts the data through several processes of interpretation—first by interpreting the noises from the tape as sentences,¹⁶ then by interpreting the sentences as speech acts. The result is a text of the world from the subject's own point of view, like the world according to Garp, or Sherlock Holmes's London—all without giving up science.¹⁷

These processes of interpretation are to be understood in terms of Dennett's intentional-stance theory of content. The value of intentional-stance theory is that it provides a way to use intentional language like 'Jill believes that she heard a siren' without commitment to any particular state of affairs in reality; the intentional language is only a tool for prediction. Intentional-stance theory 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."¹⁸ But the theory is indifferent to "the internal structures that accomplish the rational competences."¹⁹

The data to be explained, then, are a theorist's interpretations of the subject's verbal expressions of belief. The subject's expressions and reports of conscious events are never taken to *refer* to conscious events (like hearing a siren), but are taken only as data concerning what the subject thinks. ("Pain-talk," for example, "is non-referential.")²⁰ 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.).”²¹

Heterophenomenology, Dennett says, “maintains a nice neutrality: it characterizes [subjects’] beliefs, their heterophenomenological worlds, without passing judgment, and then investigates to see what could explain *the existence of those beliefs*.”²² The point of heterophenomenology is to purge the data of any reference to conscious events, and hence to purge the data of any commitments to any first-personal phenomena. This process is to render the data suitable for explanation by physical theory.

The Multiple Drafts model of consciousness then tells us what is really going on in the brain that explains the subject’s beliefs about private subjective experiences. What is really going on in the brain can also be interpreted from the intentional-stance.²³ From the intentional stance, we can say of the continual stream of processing that the brain is constantly editing and creating content. Hence, the title ‘Multiple Drafts model’.²⁴

As I mentioned, the theory of consciousness must not assume that the subject’s beliefs are true, but only that she has them. But this leads to a problem in Dennett’s account. Suppose that a subject says, “I’m seeing a continuously moving dot.” The datum would have to be “She believes (or reports) that she* is seeing a continuously moving dot.”²⁵ Then, it is the task of neuroscience to explain what is going on in her brain when she reports that she* sees a continuously moving dot. Although the neural evidence may render it doubtful or false that the subject actually sees a continuously moving dot, from the point of view of Dennett’s theory, it remains the case that she *reports* that she* sees a continuously moving dot. That’s the datum to be explained by the theory.

However, it may not be obvious, but that datum retains the commitment to a first-person perspective. The datum is ‘The subject reports that she* sees a continuously moving

dot.’ The ‘she*’ is the theorist’s expression of the subject’s attribution to herself of a first-person reference. That is, Dennett’s *theorist* must attribute to the subject a first-person reference. The theorist’s attribution is false unless the subject actually makes a first-person reference.

Consider a slightly different heterophenomenological datum, ‘The subject reports that she* saw a witch.’ The theorist who takes this to be a datum is not committed to the existence of witches, but the theorist *is* committed to the subject’s having made a first-person reference. If the subject did not really see a witch, then the subject was mistaken; but if she didn’t make a first-person reference, then there is no datum.²⁶ But if *the theorist* must suppose that the subject actually made a first-person reference, then first-personal phenomena are not eliminated from the world. The first-personal phenomenon may be concealed, but it is still there. Heterophenomenology does not succeed in purging the data of first-personal elements.

But this is not really Dennett’s last word. Although Dennett mentions that the intentional stance has been put to vigorous use by various sciences,²⁷ he takes his Multiple Drafts model to be a mere propaedeutic to a serious scientific study of the mind, for which he suggests “translat[ing] all the talk of personal access into subpersonal terms.”²⁸ Indeed, Dennett advises banishing the word ‘I’ from science that properly “treats the mind as a distributed computational system with no central controller.”²⁹

But a wholly third-person subpersonal approach cannot yield complete knowledge of reality inasmuch as the theory would render invisible the data—e.g., that the subject reported that she* saw a continuously moving dot, or a witch. Although such data do not entail the existence of moving dots or witches, they do entail the existence of first-personal phenomena. Hence, Dennett’s account of consciousness does not show how first-personal phenomena can

be replaced by wholly third-personal phenomena.

Thomas Metzinger on the Self-Model Theory of Subjectivity

Thomas Metzinger has the most detailed naturalistic account of the first-person perspective that I know of. He sees clearly that self-consciousness (expressed by I*-sentences) is importantly different from mere sentience, or the kind of consciousness that certain nonhuman animals have. Metzinger is one of the few cognitive scientists who really have a grip on what I take to be the data. He says:³⁰

As Baker points out, it is not only necessary to have thoughts that can be expressed using 'I'. What is necessary is the possession of a concept of oneself as the *thinker* of these thoughts, as the *owner* of a subjective point of view. In short, what is needed is not only reference from the first-person point of view, but the capacity to mentally 'ascribe' this act of reference *to oneself* while it is taking place.

Indeed, Metzinger goes on, "The difference between merely having a perspective and conceiving oneself as having a perspective (as indicated by 'I*') is important for cognitive science in general, and also for the philosophical notion of a true cognitive subject."

However, Metzinger disagrees with me about the ineliminability of the first-person perspective. He is an eliminativist about the first-person perspective.³¹ His view purports to eliminate reference to any genuine subject of experience in favor of subpersonal phenomena. (Whereas I speak of persons and genuine subjects of experience, Metzinger speaks of selves.)

In his book, *Being No One: The Self-Model Theory of Subjectivity*, Metzinger develops a sub-personal account of first-personal phenomena of the sort that Dennett approves.³² Metzinger proposes an ontology—an inventory of what really exists—that

includes no selves, no first-personal subjects of experience, but only information-processing systems that are understandable in wholly third-personal terms. Elsewhere, I have a detailed critique of Metzinger's view.³³ Since the view is very complicated, here I'll give only a crude sketch and then offer a comment.

On Metzinger's theory, what we really are are information-processing systems that activate mental models—both models that represent the world and self-models that represent the system generating the models. A “self-model is a model *of* the very representational system that is currently activating it within itself.”³⁴ The content of a phenomenal self-model (PSM) “is the conscious self: your bodily sensations, your present emotional situation, plus all the contents of your phenomenally experienced cognitive processes.”³⁵

It is important to see how Metzinger uses the term ‘phenomenal’. Phenomenal experience supervenes on brain states and is characterized by how it feels or “what it's like” to have it. This leaves it open whether or not a phenomenal content represents anything real, or, as Metzinger puts it, whether or not a phenomenal content is “epistemically justified”. Phenomenal content may or may not depict something in reality.

Subjective experience is phenomenal experience. It consists of activation of models of representations.³⁶ The phenomenal self-models are transparent, in the sense that in G.E. Moore's example of the sensation of blue, the sensation is transparent: we look through it and see only the blue. Similarly, phenomenal self-models are transparent: they are not experienced themselves; they are not “introspectively accessible”. Only the content properties of representations are experienced, whether the contents depict anything outside the model or not.³⁷ If subjective experience is only a matter of activation of transparent phenomenal self-models, there is no place for a “self” or a genuine subject of experience.

So, how can it seem to us that we are subjects of experience in the world, and not just elements in models? We think of ourselves as subjects of experience because we can make “cognitive self-reference”—a “very special way of higher-order self-modeling.”³⁸ We can think of “ourselves as ourselves” (i.e. , we can have I*-thoughts).³⁹ I*-thoughts, according to Metzinger, require integrating part of an opaque self-model into a preexisting transparent self-model---thus splitting the PSM into an opaque and a transparent partition.⁴⁰ The opaque self-model is a phenomenal model of the intentionality relation (PMIR) that “represents itself in an ongoing, episodic *subject-object relation*.”⁴¹

The first occurrence of ‘I’ in my thought, ‘I believe that I* am in England’ refers to a representation in my opaque self-model that represents itself as the thinker of this thought and that is integrated into my transparent self-model. Its activation manifests the subject-object relation between the representation of the thinker of this thought and the representation of myself in the transparent self-model—the representation that is introspectively inaccessible to me. The second occurrence of ‘I’ in ‘I believe that I* am in England’—the ‘I*’—refers to the self-representation in the transparent self-model. In ‘I* thoughts, there are two self-representations: the introspectively accessible ‘I’ as the thinker of this thought and the introspectively inaccessible ‘I*’. Hence, we are unable to consciously experience that “we are referring to the content of a *representation* that is ‘in ourselves’ (in terms of locally supervening on brain properties).”⁴²

That is, I* thoughts are a matter of an information-processing system that generates subjective experiences that include the experience of being a subject of experience. But the subject of experience remains phenomenal: the reference to a subject of experience is reference to an element in the self-model and, as we shall see, not to a subject existing in the

world outside the self-model.

So: Is the conscious cognitive subject *only* phenomenal? Metzinger says yes: The conscious cognitive subject is only a representation in a self-model. A cognitive first-person perspective (i.e., the ability to have I*-thoughts) is a special case of a phenomenal first-person perspective. As Metzinger puts it: “Cognitive self-reference is a process of phenomenally modeling certain aspects of the content of a preexisting transparent self-model, which in turn can be interpreted as the capacity of conceiving of oneself as oneself*.”⁴³ In cognitive self-reference, what is referred to is the phenomenal content of a transparent self-model.⁴⁴ So, the reference will be to an element of the self-model, not to a subject existing in the world. ‘I’ and ‘I*’ refer to nothing outside the self-model. In short, the conscious cognitive subject is just an element of the self-model.

The subjective experience of being someone in the world is an illusion. Just as dreams and hallucinations tell us nothing veridical about what’s really going on in the environment, so too does subjective experience tell us nothing veridical about what we are. There are no selves, just self-models. “For ontological purposes,” Metzinger says, “‘self’ can therefore be substituted by ‘PSM’ [phenomenal self-model].”⁴⁵

I want to underscore just one feature of Metzinger’s view: Metzinger’s account of first-personal phenomena is entirely at a subpersonal level. I believe that any subpersonal account of the first-person perspective courts paradox. Consider some consequences of Metzinger’s subpersonal account of the first-person.

In the first place, Metzinger says, paradoxically, “[Y]ou constantly confuse yourself with the content of the self-model currently activated by your brain.”⁴⁶ I ask: Who is doing the confusing? On the last page of his book, Metzinger says that we should not take this

metaphor too literally: “There is no one *whose* illusion the conscious self could be, no one *who* is confusing herself with anything.”⁴⁷ It is difficult to see how there can be a confusion without there being anyone who is confused.⁴⁸

In the second place, consider an epistemic consequence of Metzinger’s view. The theory cannot make sense of one’s knowing what she* is doing while she is doing it. Suppose that a teenager, on a bet with her older brother who drives fast, is accelerating a car until she hits 100 mph. Accustomed to riding with her fast-driving brother, she thinks excitedly to herself, “I can hardly believe that *I* am driving so fast!”

Our teenager supposes that she*, the subject of this thought, exists in reality. On Metzinger’s view, she is deceived. What is really going on on Metzinger’s view concerns activation of representations in self-models, without any reference to anything (a self or subject) outside the models.⁴⁹ But our teenager cannot see her own thoughts and activity in this light. Whereas she takes herself to be a subject of experience in the world, what is really happening is just the activation of part of a self-model. Metzinger’s theory would seem to make it impossible for anyone to think clearly about what she is doing while she is doing it. (Indeed, if our teenager driving 100 mph were to consider what was really going on according to Metzinger, she would probably have a wreck.) A view of subjectivity that makes it impossible for people to think clearly about what they are doing as they are doing it is dubious.

In the third place, it seems that Metzinger’s theory cannot coherently be evaluated, much less endorsed or accepted. I may have the subjective experience that I* am evaluating Metzinger’s theory. I think to myself, “I am having the experience that I* am evaluating Metzinger’s theory.” But the “I*” doing the evaluating is not an entity in the world; it is just

part of the content of a transparent self-model.⁵⁰ When I take myself to be making a first-person reference, the referent is a mental representation. It is incoherent to suppose that a mental representation can actually evaluate a theory. On Metzinger's view, all there can be is a subjectless subjective experience of evaluating his theory. There is nobody to evaluate it. So, if Metzinger is right, then it seems that his theory cannot be evaluated at all. It is paradoxical, if not outright self-contradictory, to present a theory that cannot be coherently evaluated.⁵¹ These consequences alone give us reason to resist Metzinger's or anyone else's subpersonal account of the first-person perspective.⁵²

The Upshot

We have now considered three attempts to treat first-personal phenomena in a third-personal, naturalistic way: Lewis's conceptual analysis of *de se* belief, Dennett's interpretationist/eliminativist model of consciousness, and Metzinger's subpersonal/eliminativist approach toward the cognitive subject. None of them succeeds in either reducing or eliminating robust first-person perspectives. Of course, this does not disprove naturalism, but it should give us pause.

The robust first-person perspective plays a central role in many garden-variety things we do. Consider making a vow. Vows are typically made by I*-sentences: "I promise that I* will care for you in sickness and in health." Your life may depend on my keeping my vow, which could not have been made except in the first-person. Or consider regret, which is also expressed by I*-sentences: "I regret that I* caused you such pain." Again, this is a mundane example. But these everyday phenomena could not occur at all in the absence of robust first-person perspectives. We should not embrace a metaphysics that makes mundane but significant phenomena unintelligible.

Insofar as naturalism insists that reality is wholly third-personal, it does not offer a complete view of reality.

Wittgenstein and Naturalism

‘Naturalism,’ as everyone knows, means different things to different people. On some construals of naturalism—such as Jennifer Hornsby’s naive naturalism—Wittgenstein may well have been a naturalist. Of course, naturalism of any sort is a philosophical thesis, and Wittgenstein repeatedly denied that he was proposing philosophical theses. (See PI #109 and #128.) Still, we may decide that he was a naturalist—just we may decide that Spinoza was an atheist, although he prominently discussed God.

But, even if he had not been averse to philosophical theses, Wittgenstein would have been no orthodox naturalist. From the *Tractatus* on, he never looked to science to address philosophical problems. Even if he had been inclined to propose philosophical theses, he never would have repudiated or denigrated the first-person perspective. Indeed, he wrote from the first-person perspective throughout the *Investigations*. His famous private-language argument is *not* an attack on the first-person perspective; it is an attack on a particular conception of mind and language that we associate with the Cartesian tradition. If it is a mistake to suppose that all reality is third-personal, it is not a mistake that Wittgenstein would have made.

Nor would he have made the mistake (at least, I think it is a mistake) that Dennett and Metzinger made--of seeking explanations of first-personal phenomena at subpersonal levels. As Wittgenstein said in Zettel (#611), “[I]f one allows a causality between psychological phenomena which is not mediated physiologically, one thinks one is professing belief in a gaseous mental entity.” But no! It is “perfectly possible that certain psychological

phenomena *cannot* be investigated physiologically, because physiologically nothing corresponds to them.” (#609) Eliminativists may well agree here and conclude that the psychological phenomena are illusory. But that seems to me not an option: if the psychological phenomena are illusory, so are our investigative methods (like reading a meter or a gauge).

I think that it is reasonable to conclude both that Wittgenstein would never have tried to understand first-personal phenomena from a third-personal point of view, and that he would never have supposed that subjects of experience could be fully understood in terms of sub-personal phenomena. I think that these two negative claims (that I have put in Wittgenstein’s mouth) are good starting points for a rich metaphysics. And that’s the direction that I, unlike Wittgenstein, would take them.⁵³

¹ Wilfrid Sellars, “Empiricism and the Philosophy of Mind” in *Science, Perception and Reality* (London: Routledge and Kegan Paul, 1963): 173 (127-196). (Sellars took this slogan to be a less paradoxical way of saying: “[S]peaking as a philosopher, I am quite prepared to say that the common sense world of physical objects in Space and Time is unreal—that is, that there are no such things.”)

² This follows from construing naturalism wholly in terms of science. One could hold a kind of “supervenient evolutionary emergentism.” Beginning with a scientific supervenience base, new kinds of phenomena, objects, and properties emerge over time—some by natural processes and some by human ingenuity. The original scientific supervenience base is not by itself sufficient for the all the later emergent items, and hence the emergents are not reducible to the original base. (But emergent items need not be wholly inexplicable; they may have partial explanations in terms of emergent items previously emerged along with the original base.) Then, a full account of reality would include the original subvenient base together with the emergent items. Possible emergents include self-replicating molecules, qualia, the first-person perspective. If the first-person perspective is emergent, then it belongs in a full account of reality.

³ See my *Persons and Bodies: A Constitution View* (Cambridge: Cambridge University Press, 2000) and *The Metaphysics of Everyday Life: An Essay in Practical Realism* (Cambridge: Cambridge University Press, 2007). My arguments here for the irreducibility and noneliminability of first-personal phenomena do not presuppose my view of persons. I mention my view of human persons to assure the reader that I am not concealing an appeal to an immaterial soul or mind.

⁴ Hector-Neri Castañeda introduced ‘he*’, and Gareth B. Matthews extended the ‘he*’ from sentences with a third-person subject to ‘I*’ for sentences with a first-person subject. Castañeda studied phenomena expressed by sentences like ‘The editor believes that he* is F.’ See H-N Castañeda 1967 “Indicators and Quasi-Indicators,” *American Philosophical Quarterly* 4, 1967: 85-100, and H-N Castañeda “He: A Study in the Logic of Self-Consciousness,” *Ratio* 8, 1966: 130-157. Gareth B. Matthews extended the discussion to phenomena expressed by ‘I think that I* am F’. See G.B. Matthews, *Thought’s Ego in Augustine and Descartes* (Ithaca, NY: Cornell University Press, 1992).

⁵ I also believe, but cannot argue here, that one cannot have I*-thoughts without a natural language. I*-thoughts have a grammatical complexity found only in natural language. Hence, a robust first-person perspective has social and linguistic presuppositions. Understanding I*-thoughts as self-ascription of properties (à la David Lewis) does not invalidate this claim. A nonlinguistic creature cannot self-ascribe properties.

⁶ G.E.M. Anscombe, “The First Person,” in *Metaphysics and the Philosophy of Mind: Collected Papers Vol. II* (Minneapolis: University of Minnesota Press, 1981): 21-36, and Norman Malcolm, “Whether ‘I’ is a Referring Expression,” in *Wittgensteinian Themes: Essays 1978-1989* by Norman Malcolm, Georg Henrik von Wright, ed. (Ithaca, Cornell University Press, 1995): 16-26.

⁷ For a solipsist, ‘I am in pain’ is equivalent to ‘There is pain.’ But solipsism is self-defeating in that if I were the only thing in the universe, I would not have the resources to think or say that I was.

⁸ Lewis, “Attitudes De Dicto and De Se,” in *Philosophical Papers Vol 1*. (New York: Oxford University Press, 1983): 139. (133-156) Originally published in *The Philosophical Review* 88 (1979): 513-543.

⁹ David Lewis, “Reduction of Mind,” in *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press, 1999): 291-324. The quotation is on p. 322. I take this to imply that Lewis would not consider use of ‘I*’ a user-illusion.

¹⁰ Relations of acquaintance include relations to my acquaintances, present or absent; people prominent in the news; famous dead people; authors whose works I have read; strangers I am face to

face with; strangers I am tracing (such as the driver in the car in front of me); myself. David Lewis, "Attitudes De Dicto and De Se," in *Philosophical Papers Vol 1*. (New York: Oxford University Press, 1983): 155. (133-156) Originally published in *The Philosophical Review* 88 (1979): 513-543.

¹¹ Lewis, "Attitudes De Dicto and De Se," p. 156.

¹² Lewis, "Reduction of Mind," p. 317.

¹³ Lewis, "Reduction of Mind, p. 318.

¹⁴ For the same reason that worlds were not fine-grained enough to handle self-locating beliefs, worlds plus individuals (centered worlds) seem not fine-grained enough to handle beliefs whose objects are I*-properties. This suggests that I*-properties cannot simply be construed as self-locating properties, and that the theoretical apparatus of centered worlds does not suffice for the distinction.

¹⁵ Daniel Dennett, *Consciousness Explained* (Boston: Little, Brown and Co., 1991): 71.

¹⁶ 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' Dennett, *Consciousness Explained*, p. 75.

¹⁷ Dennett, *Consciousness Explained*, p.76. 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.

¹⁸ Dennett, *Consciousness Explained*, p. 76.

¹⁹ Dennett, "Intentional Systems Theory," in *The Oxford Handbook of Philosophy of Mind*, Brian P. McLaughlin, Ansgar Beckermann, Sven Walter, eds. (Oxford: Clarendon Press, 2009), p/ 346. (39-350)

²⁰ Dennett, *Content and Consciousness* (London: Routledge, 1993), p. 96. Reprinted from Routledge and Kegan Paul, 1969.

²¹ Dennett, "The Fantasy of First-person Science," <http://ase.tufts.edu/cogstud/papers/chalmersdeb3dft.htm> p. 4. Accessed Sept. 1, 2009.

²² Dennett, "The Fantasy of First-person Science," p. 3. And: 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." Dennett, *Consciousness Explained*, 72; "The Fantasy of a First-person Science," p. 2

²³ Indeed, Dennett says that intentional-systems theory is ubiquitous in computer science, animal psychology and evolutionary biology. Dennett, "Intentional-Systems Theory," p. 349.

²⁴ Any line drawn between what is conscious and what is preconscious is arbitrary. Indeed, "the fundamental implication of the Multiple Drafts model" is this: If "one wants to settle on some moment of processing in the brain as the moment of consciousness, this has to be arbitrary....[T]here are no functional differences that could motivate declaring all prior stages and revisions to be unconscious or preconscious adjustments, and all subsequent emendations to the content...to be postexperiential memory contamination. The distinction lapses in close quarters." (p. 126) "There is no reality of conscious experience independent of the effects of various vehicles of content on subsequent action (and hence, of course, on memory)." Dennett is willing to classify the Multiple Drafts model as "first-person operationalism, for it brusquely denies the possibility in principle of consciousness of a stimulus in the absence of the subject's belief in that consciousness." Indeed, "the Multiple Drafts model makes

‘writing it down’ in memory criterial for consciousness: that is *what it is* for the ‘given’ to be ‘taken’—to be taken one way rather than another.” Dennett, *Consciousness Explained*, p. 132.

²⁵ The datum could also be “She reports seeing a continuously moving dot,” which is equivalent to “She reports that she* sees a continuously moving dot.”

²⁶ Someone may object that the datum does not require that the subject make a first-person reference. She may simply report seeing a continuously-moving dot or seeing a witch. Then, we could consider a more complex case of the subject’s reporting being embarrassed that she* won the booby prize. The attribution of a first-person reference there is ineliminable.

²⁷ “I think the fact that the concept of the intentional stance has been put to such vigorous use (and some abuse) by the relevant sciences has...shown philosophers something about the role that we philosophers can play in the interdisciplinary quest to understand the mind.” Daniel C. Dennett, “Daniel Dennett” in *Mind and Consciousness: 5 Questions*, Patrick Grim, ed. (Automatic Press, VIP, 2009): 26. (25-30)

²⁸ Bryce Huebner and Daniel C. Dennett, “Banishing ‘I’ and ‘We’ from Accounts of Metacognition,” *Behavioral and Brain Sciences* 32 (2009): 148-9.

²⁹ Huebner and Dennett, p. 149.

³⁰ Thomas Metzinger, *Being No One: The Self-Model Theory of Subjectivity* (Cambridge, MA: MIT Press, 2003): 396.

³¹ Metzinger, *Being No One*, pp. 396-7, in response to Lynne Rudder Baker, “The First-Person Perspective: A Test for Naturalism,” *American Philosophical Quarterly*, 35, 1998: 327-348, p. 331.

³² Indeed, Dennett cites Metzinger approvingly for explaining “how subpersonal mechanisms can construct ‘the self’ that is represented by the personal stance.” Huebner and Dennett, p. 149.

³³ “Naturalism and the First-Person Perspective,” ” in *How Successful is Naturalism?* Georg Gasser, ed. (Frankfurt: Ontos Verlag, 2007): 203-226.

³⁴ Metzinger, *Being No One*, p. 302. A PSM is “not a thing, but an integrated process.” *Being No One*, p. 3.

³⁵ Metzinger, *Being No One*, p. 299.

³⁶ Metzinger (*Being No One*, p. 401) says: “The content of [I] is the thinker, currently representing herself as operating with mental representations.”

³⁷ Metzinger, *Being No One*, p. 387.

³⁸ Metzinger, *Being No One*, p. 599.

³⁹ “Phenomenal representational processes...generate conceptual forms of self-knowledge, by directing cognitive processes toward certain aspects of internal system states, the intentional content of which is being constituted by a part of the world depicted *as internal*.” Metzinger, *Being No One*, p. 36. (Emphasis his).

⁴⁰ Metzinger, *Being No One*, p. 402, p. 600. Metzinger defines a minimal notion of self-consciousness as having three properties: “the content of the self-model has to be embedded into a currently active world-model; it has to be activated within a virtual window of presence; and it has to be transparent.” “Phenomenal Transparency and Cognitive Self-Reference” *Phenomenology and the Cognitive Sciences*, 2, 2003: 353-393, p. 373.

⁴¹ Metzinger, *Being No One*, p. 411. PMIRs “depict a certain relationship as currently holding

between the system, as transparently represented to itself, and an object component.” E.g., ‘I am someone who is hearing the sound of the refrigerator.’

⁴² Metzinger, *Being No One*, p. 402.

⁴³ Metzinger, *Being No One*, p. 405.

⁴⁴ “Cognitive self-reference always is reference to the phenomenal content of a transparent self-model.” Metzinger, “Phenomenal Transparency,” p. 385.

⁴⁵ Metzinger, *Being No One*, p. 626.

⁴⁶ Metzinger, *Being No One*, p. 1.

⁴⁷ Metzinger, *Being No One*, p. 634.

⁴⁸ Another example: Metzinger speaks of a system’s integrating “its own operations with opaque mental representations...into its already existing transparent self-model while simultaneously attributing the causal role of generating these representational states to itself, the system as a whole...” Metzinger, “Phenomenal Transparency,” p. 369. How can a system *attribute* anything to itself?

⁴⁹ Let R1 be the activated representation (of the opaque self-model that is integrated into the transparent self-model) that has the content ‘the thinker of this thought’. Let R2 be the inaccessible representation in the transparent self-model. R1 interacts with R2 along with other representations with content like ‘believes that’ and ‘driving so fast’.

Our teenager is an information-processing system that—as Metzinger put it—is integrating “its own operations with opaque mental representations, i.e., with mental simulations of propositional structures that could be true or false, into its already existing transparent self-model while simultaneously attributing the causal role of generating these representational states to itself.” (Metzinger, “Phenomenal Transparency,” p. 369)

⁵⁰ Metzinger, “Phenomenal Transparency,” p. 372; Metzinger, *Being No One*, p. 400.

⁵¹ This point is adapted from a long critique in my “Naturalism and the First-Person Perspective” in *How Successful is Naturalism?* Georg Gasser, ed. (Frankfurt: Ontos Verlag, 2007): 203-226.

⁵² There will also be intractable difficulties--too complicated to take up here—about the A-series of time, which, as I have argued elsewhere, depends on the first-person perspective. We require the A-series (past, present, and future) in order to hold people responsible for what they have done. But the A-series finds no place in science.

⁵³ I am grateful to Gareth B. Matthews, Phillip Bricker, and Hilary Kornblith for comments on earlier drafts.