Out of this world

Benj Hellie[*]

APA Pacific 2013

Hi Brie, Geoff, and Victor—apologies for the length! I've tried to make the conceptual hierarchy of the paper project onto the dimension of duration in such a way that readers can calibrate to their own taste the tradeoff between convenience and interest in the details without much skipping around in the MS.

I recommend the following approach: read straight through until you run out of interest or time, except—interpolate section 4.4 between sections 1 and 2.

Suppose that Ro is 'cognizing' rocks, grass, or brains: thinking about them—or looking at them, manipulating them, theorizing about them, explaining or understanding how or why they do what they do.

Suppose that Mo is 'cognizing' the conscious life of the Las Vegas short-order cook, who prepares 500 orders of eggs a day: reflecting on what it is like to be him, or grasping this, or understanding it, or 'feeling' what it is like, or explaining or understanding how or why he does what he does.

According to the *Verstehen* tradition, the kind of cognizing Mo engages and the kind Ro engages are very different animals. Ro approaches her subject matter 'from the outside': 'objectively', 'theoretically', 'structurally'. Mo, by contrast, approaches his subject matter 'from the inside': 'subjectively', 'simulationally', 'empathetically'.

This contrast is in the main unfamiliar to contemporary analytic philosophy. Or to the extent that it is familiar, it is not well understood. In particular, our formal apparatus for theorizing about meaning is relatively good at representing objective, theoretical cognizing. But we have no formal apparatus fit to represent

---

subjective empathetic cognizing. When we attempt to represent our understanding of consciousness, we use our tools for representing objective cognizing.

But that misrepresents our understanding of consciousness. No small matter: for the conceivability argument against materialism relies on this misrepresentation.

When we cognize consciousness 'as such', we are—so to speak—'struck' in ways very different from the ways we are struck when we cognize the nonconscious 'as such'. This makes for an extremely strong pull toward dualism of *some* sort. Our position is appropriately dualistic. But the dualism is not at the level of metaphysics, of ontology: we do not recommend drawing a line, in the course of objective theoretical inquiry, to segregate off some of the things we find out there from the rest. Our dualism is rather at the level of semantics, of cognitive structure: we recommend drawing a line, in the course of our subjective elucidation of the structure of our minds, to segregate *objective theoretical inquiry itself* off from our myriad and intricately interrelated ways of cognizing.

Because we must be dualists, if we *don't* draw the line where we recommend— if we think of objective theoretical inquiry as the only way we have of cognizing— we will have no choice but to push our dualism into metaphysics. But once we have seen that this line can be drawn—and how, where, and why to draw it—the motivation for dualism in metaphysics vanishes immediately.

So this paper argues. Section 1 explains, in general terms, the response I urge to Chalmers's conceivability argument against materialism. The remainder of the paper fills out the details. Sections 2 and 3 establish various foundations; section 4 builds the details of the response to the conceivability argument atop this foundation.

In particular: section 2 discusses consciousness in general terms, first from the point of view of phenomenology and then from the point of view of rational psychology. Section 3 then discusses meaning in general terms. I begin by sketching the rudiments of a formal approach I call 'mindset semantics'—in effect, a semantics built to fit consciousness as I describe it in section 2. The next subsection makes explicit the way I intend theories developed using mindset semantics to be

2

understood. And the final subsection sketches a pair of applications: first, how to use the categories of mindset semantics to sharpen the distinction between *objective theoretical inquiry* and other varieties of cognizing; and second, how to use certain devices made available by mindset semantics to explain what it is for a certain discourse to be *expressive* rather than *informational*.

Section 4 is largely formal, aside from its concluding subsection 4.4: the formal part develops mindset-semantical treatments of various phenomena central to the recent philosophical discussion of consciousness: our perceptual contact with colors (4.1); the structure of our self-knowledge and what we mean when we talk about 'what it's like' from the first person view (4.2); and the structure of our knowledge of others and the position of empathy in this structure (4.3). The concluding section 4.4 returns to a largely informal key, assembling the preceding formal discussion and linking it back to our response to the conceivability argument.

1 Meaning and the ontology of consciousness

Materialism is the following highly general hypothesis. If the *facts just about matter* are exactly a certain way, whatever *just happens* to be so is *implicit in* these facts. Perhaps the italicized phrases are so critically unclear that nothing has yet been said. Still, I find enough clarity in the statement that it seems incompatible with the following hypothesis: although, for certain facts, they are universally correlated with the facts just about matter, this correlation is the responsibility of *superadded bridge laws* connecting those facts with the facts just about matter. For any 'superadded' laws would presumably be 'over and above' [***wilson] the facts just about matter—and therefore not truly 'implicit in' the facts just about matter.

Sanding off various details,[1] Chalmers argues against materialism as follows.

---

[1] In particular, we sand off the fork of the argument addressing 'Russellian monism'. That fork requires two pieces of apparatus I find very hard to make sense of: the notion of *fundamentality*—a sort of polarized version of our 'implicitness in'—and the distinction between primary and secondary modalities.

Let *P* be a sentence stating the full truth about physical reality and let *Q* be a sentence stating the full truth about consciousness. Let *Z* entail ¬*Q*. Then:

1. (a) It is conceivable that $P \land Z$ (premiss)

   (b) It is possible that $P \land Z$ (a, CP)

   (c) $P \land Q$ (fact)

   (d) If *P*, it is contingent whether *Q*   (b, c)

   (e) If there are any psychophysical correlations, they are the product of contingent psychophysical laws   (d)

I think I can make sense of the 'contingency' of psychophysical laws being sufficient for their 'superaddition'; and so I think (e) is incompatible with materialism.

The rule (CP), of course, extracts *it is possible that φ* from *it is conceivable that φ*: this rule strikes me as valid, as I will explain shortly. Premiss (b) is stipulated to be true. Chalmers supports premiss (a) by reflection on our reaction to 'zombie' and 'inversion' cases.

In my view, the argument is valid but unsound: premiss (a) is not in fact supported by our reaction to 'zombie' and 'inversion' cases.

To see why this might be, let us fix the sense of 'conceivable' required for the validity of (CP). Modals as generally used seem to convey views about what to take seriously. What 'must' be is what not seriously to reject; what 'can' be is what, perhaps, seriously to accept. Depending on our interests, what 'must' be in discourse *A* is what 'can't' be in discourse *B*. So for example, we all know that arithmetic 'can't' be effectively first-order axiomatized—and yet the Gödel proof goes by reductio, so in the course of considering it, we consider the 'possibility' that arithmetic has been axiomatized in this way.

Still, for present purposes, the modals are linked to the kind of 'contingency' which, if psychophysical laws had it, they would be superadded to the material. In that sense, 'contingency' is evidently affiliated with empirical inquiry: with pursuit of questions that, broadly, need to be answered using evidence rather than merely linguistic/conceptual discipline. And this in turn sets up an appropriate sense of 'conceivable', if (CP) is to be valid.

When it is 'conceivable' that φ, we find that, when we try to picture circumstances in which φ, we can do it. For example, suppose I am uncertain whether goats eat cans. My uncertainty here involves two components. First, I find myself able to picture circumstances in which goats *do* eat cans and also able to picture circumstances in which goats *do not* eat cans. Second, for neither circumstance has my evidence forced upon me information about which predicament we find ourselves in which is not consistent with our being in that circumstance: where the hypothesis that we are in both that predicament and that circumstance simply does not ultimately make sense. By contrast, suppose I am uncertain whether the product of a million and a billion is a trillion or a quadrillion. My uncertainty here is of a different form. I do not find myself able to picture circumstances in which the product of a million and a billion is a trillion, and able to picture circumstances in which their product is a quadrillion. Rather, I just don't know what the output is of a computation which takes the product of one million and one billion.

In the mathematical case, my lack of an answer is not the result of a deficiency in the evidential information I possess. Rather, in posing the mathematical question, I presuppose the discipline imposed by a certain complex form of language or conceptualization. My lack of an answer is the result of my failure to submit myself to the discipline through which that form of language or conceptualization would issue an answer.[2]

My predicament regarding goats, by contrast, is not one that linguistic or conceptual discipline is fit to resolve. When we find ourselves in the 'goats' sort of predicament, the conflicting scenarios we fully picture are both 'coherent': with a full sense for what the scenarios are—with a full grasp of the representational powers of the language or concepts we use in grasping those scenarios—neither scenario is such that we find it absurd or useless or senseless to suppose that it is so. Absurdity is, moreover, not to the point in generating conviction in one rather than the other scenario. Tutelage in the proper linguistic or conceptual discipline would not be fully apt to convince those who disagree. Only the 'external' disci-

---

[2]This 'linguistic discipline' need not be articulated explicitly: indeed, if the measure of explicit articulability is [***effective], [***godel]; rather, [***more abstract norm perhaps implicit in know-how].

pline of information incompatible with one answer forced upon me by evidence is apt to ultimately generate final conviction in the other answer:[3] the 'internal' discipline of language or conceptualization is insufficient. In that sense, each scenario is 'compatible' with the discipline of our language or concepts. And in that sense, each scenario is *conceivable* (perhaps *describable* would be equally fitting terminology).

We have, then, an ideal of subjection to linguistic/conceptual discipline which we presuppose in natural-scientific inquiry. With this discipline in place, many questions still are conclusively answered only with appropriate information forced upon us through evidence such that it makes no sense for the answer and the information to be both true: that appropriate conceptual or linguistic discipline allows us to eliminate those possibilities. For such questions, all answers are *conceivable*. And in the course of such inquiry, all answers are *possible*. And, because it is this sort of inquiry that reveals the broad structures of what just happens to be so, what is in this sense *contingent* in light of certain facts is *not implicit* in those facts.

So, having fixed on the relevant spirit in which to understand the modalities, we see what (CP) means such that the case against materialism is valid; and so we locate as the only point of attack premiss (1a).

I stated that premiss (1a) is supported by reflection on our reactions to zombie and inversion stories. In these stories, both zombies and inverts are beings (in all intrinsic and extrinsic physical respects) physically just like one of us—Fred, say—but for the zombie, there is nothing it is like; while for the invert, what it is like to see a red thing is the same as what it is like for Fred to see a green thing, and so forth. If 'Q' is the total truth about what it is like for all of us, while 'Z' is the truth about what it is (or isn't) like for the zombie or the invert, then 'Q' and 'Z' are incompatible: 'Z' entails '$\neg Q$'. Recall also that 'P' is a sentence stating exhaustively just the physical facts.

The support for premiss (1a) then is this. Reflect on the story '$P \wedge Z$': is it

_____

[3]Though, of course, 'plausibility' considerations might lean us toward one answer over than the other, as a provisional matter.

coherent? Well, we find no *glaring* incoherence in a scenario in which things are physically exactly like *blah blah blah*, but such that for *that* being, there is nothing it is like for it; or such that for *that* being, what it is like is *thus* (namely, the 'opposite' of what it is like for Fred)—a scenario in which $P \land Z$. So if there is any incoherence, it must be of a sort not easy to extract.

Our existing semantic apparatus allows for really just two sorts of hidden incoherence. One is manifest in this story: 'some barber shaves all those who do not shave themselves'. Here the incoherence is hidden behind an intricate pattern of quantification, negation, relation, and reflexivity which makes its meaning difficult to compute out. The other is manifest in this story: 'a distant planet is H_2O-filled but water-bereft'. Here, perhaps, the incoherence is hidden by our deliberate shifting of semantic responsibility from ourselves to the facts.

But there is no incoherence of either sort in the '$P \land Z$' story. For logically it is simple. And externalism seems to be irrelevant: 'P' is stated solely in terms we construct for ourselves, and the meaning of 'Z' is (perhaps) utterly basic in our understanding. Having ruled out the only known sources of hidden incoherence, we should think the story is *not* incoherent (perhaps pending unforeseen semantic advances). If the story is not incoherent, then it is coherent. And that is enough, by our previous discussion, for its being conceivable that $P \land Z$—for the truth of premiss (1a).

I want now to draw some conclusions about 'Q' and 'Z' that might seem quite tedious. I don't expect the proponent of the argument, for example, to dispute them. But I think they are false, and importantly so, and therefore deserve to be made explicit.

When we picture the circumstances in which $P \land Z$, the circumstances we thereby picture are *more specific* than those we picture when *merely* picturing circumstances in which P, of course. After all, the circumstance that Z is supposed to be *incompatible* with the circumstance that Q; and (by premiss (c)), the circumstance that $P \land Q$ is coherent. Looking at it another way, when we picture circumstances in which P, we then face a range of choices about how to go on: among them are included *picturing circumstances in which Q* and *picturing cir-*

cumstances in which Z. Evidently I can picture a being such that there is nothing it is like to be it without knowing anything about its physical constitution; evidently I can picture a being such that what it is like to be it is the opposite of what it is like to be Fred without knowing anything about its physical constitution. So whether I picture circumstances in which *Z* is independent of whether I picture circumstances in which *P*. Presumably then in picturing circumstances, I can simply do this: *picture circumstances in which Z.* And so entertaining the meaning of '*Z*' counts as *picturing circumstances.* And if so the same is doubtless true of '*Q*'.

In our story about (CP), *picturing circumstances* must align with *conceiving of non-confused natural-scientific hypotheses*, *entertaining evidential information*, and *considering possibilities for the concrete world.* These notions must collectively march in step in structuring something like a 'space of empirical reasons'. If they do not, I cannot see how to distinguish the sort of cognitive activity under consideration from the sort involved in pure mathematical calculation (including the discussion of hypotheses for reductio); and if not, I lose my grasp on the notion of 'implicit in the facts just about the material' required for the dispute over materialism to make sense. That is an autobiographical claim of no philosophical interest, of course; still, my sense is that friends of the conceivability argument will generally not deny that these notions are aligned.

But if they are, then when I entertain the meaning of '*Q*', or of '*Z*', I *conceive of a non-confused natural-scientific hypothesis*, and *entertain evidential information*, and *consider a range of possibilities for the concrete world.* And so, aside from the range of possibilities for the concrete world compatible with how I conceive of things as being in entertaining them, the thought that what it is like for Flip is *thus* (the opposite of what it is like for Fred) is not significantly different from the thought that there are 19 tons of stones strewn up and down a certain segment of a certain passway, or that 14 ounces of hot-foot powder are distributed around a certain building, or that the winding chain of a certain phonograph is not intact. In each case, we simply imagine a way for the concrete world to be: it has these and those concrete things scattered around in it; they display these or those features; they are arrayed thus and so with respect to one another. Some of these

things are 'subjects of experience' while others are just tables; some of the features displayed are 'phenomenal features', while others are just shapes; perhaps in some cases a subject of experience and a table are so arrayed that the table is 'phenomenally present' to the subject—even if so, the way the table and the chair are arrayed is just of a piece. Cognition of the non-conscious and of the conscious differ in subject-matter, but do not differ structurally in any interesting way.

Say that if to entertain the meaning of 'φ' is to *consider a range of possibilities for the concrete world* (and so forth), 'φ' has *informational meaning*. We can put our concusion concisely. The friend of the conceivability argument thinks that 'P', just like 'Q' and 'Z', have informational meaning.

But the friend of *Verstehen* denies all this. Sentences like 'Q' and 'Z' do *not* have informational meaning. Cognition of the conscious (as such) is an entirely different animal from cognition of the non-conscious (as such). To consider what it is like for Flip is not to entertain a range of possibilities for the concrete world. It is not to think of Flip, the subject of experience, as 'out there', 'displaying' certain phenomenal features. It is rather to purport to *being* Flip—a *creature*—and, having done so, to therein entertain a range of possibilities for the concrete world *as from Flip's point of view*—to think, empathizing with Flip, of the *nonconscious* things out there, scattered around that creature in a *nonconscious* array, displaying these or those *nonconscious* features. For the friend of *Verstehen*, we never distinguish possibilities regarding consciousness. The distinguishing of possibilities is something that goes on *within* consciousness, of course. But the possibilities themselves are approached entirely as *objective*. Cognition of the non-conscious (as such), then, is the *distinguishing of possibilities* for the non-conscious (from a point of view). Cognition of the *conscious* (as such), by contrast, is the distinguishing of possibilities for the non-conscious *as from a point of view* (whether someone else's, taken up through empathy, or one's own, simply in occupying it). A subtle shift of emphasis, to be sure: easy to miss, and yet (so I will argue) of sufficient significance to break the case against materialism.

If this doesn't make any sense, the case against materialism goes forward, I believe. Explaining it takes extensive work, which begins in the next section. For

9

the moment, please play along, so that we may assess what it means for premiss (1a).

The case for premiss (1a) involves the following inference rule:

(DQ) Where a sentence is a permissible substituend for 'φ' just if it has informational meaning,

- The 'φ' story is coherent ⊣⊢ it is conceivable that φ

Without the restriction, the rule would be too strong: our notion of conceivability, as argued, is tied to information in such a way that we can only predicate conceivability of information; and accordingly can only predicate it of the meaning of a sentence with informational meaning.

With the rule in hand, it is argued as follows:

2.   (a) There is no superficial incoherence in the '$P \wedge Z$' story   (fact)

     (b) There is no deep incoherence in the '$P \wedge Z$' story   (best semantics)

     (c) The '$P \wedge Z$' story is coherent   (a, b)

   (1a) It is conceivable that $P \wedge Z$ (c, DQ)

Later on we will extensively enrich our semantic theory in a way that generates novel sources of deep incoherence. We will then float the hypothesis that there is in fact a deep incoherence in the '$P \wedge Z$' story. But whether or not that is true, there is a more shallow flaw in the argument: as we will now argue, the appeal to (DQ) is not valid.

We claimed that 'P' has informational meaning but 'Z' does not. In that case, '$P \wedge Z$' should also be thought of as not having informational meaning: whatever it means, it combines the meaning of 'P' and the meaning of 'Z' in a way that results in something more specific than the meaning of 'P': otherwise '$P \wedge Z$' would be compatible with '$P \wedge Q$'. But the difference between 'P' and '$P \wedge Z$' is the involvement of more than just picturing the concrete world; and our point is that varying this kind of difference need not vary the sorts of possibilities involved.

So informational meaning is a *recessive trait*: combine it with noninformational meaning and the result is noninformational meaning.

In that case, though, '$P \wedge Z$' is not a permissible substituend for φ in (DQ). And so the argument to (1a) is not valid. Indeed, if the hypothesis that $P \wedge Z$ involves more than information, it is not even easy to evaluate what (1a) means. If it simply means that the *information* is conceivable, then it means only that it is conceivable that P. While true, that is too weak to serve in the case against materialism: all it means is that a world like this one physically is conceivable—no surprise, as we knew already that such a world is possible! But if (1a) means something stronger than that, I think we do not know what that would be.

The friend of *Verstehen* can explain the allure of (1a), such as it is. Considering (1a), we are asked: *is it conceivable that $P \wedge Z$, or isn't it?* When φ presupposes ψ, so does $\neg\varphi$. So the question 'have you stopped throwing garbage out of your bedroom window or not' offers a choice between 'yes, I have stopped throwing garbage out of my bedroom window' and 'no, I haven't stopped throwing garbage out of my bedroom window'—each of which presupposes 'I used to throw garbage out of my bedroom window'. Clearly 'it is conceivable that ρ' presupposes ''ρ' has informational meaning'; so each answer to *is it conceivable that $P \wedge Z$* presupposes (3):

   3. '$P \wedge Z$' has informational meaning

And according to the friend of *Verstehen*, (3) is false.

When we face a question with a false presupposition, we get flustered: we are used to accommodating presuppositions, and so rejecting them is a departure from habit; in ordinary circumstances, we are rarely forced to do so, so most of us do not really know how. Faced with the question 'have you stopped throwing garbage out of your bedroom window or not', most of us would stammer: 'ah, well no—I mean—not that I *haven't* stopped: ah … I *never did* throw garbage out of my bedroom window!' As a community, we seem to be equally flustered regarding what to say about (1a). To judge us by (perhaps) the best (recently) among us, a close look at David Lewis's three decades of remarks on this issue reveals a

profound and apparently irremediable consternation: as early as 1966, the great analytic materialist acknowledges the 'dualism of the common man'; as late as 2005, he continues to fret that 'so many philosophers find it so very obvious' that we 'know the essences' of our 'qualia'.

The status of (1a) is consternating because of the false presupposition. This consternation is made more difficult to push through by two features of our mainstream philosophical thought. First, the *Verstehen* tradition is largely set to the side in the philosophy of mind; and second, our formal semantic resources are largely exhausted by devices for representing informational meaning. (In the case of Lewis, analytic materialism and classical modal semantics combine as mutually reinforcing pillars of his system.) So we are not really even in a position to explicitly articulate the source of our consternation. And as a result, we find ourselves compelled to assess the status of (1a).

Which side to take in this uncomfortable position? According to the friend of *Verstehen*, our discomfort is generated by our recognition that cognizing consciousness and cognizing the nonconscious are entirely different animals: that the meaning of 'P' is deeply very different from the meaning of 'Q' or of 'Z'. But if the only meaning we explicitly recognize is informational meaning, we can only acknowledge this recognition by positing a deep difference between the sorts of informational meaning entertained in cognizing consciousness (perhaps by using 'Q' or 'Z') and cognizing the nonconscious (perhaps by using 'P'). And then, faced with the compulsion to either assert or deny (1a), we are faced with a compulsion to assert either that there is *some big difference* between the meaning of 'P' as contrasted with the meaning of 'Q' or 'Z' that there is *no big such difference*. The choice here is obvious: there had better not be *no big difference*! And at this point, our only way to recognize this is to affirm (1a).

A deeply regrettable situation we are in—if this story is correct. The friend of *Verstehen* thinks it is correct and also is not in the regrettable situation (assuming the position can be made sense of). I am antecedently friendly to *Verstehen* because the characteristic views of the tradition strike me as just more or less obviously accurate to ordinary life. What about others? *Simulationists* are the

latter-day inheritors of the *Verstehen* tradition: they should find themselves happy to be in possession of the way out of the dualism debate. Those who wish they did not have to affirm (1a) but find denying it even worse should get on side: the alternative is to simply reject it. Those who find themselves constitutionally opposed to dualism but who are also averse to tampering with classical modal semantics as a way to represent information should do so as well: our approach, unlike 'type-B materialism', will not require any such deep tissue surgery. As should those who wish to preserve materialism and classical semantics, but feel uncomfortable affirming that there is really nothing distinctive about our understanding of consciousness.

A final group includes philosophers who are committed to classical modal semantics and to the theory theory, and who have reconciled themselves to accepting either antimaterialism or that nothing is special about the mind. Some of these theorists, like David Lewis and David Chalmers, whatever their own explicit views may be, think that almost everyone has explicit beliefs about the general relationships among matter, consciousness, and representation that are inconsistent with their implicit beliefs. These philosophers might worry about how conscious material beings consciously representing a material world ever manage to get anything done despite massive inconsistency in the most basic and general views about what we are like and what we are doing; and how philosophy could expect to proceed if our world view is really so permeated with incoherence. I suggest to these philosophers that technical conservatism is not in itself a positive reason to believe anything. And I suggest that rejection of simulationism is really very shallowly-rooted: where are the arguments against simulationism, and how could there be anything to them? I don't know. And it seems to me that, given a choice between abandoning phantom arguments and technical conservatism as against affirming pervasive incoherence, the wise choice would be the former.

Of course, this all hangs on whether we can make sense of the *Verstehen* position. Doing that is the work of the remainder of this paper. The approach is structured like this. I first engage in a broadly 'phenomenological' discussion, pointing out certain 'divisions' in the general structure of consciousness and de-

veloping a way of speaking that marks these divisions rather than blurring them. I argue next for a link between consciousness and rationality: rational explanation is empathetic making-sense; nothing aside from consciousness is available to this exercise; so rationality supervenes on consciousness. Our next stage is broadly 'semantical': I argue for a certain unity between rationality and meaning; but if so, the previous discussion implies that *meaning* supervenes on consciousness; so that a formal semantic theory should be a theory of the structure of consciousness; in this spirit, I introduce 'mindset semantics', a formal approach which attempts to fit that bill. I conclude by using mindset semantics to provide a rigorous statement of the doctrines of the *Verstehen* tradition; which, finally, equips us to grasp the sense in which sentences about consciousness do not have informational meaning.

2 Consciousness

2.1 The structure of consciousness

These sentences Q and Z: what kind of thing do they say? Q, perhaps, includes this conjunction:

4. Part of what it is like for Greg is this: G; and part of what it is like for Rebecca is this: R

While Z, perhaps, is this conjunction:

5. Part of what it is like for Greg is this: R; and part of what it is like for Rebecca is this: G

And what are R and G? Presumably something like this, respectively:

6. (a) This is now thus-red
    (b) This is now thus-green

Where to get a handle on (6a), pronounce it while staring at a red thing qua red, while to get a handle on (6b), pronounce it while staring at a green thing qua green.

I am inclined to think that part of what I mean when I assert (6a) is that the thing, this, is given at the present moment qua thus, namely red. So it is, in general, part of what it is like that one or more things are given at present qua one or more perceptually identified features. So part of what Q states is that, over time and to various subjects, certain things are given at the then present moment qua certain perceptually identified features.

What can be given and how can it be given qua? Look around you, feel within, reflect on your perceptual situation. Objects, events, surfaces, beams, sounds, smells, aches; floaters in the eye, doubled perception, blurredness of perception; motion; sense-memory of things that pass in and out of view.

This now qua thus is not all there is to what it is like for Greg, of course. In general, that this is now thus is a core part of my picture of the world—but still only a part. Much of belief affects what it is like for one. Think of what it would have been like for you five minutes ago to be uncertain whether the chair you are sitting in is sturdy. It would be different, of course. So your belief that the chair you are sitting in is sturdy influenced what it was like for you. More generally, one harbors at a time a 'picture of the world': a consciously registered body of perspective-independent information. This information—any certainties or uncertainties one may have about the world—is 'wrapped around' the given. Neither reduces to the other: without information about the world beyond the given, I am stuck in the here and now; without the given, I have no sense for my place in the world.

So far we are only registering within consciousness things, aspects, and purported truths not themselves at all pertinent to consciousness. That this is thus is not a fact about consciousness but a fact about this. Even if being thus is having a 'primitive quality', that is not a fact about consciousness. Chalmers's 'Eden' thought experiment shows this convincingly: if my stuffed elephant's pinkness is a primitive quality, it is a quality something could have in lifeless Eden. If this tree's

greenness is a primitive quality, it is a quality something could have in animal-free Eden. If zombies are intelligible—and here, recall, it is my *opponent* who thinks so, so that I make this supposition without risk of begiing the question—then the tree could also have primitive greenness in 'zombie Eden' without thereby turning the lights on for the zombies. A zombie labrador retriever could have a primitive-yellow coat without turning the lights on. The zombie labrador could make a primitive-barking sound without turning the lights on. A pile of bread could have a primitive-bready aroma without the lights going on. The zombie labrador could have a primitive-stretching sensation without the lights going on. The zombie human patting the zombie labrador could have a primitive-furry hand sensation without the lights going on. The zombie human could have a primitive-pain sensation without the lights going on. The zombie human, looking at the primitive yellow coat, could primitive-see it with primitive double-vision without the lights going on. The zombie human with reverse-wired visual apparatus could have the same primitive-Y visual sensation upon looking at the sky as the normal zombie has upon looking at the labrador's coat.

Why is this? When we consider a zombie, we simply *neglect to empathize*. I am inclined to think that *any quality whatsoever*, and *any individual whatsoever*, and *any relation among individuals whatsoever*—and any pattern among these things—can be stipulatively introduced into a situation together with a refusal to empathize. Suppose that, while refusing to empathize, we stipulate that *x* bears *R* to *y*'s *F*-ness. Suppose that we now stipulate that *y* is a tree and *F*-ness is *greenness*: this can't stop us from refusing to empathize. Alternatively, suppose we stipulate that *y* is a 'sense-datum' and *F*-ness is 'green-prime-ness': I do not know what that is, so I cannot tell whether this stipulation forces us to empathize.

Suppose that we now stipulate that *x* is a human being: my opponent certainly thinks this doesn't force us to empathize. Alternatively, suppose we stipulate that *x* is a 'schmubject of schmexperience'. I do not know what that is, so I am not sure whether that stipulation forces us to empathize. But I can't see how, if stipulating that something is a human being—a very familiar kind of entity—does not force us to empathize, we *could* be forced to empathize by the stipulation that something is

an entity of an unfamiliar kind. And I for one don't find 'subjects of experience' any more familiar than 'schmubjects of schmexperience', unless the former are understood as *animals*. If they are, then if we are forced to empathize with a subject of experience, we are forced to empathize with a human; if they aren't, they are in the same boat with schmubjects of schmexperience.

Similarly, suppose we go on to stipulate that *R* is *seeing*. I am inclined to think of *seeing* as an 'ecological' relation, with a nature that can be fully understood without empathizing (if that isn't obvious, think of a beetle seeing a morsel). So this stipulation doesn't force us to empathize either. And if that is so, I don't see how things could be different if we stipulate that *R* is *acquaintance*, or some other unfamiliar relation.

So if we can—as my opponent thinks—stipulate a situation in which this human sees that tree's greenness while refusing to empathize, I don't see what could prevent us from stipulating a situation in which this subject of experience is acquainted with that sense-datum's green-prime-ness while refusing to empathize.

Perhaps, though we *can* refuse to empathize, we are *wrong* to so refuse. But this mistake would only apply in certain cases and not in all cases: I can certainly correctly stipulate the existence of a table while refusing to empathize. My opponent thinks it is also no mistake to stipulate the existence of a flesh-and-blood human being in a situation just like that of Hank Aaron while refusing to empathize. But if that is OK, it is beyond credulity to deny it is OK to stipulate a highly unfamiliar situation while refusing to empathize.

So, if zombies are conceivable, it would seem that *no matter what* story I tell, it is OK to refuse to empathize—unless that story begins with the (implicit or explicit) preface 'here's what it's like (for *x*)'. If so, consciousness only ever 'gets into the act' once that preface is added. What goes on within the *scope* of the preface does of course characterize what it is like. But without the preface, nothing could have any significance whatever for what it is like.

Facts about the qualities of the given, registered as such, are important to consciousness, but they are not the whole of consciousness, and they are not sufficient for consciousness. Facts about consciousness are not the facts we state *after* say-

ing 'here's what it's like'. Rather, to state these facts, we have to include that preface.

Suppose that at *t*, I state every truth of form 'here's what it's like: φ'. Have I left any latitude remaining regarding what it's like for me? Evidently not. Have I left any latitude remaining regarding how my stream of consciousness then is? If the question simply repeats the previous question, it gets the same answer; otherwise, I am not sure what is being asked. So it would certainly seem that the totality of truths regarding how my stream of consciousness at a time is is the totality of truths I would then state of form 'here's what it's like: φ'.

In a lot of respects, saying 'here's what it's like: φ' does not differ from just saying φ. Whenever I am willing to say one I am willing to say the other. That goes without exception: whenever I am willing to say the former, I am willing to say 'here's what it's like: here's what it's like: φ'. And whenever I am willing to say $\neg\varphi$, I am willing to say 'here's not what it's like: φ'. The difference is this: I am willing to say 'here's not what it's like: φ' even when I am not willing to state either φ or $\neg\varphi$. There aren't any gaps in the world: only ever my own ignorance. But there are gaps in consciousness: my ignorance is itself reflected within consciousness.

So in that sense, consciousness is *almost transparent* to itself. For those aspects of the world that are within consciousness, consciousness doesn't find itself getting in the way. But when the world is not within consciousness regarding a certain matter, that very absence is itself within consciousness: the absence of an answer gets in the way of the answer, whatever it may be.

Nor for that matter is consciousness limited to registration of information. Just for one, the presentness of the present moment is not a piece of information: otherwise we would contradict ourselves as we felt time pass. Nor, perhaps, is the givenness of the given a piece of information. Perhaps if *j* is given as *G* at *t*, one has the information that *j* is *G* at *t*. But one can have that information whether or not *j* is given as *G* at *t*. Does one then have the information that *j* is given as *G* at *t*? It can seem not: if *j* is given as *G* to Sam at *t* but *k* is given as *H* to Fred at *t*, Sam would then have the information that *j* is given as *G* while Fred

would have the information that *k* is given as *H*. They would then have something to argue over—but they manifestly do not. Or does Sam have the information that *j* is given as *G to Sam*? If this is information Fred could have, possession of this information is insufficient for the phenomenology of givennness. If this is information Fred couldn't have, this must remain forever a mystery to Fred. The world seems oddly fragmented into private chambers.

Nor for that matter must we think what it is like is exhausted by the broadly 'cognitive'. Plausibly 'conative' and 'affective' matters are registered in what it is like. But this registration is not by itself registration of the fact that what it is like is such and such: it is rather a registration of the fact that such and such. The transparency of consciousness is not a transparency of 'the mental', if the conative and affective are included in the mental. It is rather a transparency of consciousness *to* these facts about the mental. (It may be, however, that these 'facts about the mental' are not themselves matters about which we could be informed.)

I have been speaking of 'consciousness' to mean 'my consciousness now'. But the latter locution is not 'more basic' than the former locution. Nothing about consciousness marks itself as one among many, or this time as but one among many. That is not to say of course that I am unaware of other times and people. But my awareness of them is of a radically different variety than my self-awareness. In order to engender such awareness, I need to *modify* my present consciousness. I need to *recollect* the past, or *anticipate* the future, or *empathize* with the other. These are modifications of consciousness, in something like the way that having information, or being given something, or undergoing affect, are modifications of consciousness. These modifications of (my own) consciousness are requisite for my recognition of the consciousness of the other. But no such modification is requisite for my recognition of my own consciousness. That comes along 'for free'. Consciousness is, we might say, *perpetually self-permeating*.

2.2 Consciousness and rationality

Consciousness, then, includes a fair bit. But how shall we decide what does and does not go into consciousness? What are the upper and lower limits? I suggest

that consciousness must include at least enough to explain rationality; and I doubt that consciousness needs to include anything else.

More specifically. Let a *magnitude relevant to rational assessment* be a family of features, such as the many varieties of belief state, to which appeals are made in rational explanation. Then:

7. **Reason reduces to consciousness**

   One's magnitudes relevant to rational assessment are constituted by what it is like for one.

It follows from (7) that there must be at least enough in consciousness to accommodate rational assessment: in particular, that reason *supervenes on* consciousness, so that any differences relevant to rational assessment must be accompanied by differences in what it is like.

Why believe (7)? I once defended the supervenience thesis by thought experiment:[4] let *A* and *B* be the same in what it is like; let *C* and *D* be the same in what it is like; let *A* and *C* (therefore *B* and *D*) differ in what it is like; let *A* and *C* be the same in all remaining respects; let *B* and *D* be the same in all remaining respects; let *A* and *B* (and therefore *C* and *D*) differ in respect of what it is like: we then judge that in whatever ways these subjects might be rationally in order or out of order, these ways track similarity in what it is like rather than other respects of similarity (so *A* and *B* are rationally alike, as are *C* and *D*, but neither pair is rationally like the other).

A more effective (and more illuminating) line of argument explains these judgements. Rational norms are active only at the first-person. If I want to know—rationally, as opposed to efficient causally—why Fred came to believe that goats eat cans, or why Phyllis moved this widget to the right rather than the left, I want to know what it was like for Fred, or for Phyllis, such that updating in this way was appropriate in its light. I want *verstehen*: an empathetic grasp of the recent course of Fred's or of Phyllis's conscious life such that forming the belief or moving the widget *made sense* against its historical background. I don't care about

---

[4][***cite], submitted for review in September 2004.

microwave beams impinging the pineal gland or any computational structure realized in the brain; I don't care about his relationship to his mother or whether ten years hence either will be president of the JCs; I don't care about goings on on Twin Earth or in anyone else's conscious life—or at least I care about these only insofar as they are manifest, at least as a matter of regulative ideal, within what it is like for Fred or for Phyllis: if not, in what sense could these issues be in any way relevant to *making sense*? Any explanatory force they would possess must be of some other variety: presumably of the efficient causal variety.

So in particular, consider the zombie: its arms and legs manipulate objects and transport it about; its sense organs transduce signals from the environment; its face contorts; its vocal apparatus emits acoustic blasts. Why? Not because any of this *makes sense*—makes sense of anything or itself makes sense. Thinking about what it is like and finding nothing, I can never nod and say 'yep, me too'. Rather, only efficient causal explanation will ever be available.

One might object to (10): consciousness is internal and rationality is external, so that rationality cannot supervene on consciousness. But rational explanation goes by *verstehen*, which requires empathy, which displays consciousness. Some might take the lesson that rationality must be internal. But the language of rationalization discusses externalia ('he threw the stone at them to protect his mother'), and we lack a fallback language that does not; so if this language discusses rationality, means what it seems to, we understand it, and we mean it literally, rationality must be external. This is the classical deadlock between internalism and externalism.

But *modern externalism* breaks the deadlock. Rationality supervenes on consciousness while being external because consciousness is external—so says the modern externalist. And fortunately there is plenty of reason to assume consciousness is external and no reason to assume consciousness is internal.

Here is an argument that consciousness is external. We begin by arguing that consciousness pays attention to the particularities of things. Consider Lewis's two gods: Hera lives on the highest mountain and throws down manna; Zeus lives on the coldest mountain and throws down thunderbolts; each of them knows exactly

21

what the world is like 'as a whole'; and yet each is utterly uncertain where or when he or she is located in the world.

Suppose Hera's followers are suffering a famine; her body transduces information pertaining to this famine, performs some computation, and behaves in such a way that manna is hurled down by the arms of the body. Zeus's body does nothing in response. Can we rationally explain these reactions? No. What is it like for Hera? Take a zombie and give it a total picture of a world containing two gods of an appropriate nature. And that is the same as what it is like for Zeus. We find nothing to make sense of their differential reactions.

Giving them just a bit of 'self-locating information' won't help. Introduce a third god—Pham. Each of Hera and Zeus is certain he or she is not Pham, but has no idea whether he or she is Hera or Zeus. We still cannot make sense of the differential reaction: what it is like for them does not differ.

Nor can Hera ever then learn which of Zeus or Hera she is. To do so would require 'self-locating evidence'. But if there is any such thing, presumably Zeus's evidence establishes he is Zeus, while Hera's evidence establishes she is Hera; so their self-locating evidence is jointly inconsistent. But which evidence should Hera abide? Perhaps she *would* abide Hera's evidence. But why *should* she do that? Uncertain about which god she is, it would seem to make no sense to abide *her* rather than *his* self-locating evidence.

Suppose Hera mistakenly thinks she is Zeus. Can we make sense of her reactions? No: presumably to think she is Zeus is to think she is on the coldest mountain, stocked with thunderbolts rather than manna, and unconcerned with the followers of Hera.

To make sense of Hera's reaction, we need her to be confident, at some level, that she is Hera. But, in the absence of evidence to discriminate this prospect from the prospect that she is Zeus, such confidence would be irrational were error possible. And as we have seen, there can be no such evidence. So we must conclude that it is impossible for Hera to be in error at this level about which of Zeus and Hera she is. This infallible confidence must be, moreover, part of what it is like. So one's own particularity is built into what it is like; to the extent that

this counts as 'external', consciousness is external. (Notably, the view is in line with Frege's doctrine that each of us is present to him or herself in an absolutely distinctive way.)

A similar argument could be made for building particular *times* into consciousness: why *scoop* manna rather than *hurl* it at this moment? Unless Hera at *t* has some sense of which interval of her life she occupies, no reason at all. Because action is finely grained without limit, there is no lower limit to how finely the present moment is manifest in consciousness.

A full *externalism* properly so-called is reached by thinking of behavior and sensation as fundamentally intertwined and 'ecological'; and of action and perception as requiring the manifestation of aspects of this ecology without a priori limits on spatial or temporal distance.

And I see no compelling reason to deny consciousness is external even in this expansive sense. If asked what it is like for me, I would say 'I am writing my APA talk'. That means I will in the future finish this APA talk. So what it is like for me supervenes on the future. Of course I may be wrong: I may be hit by a bus and never finish the talk. But if so, I am wrong about what it is like for me. It *seems* as if this is what is like for me, but it is not. My mistake about consciousness is explicable in virtue of another mistake I have made: my assumption that I will not be hit by a bus.

This might seem to conflict with the view that consciousness is 'luminous': that one has the correct answer to every question about what it is like for one. But luminosity is compatible with one's occasional possession of an *incorrect* answer to a question about what it is like for one. False belief and true belief are sometimes compatible. Still, sometimes the information falsely believed and the information truly believed are incompatible. In that case, we have *fragmentation*: a multiplicity of individually coherent but jointly incoherent mindsets. Under those circumstances, there is no way to make sense of what it is like for one that is *transparent*, *unified*, and *coherent*. The external empathizer must choose from among several imperfect strategies; the subject him or herself is in certain respects in the dark.

3 Mindset semantics

Linguistic activity is an activity and therefore rationally guided. The point of linguistic activity is to mean things. Semantics studies meaning. So semantics studies something rationally guided, and therefore studies something that supervenes on consciousness. Mindset semantics is a formal approach for representing these ideas.

3.1 Information, support, and context

The core doctrines of mindset semantics are a principle (8) relating *context*, *information*, and *modality*; a principle (9) relating *reason*, *context*, and *information*; and a principle (10) relating *context* and *consciousness*.

We can recognize at least two aspects to the meaning of a declarative sentence φ against a context c. There is its *content* relative to c, $[\![\varphi]\!]^c \subseteq W$—namely, a proposition. And there is φ's context invariant *support condition*, $\lfloor \varphi \rfloor$, which is the set of contexts at which φ is accepted.

Here the notion of 'acceptance' is of *implicit* acceptance by virtue of *competence* with the sentence rather than of *explicit* acceptance by virtue of *performance* with the sentence. The theory of competence is a chapter of rational psychology. Failures to act in accord with competence due to performance errors are episodes of rational incoherence, requiring for their explanation 'special pleading' outside of rational psychology.

The content of φ against c, to reiterate, marks the *information* φ encodes against c. The information encoded by a sentence (against a context) is that which one believes in accepting the sentence (against the context). We represent one's belief state at a context as the total increment of information one has. An increment of information, in turn, is represented as a region of modal space. Accordingly, where i_c represents the belief information available in c:

8. **Information is modal**

   $i_c \subseteq W$

Context may include further parameters as well: just which parameters it does include, we discuss shortly.

The support condition for a sentence is the set of contexts at which it is accepted. To accept a *declarative* sentence (against a context) is to believe the information the sentence encodes (against the context). (For a nondeclarative sentence, what it is to accept it is not that: more below.) To believe at c what a sentence φ encodes against c is for the increment of information one possesses to be more specific than what is encoded: for $i_c \subseteq [\![\varphi]\!]^c$. Accordingly, the support condition for φ is this:

9. **Support is belief**

   $\lfloor \varphi \rfloor = \{c : i_c \subseteq [\![\varphi]\!]^c\}$

Namely, φ is accepted at c just if c believes the content of φ against c; the support condition of φ is the set of contexts for which this is so. It follows from (8) and (9) that content is modal: $[\![\varphi]\!]^c \subseteq W$.

The apparatus of support conditions as an element of meaning is somewhat unfamiliar: many theories of meaning concentrate solely on 'sentence meaning', a matter of what is depicted 'in the abstract', considered to some degree in isolation from psychology. But what is truly distinctive of our position is our view regarding the limits of context:

10. **Context supervenes on consciousness**

    If c and c' are predictively distinct, then what it is like for the subject of c is distinct from what it is like for the subject of c'

The core of mindset semantics, then, is the conjoint use of (8), (9), and (10) in explaining linguistic meaning.

On our view, contexts represent *mindsets*: aspects of streams of consciousness. Standardly, context is understood as a point of spacetime—and therefore

as without any psychological significance to speak of. Our position could not be more starkly opposed.[5]

Why believe (10)? The arguments of the previous section show that consciousness *can* do the job, required by (9), of providing an information state. They show also that consciousness *might* provide a range of further aspects of context—as is, more generally, motivated by the desire to preserve what is virtuous from existing semantical theories in light of (8). For consciousness is multifarious; context might therefore also involve much more than an information state. Finally, they show that consciousness *must* do the job: the core notion of mindset semantics, *support*, pertains to rational sentence acceptance; and as argued in the previous section, rationality supervenes on consciousness. Whether a sentence is supported in a context depends on its content against that context; but for this to be manifest within consciousness, so must be both how the sentence's content depends on context and which context one is in.

3.2   Using mindset semantics

Suppose the mindset semanticist assigns to Fred the context c. What has she thereby done? Has she provided *information* about Fred, to the effect that there is this thing, context c, hovering around him—or if not that, that Fred has some *property* such that context c is a 'measure' of that property?

If we said that, it would be 'game over'. For in assigning to Fred the context c, our theorist aims to convey what it is like for Fred. And understanding what it is like for Fred, according to the *Verstehen* tradition, is not possessing information about Fred.

Rather, c is used as an *instrument*. Our objective in semantics is to state, with mathematical sharpness, our views about someone's conscious life. Mathematical sharpness requires the use of 'magnitudes' to state our opinions. But ulti-

---

[5]The need to represent non-idiolectic or *conversational* meaning does not show that context floats free of consciousness. Our options for addressing, for example, 'common ground' include: (i) instrumentalism *by the semanticist*; (ii) instrumentalism *by the conversationalists*; (iii) realism about a temporary 'collective consciousness' in which individuals overlap over periods of conversational entrainment.

mately the magnitudes are instruments: what we are ultimately interested in can be expressed (perhaps with grave inconvenience or obfuscating unclarity) without them. Our *pre-semantical* story about Fred should mean the same as our *semantical* story. So if the former is a story that expresses our empathy with Fred, the same is true of the latter. So if expressing a certain empathetic posture is not providing information, neither is telling a certain story with the semantics.

3.3   Work for mindset semantics

Objective and subjective meaning   A long-standing ambition in philosophy has been the distinction of the more 'objective' aspects of meaning from its more 'subjective' aspects. These are captured in the famous distinctions of 'reference' versus 'sense'; of 'content' versus 'character'; of 'secondary intension' versus 'primary intension'. In our system, the content of a sentence against a context is on the objective side, while the context-invariant support condition is more on the subjective side. A first important job for mindset semantics is in its making this distinction.

Informally, we think of objectivity along the following lines:

11. When our answers to a stock of questions are intended objectively, we aim to . . .

    (a) Avoid above all accepting both answers to a yes-no question

    (b) Register our evidence rather than answering questions in ways that conflict with evidence

    (c) Seek coherence and completeness: it is my failing when some of my answers collectively answer one of my questions but I do not accept the answer

    (d) Grow our stock of answers monotonically: it is my failing when I reopen a question that has been closed

    (e) Agree with one another: it is our failing when one of us knowingly fails to accept the answer to an interesting question given by the other

27

Modal space allows us to mathematize this story. Let $i_c \subseteq W$ be the information state present at context c; let $c(j, t)$ be the context for subject j at time t; let $e_c \subseteq W$ be the total stock of evidence accumulated in c. Then our aims can be phrased as follows.

12. Ideally,

   (a) $i_c \neq \emptyset$

   (b) $i_c \subseteq e_c$

   (c) $i_{c(j,t)}$ is a function of j and t

   (d) $i_{c(j,t)} \supseteq i_{c(j,t^+)}$ where $t \leq t^+$

   (e) $i_{c(j,t)} = i_{c(j',t)}$ (blurring up our eyes at matters not of mutual interest)

Combined with an appropriate semantic theory for a stock of sentences, we can separate cases in which patterns of acceptance and rejection of those sentences mark genuine agreement and disagreement, genuine diachronic coherence and incoherence, genuine learning, failing to learn, and forgetting.

On the subjective side is that which is responsible for entailment and a priority: one sentence entails another just if, intuitively, one's psychological position in accepting the former but not the latter is incoherent; a sentence is a priori just if, intuitively, one's psychological position in failing to accept it is incoherent.

A sentence like 'I am at work' has both objective and subjective aspects to meaning. Neither is fully dependent on the other (by contrast, the subjective aspect of the meaning of 'Fred is at work at t^*' seems to be entirely dependent on its objective aspect). And yet nor are these aspects isolated from one another. It would be nice to parcel out those aspects of its meaning which are objective from those which are subjective. And it would be nice to do so in a way that allowed both to be genuine aspects of meaning: genuinely rationally normative.

Mindset semantics explains both the distinction between these aspects and how they are related: the 'objective' meaning of a sentence against a context is its content; the 'subjective' meaning of a sentence is its context-invariant support

28

condition. Support condition plus context determines content; content abstracted appropriately from context determines support condition.

To begin with, tying the subjective side of meaning to the support condition is attractive. Consider the following analysis of entailment:

13. $\varphi \vdash \psi$ just if $\lfloor \varphi \rfloor \subseteq \lfloor \psi \rfloor$

Namely, φ entails ψ just if any context supporting the former supports the latter. Our 'support-preservation' account is superior to 'truth-preservation' accounts on the grounds that (a) the notion of truth is either obscure or trivial; (b) a sentence either is or isn't true, so it is not clear what it would mean to 'preserve' its truth; (c) our account makes explicit what the truth-preservation account merely presupposes, namely that we explicitly seek to accept entailments of what we explicitly accept because we already implicitly accept them, so that an explicit failure to do so would introduce rational incoherence.

Moreover, tying the objective aspect of meaning to content is also alluring: our modal-space based account of the structure of objective enquiry provides just the right level of grain to distinguish sentences that objectively present the world in distinct ways.

Armed with an appropriate semantic theory for 'I am at work', we might then explain why it is never OK to simultaneously accept and reject it; and when and why it is OK for me to accept the sentence while you reject it, or for me to accept it at one time while rejecting it at another. Armed with an appropriate semantic theory for 'at t', we might explain why it is never OK for me to accept 'at t, I am at work' but abandon it later; and why nevertheless it is OK for me to accept it while you do not accept it. Armed with an appropriate semantic theory for 'Sam is at work', we might explain why it is never OK for me to accept 'Sam is at work at t' while you do not accept it (in full knowledge of my acceptance). We would then be able to parcel out the objective from the subjective aspects of the meanings of these sentences.

For example, let j_c be the subject of c and let t_c be the time of c. Plausibly $[\![$I am at work$]\!]^c = \{w : $ in w, j_c is at work at $t_c\}$. Then the support condition of 'I

am at work', ⌊I am at work⌋, is $\{c : i_c \subseteq \{w : \text{in } w, j_c \text{ is at work at } t_c\}\}$. So 'I am at work' has the same support condition against all contexts. This is because a support condition is a property of contexts: in general, the support condition of a sentence is a set of contexts distinguished by a pattern of relationships between information possessed and (perhaps) other aspects of context. In the present case, the support condition of 'I am at work' admits a context just if the context has the property F of identifying oneself at present with a creature whom one is confident is at work.

But at the same time, the fixed support condition of a sentence is compatible with its context-dependent content. That in virtue of which 'I am at work' is supported at *this* context rather than *that* context varies with the constituency of the context. For my context to have the property F at t^* is for it to identify *Hellie at* t^* as myself now and to be confident that subject is at work; for Sam's context to have F at t' is for it to identify *Sam at t'* as herself now and to be confident that that subject is at work. Change the time, change the subject, and the proposition changes. That is why there need be no conflict between my acceptance at a time and my later rejection of the sentence, or between my acceptance and Sam's simultaneous rejection. If 'at τ' somehow 'saturates' the parameter t_c, however, 'I am at work at noon of the first day of 2001' encodes a proposition 'eternally', such that if I accept the sentence at a time and then later reject it, I am incoherent diachronically.

This requires context to be 'external' rather than 'narrow'. In the classical discussions, the required externality was taken as a sign that context could not be luminous: that the if the support condition of a sentence is psychologically present, it must somehow 'blot out' its content against a context. This set up an unhappy dilemma: either there is nothing present to psychology that is common cross-contextually to cases in which 'I am at work' is accepted, or there is never any propositional content to such sentences present to psychology. The first horn of the dilemma is at odds with 'essential indexicality'; the second is at odds with our striving toward objectivity. But again, the dilemma is false: modern externalism shows the way out. For if context draws on resources present to consciousness,

and consciousness is external, then both the subjective and the objective aspects of the meaning of a sentence may be psychologically present. We do not need to choose.

Expressivism   A second application makes sense of *expressivism*. Here is a puzzle motivating the need for this, which I call the *curious problem*. 'Internalists' about curiosity detect a psychologically necessary connection between curiosity and belief in a certain sort of fact—the facts concerning what one wonders about. The alleged *incurousist* who believes *I wonder whether goats eat cans* but does not ask *do goats eat cans?* seems incoherent, bizarre (engaging from time to time in such weird exchanges as 'I wonder whether goats eat cans'; 'wait, you're asking whether goats eat cans?'; 'where did you get that idea? I could never care about such a stupid question!'). So, conversely, does the alleged *clueless questioner*, who asks *do goats eat cans?* while failing to believe *I wonder whether goats eat cans* (the clueless questioner from time to time issues in such weird speeches as 'do goats eat cans? wait, why did I just ask that, I don't care whether goats eat cans! but still, do goats eat cans? stop it!').

On the other side, 'externalists' about curiosity point to the 'logical mismatch' between belief and curiosity: it is our acceptance of interrogative sentences in which curiosity consists; and it is our acceptance of declarative sentences in which belief in wonderment consists. But the ancient and noble principle of the impossibility of deriving a *what?* from an *is* establishes that declarative sentences can stand in logical relationships only to other declarative sentences—never to interrogative sentences: the more sophisticated externalists put a formal cast on the objection, known to cognoscenti as the 'Chiffre-Geg Embarassment'. If it is not *illogical* to be an incurousist, then, in what could the incoherence of this position consist? We may be unlike them, but hey, it takes all kinds.

This is, of course, a parody of the so-called 'moral problem'. Unlike in the parallel case for practical motivation, there is no storied tradition behind the alleged impossibility of deriving a *what?* from an *is*. And few things could be more obvious than that one who asserts 'I wonder whether goats eat cans' simply expresses

the attitude of wonderment whether goats eat cans—that attitude the content of which is more typically encoded in the querying of 'do goats eat cans?' And yet, the Chiffre-Geg Embarassment persists: *how could* 'I wonder whether goats eat cans' simply express an attitude? It is a declarative sentence, and the content of a declarative sentence must be a proposition; and since one might easily wonder whether goats eat cans without wondering whether lambs eat ivy, 'I wonder whether goats eat cans' and 'I wonder whether lambs eat ivy' might be one true and the other false; so they must express *different* propositions. So the sentence must after all simply do what the externalist says: express a belief with the content being a contingent proposition about one's own psychology. But since there is in general no guarantee that one is right about a contingent proposition, we can just frame the alleged incurousist as someone who mistakenly thinks he wonders whether goats eat cans when he doesn't—match to the externalist.

The Chiffre-Geg Embarassment is, at some level, a technical puzzle: how to describe a kind of expression which 'logics' like a declarative but 'means' like an interrogative. But it is the technical face of an extremely deep psychological puzzle: in what does the unity of the mind consist? After all, the bizarreness of the clueless questioner and the incurousist suggests that we do not think of wondering and believing as machines that somehow run on in parallel. But until we can resolve the Chiffre-Geg Embarassment, that is how we must represent ourselves.

Mindset semantics provides a philosophically satisfying resolution of the Chiffre-Geg Embarassment. According to mindset semantics, all meaning is linked, for both contextual relevance and comprehension, to context. Context is therefore always present whenever meaning crops up. And context supervenes on (externally supervenient) consciousness. There is, accordingly, no reason whatever to think delcarative meaning must—as it were—always look outward, to the world, and never simply remain at home, reflecting solely on consciousness. If consciousness involves both opinion and curiosity, declarative meaning reflecting solely on consciousness is in a position to reflect solely on curiosity. And our claim then is that this is exactly what wonderment-avowals do. Let us now sketch the details.

Suppose that one aspect of consciousness is what one wonders about: not that about which one is simply uncertain, but those aspects of one's uncertainty one aims to resolve. If one wonders whether goats eat cans, one is uncertain whether goats eat cans and seeks to resolve this uncertainty. And one then 'accepts' the interrogative sentence 'do goats eat cans'—a state one manifests in introducing (at appropriate moments) into the conversation the query 'do goats eat cans?' We may say, somewhat standardly, that the content of an interrogative—a *question*, in the technical sense—is a *partition of modal space*: a set q of mutually exclusive and jointly exhaustive propositions, each of which is such that if one believed that proposition, one would have a distinctive answer to a query posed using that interrogative. For example, letting G be the set of worlds at which goats eat cans, $[\![\text{do goats eat cans}]\!]^c = \{G, \overline{G}\}$: the set containing the proposition that goats do eat cans and the proposition that goats do not eat cans.

We may say that one aspect of mindset is what one wonders about, and that one aspect of context is the product of all questions one asks, a partition q_c of modal space. Let $q \leq q'$ just if q is more 'finely grained' than q': just if whenever $P \in q$, for some $P' \in q'$, $P \subseteq P'$—no line drawn through modal space by q cuts across any line drawn by q'. We may then set up this support condition for interrogatives:

14. $\lfloor \omega \rfloor = \{c : q_c \leq [\![\omega]\!]^c\}$

A context c supports an interrogative ω just if in c, one wonders about matters more finely grained than the matters ω concerns. Both here and in the support condition for declaratives, we see that a context supports a sentence just if in it, one has 'already done enough' so that coming anew to accept the sentece would not mark a change in one's psychological position.

We may then establish an expressivist semantics for 'wonderment-avowals':

15. $[\![\text{I wonder } \omega]\!]^c = \{w : \text{in } w, q_c \leq [\![\omega]\!]^c\}$

Namely, the content of 'I wonder whether goats eat cans' relative to c is the set of worlds w at which, in c, one wonders whether goats eat cans.

Note that in this case, the world binder binds vacuously. In such a case, the set abstract can be read along the lines of 'the set of humans h such that roses are red': namely, the set of all humans if roses *are* in fact red, and the empty set if roses are *not* red. Or in the present case, the set of all worlds if c does wonder whether goats eat cans, and the empty set if c does not wonder whether goats eat cans.

As announced, then, this is a case in which a declarative sentence does not look outward from context to the world for its meaning, but stays for its meaning at home, within context. An avowal of wonderment whether goats eat cans has a different meaning from an avowal of wonderment whether lambs eat ivy. But this is not because they encode different information about the world. It is rather because each sentence's pattern of response to the context it is fed differs: the former avowal issues in the set of all worlds as its semantic value just when the context wonders whether goats eat cans, whereas the latter does so just when the context wonders whether lambs eat ivy.

The support condition for a wonderment-avowal is then this:

16. $\lfloor$I wonder $\omega\rfloor = \{c : i_c \subseteq \{w : \text{in } w, q_c \leq [\![\omega]\!]^c\}\}$

That set of worlds is either all of modal space or none of it, depending on whether c wonders ω or does not wonder ω. So whatever i_c may be (assuming nonvacuity), it is a subset of that set of worlds just if c wonders ω. So the support condition of the wonderment-avowal is equivalent to this:

17. $\lfloor$I wonder $\omega\rfloor = \{c : q_c \leq [\![\omega]\!]^c\}$

But that is just the same as the support condition (14) for ω! So on this semantics, 'I wonder ω' and ω are *equivalent* in the sense of possessing the same support condition. They therefore entail one another—which, as it happens, they do in fact seem to.

'I wonder ω' is an interrogative in declarative clothing. Unlike your typical declarative, its 'most basic' semantic function is not to encode information—it is rather to parameterize information worth seeking out. In the traditional sense,

then, this semantics for wonderment avowals is 'expressivist' rather than 'informational'. But note that we have provided a proposition as the content of the wonderment avowal. Accordingly it can interface perfectly cleanly with Boolean connectives—thereby shouldering away the technical aspect of the Chiffre-Geg Embarrassment. As far as its more philosophical side is concerned, though, we are not home free: although we have represented curiosity and opinion as somehow unified through the declarative sentence, we have not yet fully explained what the significance of this unification is.

4 Mindset semantics and our understanding of consciousness

The application motivating this article is a third: using mindset semantics to illuminate the structure of *our understanding* of consciousness. This application will draw on the previous two. It will complete our rebuttal of the semantic case against materialism.

Our target explananda are the sentences (4) and (5)—Q and Z from the conceivability argument—repeated here:

(4) Part of what it is like for Greg is this: G; and part of what it is like for Rebecca is this: R

(5) Part of what it is like for Greg is this: R; and part of what it is like for Rebecca is this: G

Where G is 'this is thus-green' (to understand which one pronounces it while directing attention on a green thing 'qua' green), R is 'this is thus-red' (to understand which one pronounces it while directing attention on a red thing 'qua' red), and so forth.

These sentences have the following key constituents: the 'complement' of 'what it is like is this:', namely G and R; the 'what it is like is this:' prefix itself; and the 'for Greg' part. We address each of these in its own subsection; in the final subsection we sum up.

4.1 This is thus

The sentence 'this is thus' predicates 'is thus' of the term 'this'. We discuss predication in general, then 'this', then 'is thus'.

Predication

We assume that the semantic value of a predicate is a property and that a property is a set of world-time-individual triples. Let T and J are the sets of times and of individuals, respectively. Let $X_1 \times \ldots \times X_n$ is the set of all n-tuples drawn, in order, from the sets $X_1, \ldots, X_n$ through free arbitrary scrambling. Then:

18. $[\![\Gamma]\!]^c \subseteq W \times T \times J$

We assume next that the semantic value of a term is an individual:

19. $[\![v]\!]^c \in J$

A predicate and a term together have enough material by themselves to result in the saturation of the *individual* place in the relevant property, but getting a proposition out the other end requires saturation of the time place as well. For simplicity, we assume that context c registers a time $t_c \in T$ and that the temporal saturation is with this contextually supplied time. Then:

20. $[\![\Gamma(v)]\!]^c = \{w : \langle w, t_c, [\![v]\!]^c \rangle \in [\![\Gamma]\!]^c\}$; where, assuming homophony between object language and metalanguage, we may say that in general, $\langle w, t, j \rangle$ is a member of this semantic value just if in w, at t, j has Γ-ness

The proposition encoded by 'Fred is jolly' against c is the set of worlds such that whoever 'Fred' names—namely, Fred—is, in that world, at the time of c, an instance of that property 'is jolly' designates—namely, *jollity*. So 'Fred is jolly' against the first moment of 2013 is the proposition containing a world just if, in that world, Fred is jolly at the first moment of 2013—namely, the proposition that Fred is jolly at the first moment of 2013.

'This'

We will tie the referent of 'this' to *attention*: not all that is seen is given; rather, givenness requires promotion to consciousness of an excerpt from what is seen. We assume that context c registers an individual $a_c \in J$, where a_c is whatever individual is targeted with attention in the mindset c represents. Then:

21. $[\![\text{this}]\!]^c = a_c$

The denotation of 'this' against c is the target of attention in c.

'Thus'

Contextually evaluated predicates are more complicated than contextually evaluated terms. To see why, consider the following *incorrect* proposal. Suppose we said the exactly analogous thing about 'thus' as we have said about this. That would be to say that context registers a property $F_c \subseteq W \times T \times J$, where F_c is the property 'qua' which attention targets on a_c. We would then provide the following semantic-valuation clause:

22. $[\![\text{Thus}]\!]^c = F_c$

And then, together with (20) and (21), this would yield the following context-relative content for 'this is thus':

23. $[\![\text{this is thus}]\!]^c = \{w : \langle w, t_c, [\![\text{this}]\!]^c \rangle \in [\![\text{Thus}]\!]^c\}$
    $= \{w : \langle w, t_c, a_c \rangle \in F_c\}$

No problem as yet: it does seem as though that is the proposition expressed by 'this is thus'.

But when we feed (23) into our theory of the determination of support conditions for declaratives (9), the predicted support condition for 'this is thus' is implausibly weak:

24. $\lfloor \text{this is thus} \rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c \rangle \in F_c\}\}$

That is, 'this is thus' is supported in c just if the target of attention in c is believed in c to have the feature 'qua' which attention targets it in c.

It is of course true that *if* c supports 'this is thus', c meets that condition. But it does not suit our purposes to say that meeting that condition exhausts the meaning of any given use of 'this is thus'. For consider our examples in (4) and (5): at the very least, Greg's sentence G and Rebecca's sentence R have different meanings. For in accepting G, Greg thereby recognizes that the object he sees is green *and not red*; and in accepting R, Rebecca thereby recognizes that the object she sees is red *and not green*. Greg's and Rebecca's sentences do not entail one another: indeed, they entail one another's *negations*. According to (24), they entail one another. So (24) is wrong. The culprit is (22); so we require an alternative.

The solution cannot be to simply modify (24), by providing a different semantic clause for the predicate 'thus'. The surgery required is more drastic than that. For all occurrences of the same sentence have the same support condition. If we wish Greg's G and Rebecca's R to have differing support conditions, they must therefore be different sentences: Greg's predicate 'thus' and Rebecca's predicate 'thus' must be *orthographically distinct predicates*.

These predicates must, moreover, have the following further properties. Orthographically, they must be 'tied to the perceptual state': it must be the case that Greg's 'thus' is not available to Rebecca, and vice versa. Semantically, the spirit of the direct realism of section 2.2 requires each to subserve the infallibility and certainty of the information possessed in accepting 'this is thus'. And in order to serve the pragmatic demand for distinct support conditions for Greg's and Rebecca's sentences, the semantics must not do so via the contextual saturation of a free parameter.

The apparatus of a 'Lagadonian' language, in which the property 'qua' which the target of attention is so targeted is a *constituent* of the orthographic type of 'thus', satisfies all these desiderata at once. Greg targets an object 'qua' green: so *greenness* is a constituent of the orthographic type of his 'thus' (and *redness* is not). Rebecca targets attention 'qua' red: for her, *redness* is a constituent of the orthographic type of 'thus' (and *greenness* is not). So Rebecca and Greg accept

38

orthographically distinct sentences, as desired. More generally, we may say this:

25. When, in c, attention is targeted 'qua' F, the orthographic type of 'thus' against c includes F-ness (as well, perhaps, as other material): in such a case, we write 'thus$[F]$' as the name of predicates of that orthographic type

Semantically, we may evaluate their predicates as follows:

26. $[\![\text{thus}[F]]\!]^c = F$

This approach secures the direct realist's desired infallibility. Assume that something can be targeted with attention 'qua' F only if it is F; then, in c, one targets attention on a_c 'qua' F only if a_c is F. If one does, then 'this is thus' against c is of the orthographic type 'this is thus$[F]$'. And then, by appeal to (20), (21), and (26):

27. $[\![\text{this is thus}[F]]\!]^c = \{w : \langle w, t_c, [\![\text{this}]\!]^c\rangle \in [\![\text{thus}[F]]\!]^c\}$
    $= \{w : \langle w, t_c, a_c\rangle \in F\}$

Namely, the content against c of 'this is thus' is the proposition that, at t_c, a_c has the property 'qua' which it is targeted with attention—which, as we have just argued, is in fact true.

Finally, assembling (9) and (27) yields the following support condition for 'this is thus' against a context in which attention is targeted 'qua' F:

28. $\lfloor\text{this is thus}\rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c\rangle \in F\}\}$

We observe that, while the identities of the individual and temporal parameters are blurred out, the identity of the property is not. Accordingly, Greg's sentence G and Rebecca's sentence R may fail to share support conditions. And, indeed, if nothing is possibly simultaneously both red and green—if *redness* and *greenness* have a null intersection—then $G \dashv\vdash \neg R$. Whether that is a desirable result is not to the point: what is desirable is the ability to generate that result if necessary.

We are not quite out of the woods. Assume that *F* is *redness*. Then, as should be clear upon inspection, the support condition (28) is the same as the support condition for 'this is red'. But a sentence's support condition is to represent its 'internal' significance. And it would be desirable to predict that 'this is red' and 'this is thus[red]' *differ* in their internal significance. After all, one who sees only in black and white might well accept the former; but we might wish to recognize a sense in which one would yet not accept the latter.

Here I think we have no choice but to appeal to *presupposition*. Let us say that φ presupposes that context is a certain way just if c cannot be updated by *adding* φ to the stock of sentences it accepts unless c already is the way presupposed. An extensive defense and detailed explication of this analysis would take us far afield, but we can briefly illustrate. Return to our question 'have you stopped throwing garbage out of your bedroom window': that question presupposes a shared belief that the audience did in the past have a practice of throwing garbage out of their bedroom window. If the audience does not share this belief, they will refuse to add the question to the stock they 'accept': as, I think, the confused spluttering discussed above may plausibly reflect.

In a case in which φ presupposes that context is among a set $\mathcal{P}$, we can distinguish a stronger and a weaker support condition. The weaker support condition is our standard $\lfloor\varphi\rfloor$. The stronger support condition, which we shall write $\lfloor\lfloor\varphi\rfloor\rfloor$, is $\lfloor\varphi\rfloor \cap \mathcal{P}$. These are generally distinct if the content of a sentence is not in general stronger than that of any belief it presupposes: for example, it may be that the content of 'Fred has stopped throwing garbage from his bedroom window' is the same as that of 'Fred does not have a practice of throwing garbage from his bedroom window'; but they differ in that the former presupposes more than the latter. Inasmuch as they clearly have different meanings, the presupposition of a sentence is part of its meaning, and sentences with the same support condition may yet differ in meaning by having differing strong support conditions: even if $\lfloor\varphi\rfloor = \lfloor\psi\rfloor$, if $\lfloor\lfloor\varphi\rfloor\rfloor \neq \lfloor\lfloor\psi\rfloor\rfloor$, φ and ψ will yet differ in meaning.

Now surely if φ cannot even be entertained by contexts outside of $\mathcal{X}$, no context outside of $\mathcal{X}$ can be updated by adding φ to the stock of sentences accepted at

it. If so, φ presupposes that context is among $\mathcal{X}$. And in that case, whatever $\lfloor \varphi \rfloor$ may be, we can nevertheless be sure that $\lfloor \lfloor \varphi \rfloor \rfloor \subseteq \mathcal{X}$.

We can now assemble our final theory of meaning for Greg's sentence G and Rebecca's sentence R. Their contents and support conditions are as above. But Greg's sentence presupposes that context is among $\mathcal{R}$, the set of those contexts in which attention is targeted 'qua' green, while Rebecca's presupposes that context is among $\mathcal{G}$, the set of those contexts in which attention is targeted 'qua' red. Accordingly:

29. $\lfloor \lfloor G \rfloor \rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c \rangle \in \text{GREENNESS}\}\} \cap \mathcal{G}$
    $\subsetneq \lfloor \lfloor \text{this is green} \rfloor \rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c \rangle \in \text{GREENNESS}\}\}$

30. $\lfloor \lfloor R \rfloor \rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c \rangle \in \text{REDNESS}\}\} \cap \mathcal{R}$
    $\subsetneq \lfloor \lfloor \text{this is red} \rfloor \rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c \rangle \in \text{REDNESS}\}\}$

As desired—and we are done.[6]

Should we regret the move to *strong support* to distinguish the meanings of 'this is thus[red]' and 'this is red'? We really should not. Assigning these sentences different support conditions would acknowledge that they 'logic' differently. Logic seems to care about content, and it seems to care about context-sensitivity. It is hard to see how influences of other sorts could fail to amount to simple performance errors. The distinction between 'this is thus[red]' and 'this is red' seems to be a pure difference in mode of presentation if there ever were one. To inflate it into something larger would unleash a much wider intrusion into matters logical of matters pedagogical, aesthetic, or practical.

Reflections

Let us ascend from the welter of detail to occupy a broader vantage point on what just happened. We have presented an account of the structure of our dis-

---

[6]If attention has extensive internal structure beyond the raw feature qua which it alights—as it almost certainly does—this internal structure may be called upon to further restrict strong support conditions for 'this is thus' sentences. We might imagine that patterns would begin to emerge in the field of these strong support conditions. Those patterns would be something like the Kantian a priori intuitions of space and time—and perhaps also of color, of pitch, ...

tinctively perceptual understanding of perceptible qualities—that understanding which we attain only through finding them in our sensory manifold, and cannot attain through book learning (recall here that the sensory manifold includes not just external features but also our own bodily features and aspects of our perceptual relation to such features, such as *seeing red* or *seeing green with blurry double vision and an afterimage*). This distinctive understanding is widely discussed. It is, I would say, the near-universal practice in this discussion to posit 'phenomenal qualities' to account for this understanding: a creature can only distinctively understand redness, for example, by instantiating 'phenomenal red'. These phenomenal qualities are often thought to be 'intrinsic qualities of experience'—or at least once upon a time they were thought to be so. But that may be optional: those inclined to direct realism might well say phenomenal qualities are relations of attention-based acquaintance with external qualities.

That is not the story we tell. As far as qualities and their pattern of instantiation is concerned, we posit only the sensory manifold. Almost all aspects of that are not targeted with attention, and accordingly it is the way it is largely independently of our possession of any distinctive grasp of it (and given that we could have directed attention otherwise than we did, it is modally entirely independent of this distinctive grasp). If so, our story accommodates our distinctive perceptual understanding without the posit of phenomenal qualities.

4.2 What it's like

We remarked early on that

> Cognition of the non-conscious (as such), then, is the *distinguishing of possibilities* for the non-conscious (from a point of view). Cognition of the *conscious* (as such), by contrast, is the distinguishing of possibilities for the non-conscious *as from a point of view* (whether someone else's, taken up through empathy, or one's own, simply in occupying it).

Our tradition marks this subtle distinction with the language 'what it's like'. No harm in this, I suppose, insofar as it allows us to understand one another—so long

as we do not seek to read too much off of the grammar here.

The Veltman Box

To sidestep the grammatical morass, I will mark the subtle shift of emphasis as follows. Suppose φ has informational meaning: to highlight accepting φ is then to (from one's point of view) *distinguish possibilities for the nonconscious* (as such). Suppose that we then shift emphasis to *as from one's point of view* distinguishing possibilities for the nonconscious (as such). The Frege-Geach problem forces us to have a sentence the acceptance of which we highlight in this latter case. This sentence will be the following:

31. $\Box_v \varphi$

Where we call $\Box_v$ the *Veltman Box*. Acceding in the idiom of our tradition, $\Box_v \varphi$ may then be regarded as an abbreviation for 'here's (part of) what it's like: φ'— though again, that interpretation may be dispensible, for it is the aspect shift that is crucial rather than the language which underlies it.

The Veltman Box has the following semantic valuation clause:

32. $[\![\Box_v \varphi]\!]^c = \{w : i_c \subseteq [\![\varphi]\!]^c\}$

Like our semantic valuation clause for wonderment avowals, this provides the Veltman Box with an *expressive* semantics: it looks for its meaning not outward to the world but inward to context. In the present case, that means $\Box_v \varphi$ distinguishes contexts into those which support φ and those which do not: at the former, $\Box_v \varphi$ has as its content the set of all worlds, while at the latter, $\Box_v \varphi$ has as its content the empty set.

A transparent window

From (32) and (9), we extract the following support condition for $\Box_v \varphi$:

33. $\lfloor \Box_v \varphi \rfloor = \{c : i_c \subseteq \{w : i_c \subseteq [\![\varphi]\!]^c\}\} = \{c : i_c \subseteq [\![\varphi]\!]^c\} = \lfloor \varphi \rfloor$

The support conditions for φ and $\square_\mathsf{v}\varphi$ are the same. That yields this equivalence:

34. $\varphi \dashv\vdash \square_\mathsf{v}\varphi$

Whatever sentence I accept, according to the results in (34), I thereby also accept its embedding under 'here's what it's like'. What I think I find, I think I find within consciousness; and what I think I find within consciousness, I simply also think I find. Consciousness in general is a 'bare container'.

Similarly, it can be easily seen that $\lfloor\neg\varphi\rfloor \subseteq \overline{\lfloor\varphi\rfloor}$: if a context supports the negation of a sentence, it is not among the contexts supporting the sentence. Moreover, because $\lfloor\square_\mathsf{v}\varphi\rfloor = \lfloor\varphi\rfloor$, $\overline{\lfloor\varphi\rfloor} = \overline{\lfloor\square_\mathsf{v}\varphi\rfloor}$; accordingly, $\lfloor\neg\varphi\rfloor \subseteq \overline{\lfloor\square_\mathsf{v}\varphi\rfloor}$. And finally, $\overline{\lfloor\square_\mathsf{v}\varphi\rfloor} = \lfloor\neg\square_\mathsf{v}\varphi\rfloor$.[7] From all this it follows that $\lfloor\neg\varphi\rfloor \subseteq \lfloor\neg\square_\mathsf{v}\varphi\rfloor$:

35. $\neg\varphi \vdash \neg\square_\mathsf{v}\varphi$

By (34), whatever sentence I reject, I thereby also accept the embedding under 'here's what it's like' of its negation; and by (35) I thereby also accept the negation of its embedding under 'here's what it's like'. I never find consciousness to disagree with what I find: if I find something to be not so, I find it not to be the case that, within consciousness, it is also so. Good thing that: otherwise consciousness would contradict itself.

It is not, however, generally the case that $\overline{\lfloor\varphi\rfloor} \subseteq \lfloor\neg\varphi\rfloor$: that if a context fails to support a sentence, it thereby supports the negation of a sentence. Whenever uncertainty about the status of φ is an option—whenever φ encodes information—those contexts which are in fact uncertain about this status will fail to support either φ or $\neg\varphi$. Because, as we have seen, $\overline{\lfloor\varphi\rfloor} = \overline{\lfloor\square_\mathsf{v}\varphi\rfloor} = \lfloor\neg\square_\mathsf{v}\varphi\rfloor$, it follows that it is not generally the case that $\lfloor\neg\square_\mathsf{v}\varphi\rfloor \subseteq \lfloor\neg\varphi\rfloor$. And so:

36. $\neg\square_\mathsf{v}\varphi \nvdash \neg\varphi$

The meta-rule of contraposition fails. According to (36), there is at least room for recognizing uncertainty about matters addressed in objective inquiry.

---

[7]Proof: $\lfloor\neg\square_\mathsf{v}\varphi\rfloor = \{c : i_c \subseteq [\![\neg\square_\mathsf{v}\varphi]\!]^c\} = \{c : i_c \subseteq \overline{[\![\square_\mathsf{v}\varphi]\!]^c}\} = \{c : i_c \subseteq \overline{\{w : i_c \subseteq [\![\varphi]\!]^c\}}\} = \{c : i_c \nsubseteq [\![\varphi]\!]^c\} = \{c : i_c \subseteq \overline{[\![\varphi]\!]^c}\} = \{c : i_c \subseteq \{w : i_c \subseteq \overline{[\![\varphi]\!]^c}\}\} = \{c : i_c \subseteq [\![\square_\mathsf{v}\varphi]\!]^c\} = \lfloor\square_\mathsf{v}\varphi\rfloor$.

Moreover, whenever i_c overlaps both $[\![\varphi]\!]^c$ and $[\![\neg\varphi]\!]^c$—so that c is in fact uncertain about the status of φ—$c \in \lfloor\neg\square_\mathsf{v}\varphi \wedge \neg\square_\mathsf{v}\neg\varphi\rfloor$: c supports 'here's not what it's like: φ; but also here's not what it's like: $\neg\varphi$'; genuine uncertainty is recognized as such. So it is through ignorance that consciousness distinguishes itself from reality. For a way the world might or mightn't be, when consciousness takes a stand on that issue, consciousness does not interpose itself before the world. But when consciousness fails to take a stand, it does interpose itself: the positive failure of consciousness to take a stand intrudes even when no stand one way or the other is present. In that sense, consciousness is found to be a transparent but partial window on the world, with a visible frame.

Self-permeation

The contraposition failures in (36) are restricted to sentences about which uncertainty is an option: to sentences encoding information about matters addressed in objective inquiry. When uncertainty is not an option, support commutes with negation: $\lfloor\neg\varphi\rfloor = \overline{\lfloor\varphi\rfloor}$. But because $\lfloor\varphi\rfloor = \lfloor\square_\mathsf{v}\varphi\rfloor$, it follows that $\lfloor\square_\mathsf{v}\neg\varphi\rfloor = \overline{\lfloor\square_\mathsf{v}\varphi\rfloor}$; and, as we have seen, $\overline{\lfloor\square_\mathsf{v}\varphi\rfloor} = \lfloor\neg\square_\mathsf{v}\varphi\rfloor$; and so $\lfloor\square_\mathsf{v}\neg\varphi\rfloor = \lfloor\neg\square_\mathsf{v}\varphi\rfloor$: $\square_\mathsf{v}$ also commutes with negation.

For example, as we have already seen, $\lfloor\neg\square_\mathsf{v}\psi\rfloor = \overline{\lfloor\square_\mathsf{v}\psi\rfloor}$: it follows that $\lfloor\square_\mathsf{v}\neg\square_\mathsf{v}\psi\rfloor = \lfloor\neg\square_\mathsf{v}\square_\mathsf{v}\psi\rfloor = \overline{\lfloor\square_\mathsf{v}\square_\mathsf{v}\psi\rfloor} = \overline{\lfloor\square_\mathsf{v}\psi\rfloor} = \lfloor\neg\square_\mathsf{v}\psi\rfloor$—what isn't evident within consciousness evidently isn't so. And because $\lfloor\varphi\rfloor = \lfloor\square_\mathsf{v}\varphi\rfloor$, it follows also that $\lfloor\square_\mathsf{v}\psi\rfloor = \lfloor\square_\mathsf{v}\square_\mathsf{v}\psi\rfloor$: what is evident within consciousness evidently is so. Consciousness is absolutely clear to itself: in that sense, consciousness is *self-permeating*.

The Chiffre-Geg Embarassment

For another example, let φ be 'I wonder ω':

37. (a) $\lfloor$I wonder $\omega\rfloor = \{c : i_c \subseteq [\![$I wonder $\omega]\!]^c\}$
     $= \{c : i_c \subseteq \{w : q_c \leq [\![\omega]\!]^c\}\} = \{c : q_c \leq [\![\omega]\!]^c\} = \lfloor\omega\rfloor$

(b) $\lfloor\neg(\text{I wonder } \omega)\rfloor = \{c : i_c \subseteq \llbracket\neg(\text{I wonder } \omega)\rrbracket^c\}$

$= \{c : i_c \subseteq \overline{\{w : q_c \preceq \llbracket\omega\rrbracket^c\}}\} = \{c : q_c \npreceq \llbracket\omega\rrbracket^c\} = \overline{\lfloor\omega\rfloor} = \overline{\lfloor\text{I wonder } \omega\rfloor}$

And so, when it operates on 'I wonder ω', $\Box_v$ also commutes with negation:

(c) $\lfloor\neg\Box_v(\text{I wonder } \omega)\rfloor = \lfloor\Box_v\neg(\text{I wonder } \omega)\rfloor$

What I wonder about is settled within consciousness with absolute clarity.

This resolves the Chiffre-Geg Embarassment: the 'wonderment machine' is unified with the 'belief machine' by being wrapped up under consciousnesness. While the 'semantic form' of curiosity is most fundamentally encoded in interrogatives, curiosity can also be encoded in declarative wonderment avowals. The reason for this is the fundamentally assertoric character of consciousness: *this is what it's like*, consciousness insists. It is through this processing into wonderment avowals that consciousness is able to perform the relevant insistence with regard to one's curiosity—and therefore that one's curiosity is brought under the umbrella of unified consciousness.

What it's like: this is thus

Recall the support and strong support conditions discussed in the previous section for 'this is thus'-sentences:

(28) $\lfloor G\rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c\rangle \in \text{GREENNESS}\}\} = \lfloor\text{this is green}\rfloor$

$\lfloor R\rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c\rangle \in \text{REDNESS}\}\} = \lfloor\text{this is red}\rfloor$

(29) $\lfloor\lfloor G\rfloor\rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c\rangle \in \text{GREENNESS}\}\} \cap \mathcal{G}$

$\subsetneq \lfloor\lfloor\text{this is green}\rfloor\rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c\rangle \in \text{GREENNESS}\}\} = \lfloor G\rfloor$

(30) $\lfloor\lfloor R\rfloor\rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c\rangle \in \text{REDNESS}\}\} \cap \mathcal{R}$

$\subsetneq \lfloor\lfloor\text{this is red}\rfloor\rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c\rangle \in \text{REDNESS}\}\} = \lfloor R\rfloor$

By the results of this section together with (28), the support conditions for 'here's what it's like: this is thus' look as follows:

38. (a) $\lfloor \Box_v G \rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c \rangle \in \text{GREENNESS}\}\} = \lfloor \text{this is green} \rfloor$

   (b) $\lfloor \Box_v R \rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c \rangle \in \text{REDNESS}\}\} = \lfloor \text{this is red} \rfloor$

If that is all there is to be said about the meanings of these sentences, we would be in big trouble: Greg and Rebecca are paradigms of distinctiveness in what it's like; no credible theory could equate what it is like for them with what it is like for a color-blind person thinking the same thing.

Of course, this surely is at least *part* of what these sentences mean: after all, we do find common cause in our picture of the world with the color-blind. We understand things to be the same way and ourselves to be similarly located, even if our local internal reactions to things may be distinct. The superficial difference in 'aroma' should not blind us to the deep commonality of form. In light of all of our manifest internal differences, a Nagel-type thesis of the incomprehensibilty of the other would be hard to resist were we to focus only on the superficial.

But we still need to generate a *difference* between the meanings of 'here's what it's like: this is thus[red]' and 'here's what it's like: this is red'. It would be straightforward to do this if the presuppositions of φ and $\Box_v \varphi$ are the same: in that case, $\lfloor\lfloor \varphi \rfloor\rfloor = \lfloor\lfloor \Box_v \varphi \rfloor\rfloor$, so that $\lfloor\lfloor \Box_v G \rfloor\rfloor = \lfloor\lfloor G \rfloor\rfloor \neq \lfloor\lfloor \text{this is red} \rfloor\rfloor = \lfloor\lfloor \Box_v(\text{this is red}) \rfloor\rfloor$. Are they?

Note that, if c updates with $\Box_v \varphi$, the result collapses into defectiveness if c does not support φ. In that sense, $\Box_v \varphi$ presupposes that c accepts φ. But that is a weaker claim than we need, which is that 'here's what it's like: φ' presupposes that c accepts φ *and satisfies all presuppositions of φ*. What we need is for the presuppositions of φ to 'project' through 'here's what it's like'. That they do is quite plausible in light of the job we have assigned for 'here's what it's like'. For surely if Fred does not believe he ever threw garbage out of his bedroom window, he cannot also accept that part of what it's like for him is that he has stopped throwing garbage out of his bedroom window. For he would then be just an aspect shift away from accepting that he has stopped throwing garbage out of his bedroom window—which would be out of coherence with his failure to affirm that he ever did so. (Similarly, it seems as if for 'Moorean' psychological verbs such as 'believe', avowals of belief project the presuppositions of their complements:

'I believe Fred has stopped throwing garbage from his bedroom window' presupposes that Fred used to do so. So this projection would be in line with a more generally observed pattern.)

So, in particular, c is in the strong support condition of 'here's what it's like: R' only if, in c, attention is targeted 'qua' red.

The presupposition seems also to be part of what it is like. We might capture that by introducing further expressive vocabulary, along the following lines:

39. $[\![$I target attention 'qua' $\Gamma]\!]^c = \{w : $ in c, attention is targeted 'qua' $[\![\Gamma]\!]^c\}$

40. $\lfloor$I target attention 'qua' $\Gamma\rfloor = \{c : i_c \subseteq \{w : $ in c, attention is targeted 'qua' $[\![\Gamma]\!]^c\}\} = \{c : $ in c, attention is targeted 'qua' $[\![\Gamma]\!]^c\}$

In that case, the strong support condition of 'here's what it's like: this is thus[red]' is a subset of the support condition of 'I target attention 'qua' red', and therefore of the support condition of 'here's what it's like: I target attention 'qua' red'.

4.3 For Fred

As we have framed it, 'here's what it's like: φ' does not have an argument place for a subject. This is in line with the way we have been speaking: 'it is within consciousness that blah blah', 'it is not within consciousness that yada yada', and so forth. We have not been saying, by contrast, 'it is within consciousness *for me* that blah blah' or 'it is not within consciousness *for me* that yada yada'.

We have been speaking this way hoping to preserve the phenomenological aspect of things. Phenomenologically, there is no subject here. Mentioning a subject in a story that attempts to characterize the structure of our understanding of consciousness would attribute to this understanding grain that is not legitimately a part of it.

The claim that, phenomenologically, there is no subject here has been made before, but what does it mean? The claim is hard to grasp without a sense for what is being denied. So let us ask what it would be like for there to *be* a subject here.

For there to be a subject here would be for assessment of consciousness to be a kind of *predication*: that in assessing what it is like as being thus and so, I would be judging concerning this thing, that it is that way. For example, when I want to say of something that it is red, or that it is round, or that it is an elephant, I need to select which thing it is that I am going to discuss. *That apple* is red; *this painting* is round; *Olga* is an elephant. Predication and topicalization run together. Classification requires a classificandum.

I do not, by contrast, need to select which consciousness it is that I am to classify in order to assess what it is like. There is only one to choose from: mine. I never encounter any other consciousness than this one. What would it even *be* to do that? The idea boggles the mind.

One might object that my concern squeezes the data too hard. Evidently I can *in some sense* cognize other consciousnesses, even if I cannot 'encounter' them in that special manner in which I 'encounter' this one. Perhaps then my concern inflates a 'defaulting' on encountering *this* consciousness into an *impossibility* of encountering other consciousnesses. Perhaps when there is no explicit 'for Fred' specification of an argument, 'here's what it's like: φ' defaults to evaluation against the subject of this context, where this evaluation can be made explicit by specifying the argument with 'for me'.

More explicitly. The proposal is that 'here's what it's like for v: φ' has the logical form $[W : \varphi](v)$, while 'here's what it's like: φ' has the logical form $[W : \varphi](I)$. Here 'I' supplies the silent subject, and acts as the first-person pronoun: $[\![I]\!]^c = s_c$, the 'subject of experience' for c.

The question, then, is which property $[\![[W : \varphi]]\!]^c$ is supposed to be. Assuming the data of the previous subsection is correct, the following are constraints. For φ informational:

41. (a) $\lfloor \varphi \rfloor = \lfloor [W : \varphi](I) \rfloor$

    (b) $\lfloor \neg\varphi \rfloor \subsetneq \lfloor \neg[W : \varphi](I) \rfloor$

I can't think how to do this unless

42. $[\]\!]^c = \{w : i_c \subseteq [\![\varphi]\!]^c\} = \{w : \langle w, t_c, s_c \rangle \in [\![[W : \varphi]]\!]^c\}$

So that, in particular, if $i_c \subseteq [\![\varphi]\!]^c$, then for every w, $\langle w, t_c, s_c \rangle \in [\![[\mathsf{W} : \varphi]]\!]^c$, while if $i_c \nsubseteq [\![\varphi]\!]^c$, then for no w is $\langle w, t_c, s_c \rangle \in [\![[\mathsf{W} : \varphi]]\!]^c$. Then $\mathfrak{W} = (\lambda S)(\lambda c)[\![[\mathsf{W} : S]]\!]^c$ is a function mapping a proposition and a context into a property. Applied to the proposition that goats eat cans, the function $\mathfrak{W}$ maps a context into a property the subject of the context has necessarily at the time of the context just if the subject believes that goats eat cans, and that the subject could not then possibly have otherwise.

Of course, that massively underspecifies the nature of $\mathfrak{W}$. But a cursory examination of the most basic parameters any specification would have to settle on reveals that no answer will adequately capture the initial idea behind the approach: that in assessing the distinctive character of consciousness, I somehow find myself as one subject among many, and then categorize myself as having some features and lacking others.

The parameters in question are these. Fixing a context c and a proposition P, which proposition is the output when $\mathfrak{W}(c, P)$ is applied to a subject and time distinct from those of the context—Harlan Sanders at the first moment of 1960, perhaps? And is the answer the same when, for $c' \neq c$, $\mathfrak{W}(c, P)$ is applied to Harlan Sanders then? If the proposition is contingent, $[\mathsf{W} : \varphi]$ would seem to be a 'semantic gerrymander': it behaves expressively at home but informatively abroad. It is hard to believe that our understanding of consciousness could be so bizarrely disunified. And recall that we are attempting to preserve the idea that in reflection on consciousness, I 'find myself' as but one subject among many, and then go on to just sort myself together with some and apart from others through relatively ordinary strategies of predication. The 'semantic gerrymander' strategy is not faithful to this motivating idea. I 'classify' others using a constant strategy but myself using an inconstant strategy; it is bad theory to stretch vague notions such as 'classification' to a point at which they lose elasticity. And my radically distinct classificatory strategy regarding myself as against all others suggests that any sense in which I 'find' both self and others is again a sense drained of all connection to the ideas that got us started. (Note that when we think about the past and the future as against the present, the *property* of times does not change

50

with the time: only the time against which the property is set does.)

We could avoid the semantic gerrymander approach by allowing [W : φ] to just be purely expressive. That is fine as far as predictions go: our approach too will be purely expressive, and as far as predictions go it does not matter where the parentheses are found in the guts of the theory. Still, the approach seems unnaturally complex, parasitic on our approach, and out of line with the 'aspect shift' story we told to motivate the view. I don't 'find' myself when thinking just about the cuisine of Italy; why assume I pop up when I shift aspect to taking a point of view on the cuisine of Italy? And given that mindset semantics has extensive independent motivation, and allows the relatively trivial definition of the Veltman Box, which yields the transparency predictions right away, and does so without raising further questions about what happens beyond me now—and given that the aim of semantics is to illuminate the structure of consciousness, and given that theoretical activity in general prefers to avoid positing unmotivated complexity: well, if we have the option *not* to introduce subjects of experience, shouldn't we take the option?

How then do we avoid introducing subjects of experience? Framing the technical specification more directly, it is this. We want 'here's what it's like: φ' and 'here's what it's like *for Fred*: φ' to have extensive 'semantic overlap'—to avoid, as it were, 'changing the subject' between the former and the latter. But this overlap can't involve a common predicate of distinct subjects, as we have argued; and more generally, we should not think of both as distinct species of a common semantic genus. The alternative would be to think of 'here's what it's like: φ' as *semantically basic* and 'here's what it's like for Fred: φ' as *semantically derivative*: that, in some sense, the entirety of the meaning of the former is present in the latter—but rather than in its pure form, perhaps as somehow 'adjusted' by additional semantic material.

The device that does this sort of work is the *index-shifter*. In general terms, an index-shifter S acts like this:

43. $[\![S^\pi \varphi]\!]^c = [\![\varphi]\!]^{F(c,p)}$

     • Where $p = [\![\pi]\!]^c$

That is, S takes an internal argument π and makes a sentential operator. The effect of this operator is to adjust the context against which the embedded sentence is evaluated from the 'home' context c to a different context c'. The features of this new context c' are the product of three factors: the features of the home context c; the general activity F associated with S; and whatever the distinctive contribution of the parameter p, determined as the semantic value against c of the expression π. An index-shifted sentence $S^{\pi}\varphi$ introduces parametrization, by varying π, into a family of sentences including the operand φ while retaining the semantic basicness to that family of the operand.

Let us illustrate. We have been assuming (solely for convenience) that a subject-predicate sentence Γv has 'natural present tense': against a context, the time parameter in the semantic value of Γ is saturated by the time *of the context*. 'Sam is clever', against c, encodes the proposition that, at t_c, the person named 'Sam' has the feature called 'cleverness'. But we might want to say things like 'at t, blah', to get out of the present. One approach would be to introduce an index-shifter $[\text{AT}\,\tau]$, which binds the time parameter in the predicate of its operand: which supplies a time named by τ to that parameter 'internally', screening off the attempt by context to supply that value. Formally:

44. $[\![[\text{AT}\,\tau]\,\varphi]\!]^c = [\![\varphi]\!]^{\text{T}(c,t^*)}$

    (a) $t^* = [\![\tau]\!]^c$

    (b) Where $c' = \text{T}(c,t^*)$: $t_{c'} = t^*$; otherwise $x_{c'} = x_c$

45. Example:

    (a) $[\![[\text{AT noon, the Queen's 75th birthday}](\text{Fred yawns})]\!]^c = [\![\text{Fred yawns}]\!]^{c'} = \{w : \langle w, t_{c'}, \text{Fred}\rangle \in [\![\text{yawns}]\!]^c\}$—namely, the proposition that Fred yawns at $t_{c'}$:

    (b) Where $t_{c'}$ is $[\![\text{noon, the Queen's 75th birthday}]\!]^c$—namely, noon on the Queen's 75th birthday; accordingly

    (c) Against c, 'Fred yawns at noon on the Queen's 75th birthday' encodes the proposition that Fred yawns at noon on the Queen's 75th birthday

So we note that an 'at-shifted' sentence against c does not look to c to provide the moment it concerns: rather, that moment is provided by the parameter embedded within the index-shifter. So the family including φ and its 'at-shifted' embedding sentences both involves explicit parametrization and in some sense recognizes the semantic primacy of φ. 'At-shifting' allows our minds to range throughout time even while our feet stay firmly rooted in the present.

We want to use a 'for-shifting' operator to accommodate talk about other minds: $\text{for}^\sigma \varphi$ would allow us, as it were, to express things 'from the point of view of σ', as when we say 'here's what it's like *for Fred*: blah blah blah'. An approach precisely analogous to the one developed for $[\text{AT } \tau]$ runs as follows.

46. $[\![\text{for}^\sigma \varphi]\!]^c = [\![\varphi]\!]^{S(c,s^*)}$

   (a) $s^* = [\![\sigma]\!]^c$

   In particular, assume s^* to be a *creature*, in line with materialism as the best candidates for subjects

   (b) $t_{S(c,s^*)} = t_c$

   This enables the speaker to have control over which stage of the creature at issue is under consideration: without wide scope 'at-shifting', the default present tense means that the present stage is under consideration ('here's what it's like for Fred'—what it *is* like, as in *now*); temporally distinct stages may be considered through at-shifting ('at t', here's what it's like for Fred'—in the sense of what it is/was/will be like, as t' is present/past/future)

   (c) $s_{S(c,s^*)} = s^*$

   Namely, S moves the subject of the shifted context from the subject of the home context over to the referent of σ

   (d) Otherwise, $x_{S(c,s^*)} = x_c$

   Namely, S moves nothing else

A question to flag for the moment: what is the nature of the context S extracts from c and s^*, and how is it related to c and s^*? Set that aside for a

bit.

47. **Notational convention:** when σ is the internal argument of for, let $c' = S(c, [\![\sigma]\!]^c)$

48. Example:

    (a) $[\![\text{for}^{\text{Fred}}(\text{I yawn})]\!]^c = [\![\text{I yawn}]\!]^{c'} = \{w : \langle w, t_c, s_{c'} \rangle \in [\![\text{yawn}]\!]^{c'}\}$—namely, the proposition that $s_{c'}$ (the subject of c') yawns at t_c:

    (b) Where $s_{c'}$ is $[\![\text{Fred}]\!]^c$—namely, Fred; accordingly

    (c) Against c, 'for Fred, I yawn' encodes the proposition that Fred yawns at t_c, the time of c

49. More germane example:

    (a) $[\![\text{for}^{\text{Fred}}\Box_{\text{v}}(\text{I yawn})]\!]^c = [\![\Box_{\text{v}}(\text{I yawn})]\!]^{c'} = \{w : i_{c'} \subseteq [\![\text{I yawn}]\!]^{c'}\}$
        $= \{w : i_{c'} \subseteq \{w : \langle w, t_c, s_{c'} \rangle \in [\![\text{yawn}]\!]^c\}\}$ (which, because $i_c = i_{c'}$)
        $= \{w : i_c \subseteq \text{the proposition that } s_{c'} \text{ yawns at } t_c\}$
        $= \{w : i_c \subseteq \text{the proposition that Fred yawns at } t_c\}$

    (b) Accordingly, against c, 'for Fred, here's what it's like: I now yawn' has as its semantic value the set of all worlds if **according to c,** Fred now yawns and otherwise the set of no worlds

This approach is bad. First, it generates 'monsters begat by elegance'. Whether or not this is in itself so dreadful, it is a side effect in which we had no interest. But more importantly, the approach just does *not* do what we most wanted: it assigns as its meaning to 'for Fred, here's what it's like: I yawn' what should *at best* be the meaning of 'here's what it's like: for Fred: I yawn' (even setting aside the worry over monsters begat by elegance).

Let us take another whack at it. We make the minimal revision needed to push the information state considered by the Veltman Box away from the home context and over to a context somehow associated with the subject s^*:

50. $[\![\text{for}^\sigma \varphi]\!]^c = [\![\varphi]\!]^{S(c, s^*) = c'}$

(a) $s^* = [\![\sigma]\!]^c$

(b) $t_{c'} = t_c$

(c) $i_{c'} = i_{cs^*}$

Namely, S moves the belief state of the shifted context from the belief state of the home context over to the belief state of the context of the referent of σ

(d) Otherwise, $x_{c'} = x_c$

Namely, S moves nothing else

51. Example:

(a) $[\![\text{for}^{\text{Fred}}(\text{I yawn})]\!]^c = [\![\text{I yawn}]\!]^{c'} = \{w : \langle w, t_c, s_c\rangle \in [\![\text{yawn}]\!]^c\}$—namely, the proposition that s_c yawns at t_c; accordingly,

(b) Against c, 'for Fred, I yawn' encodes the proposition that s_c (the subject of c) yawns at t_c (the time of c)

52. More germane example:

(a) $[\![\text{for}^{\text{Fred}}\Box_{\mathsf{v}}(\text{I yawn})]\!]^c = [\![\Box_{\mathsf{v}}(\text{I yawn})]\!]^{c'} = \{w : i_{c'} \subseteq [\![\text{I yawn}]\!]^{c'}\}$
$= \{w : i_{c'} \subseteq \{w : \langle w, t_{c'} = t_c, s_{c'} = s_c\rangle \in [\![\text{yawn}]\!]^{c'}\}\}$
$= \{w : i_{c'} \subseteq \text{the proposition that } s_c \text{ yawns at } t_c\}$

(b) Accordingly, 'for Fred, here's what it's like: I yawn' has as its semantic value against c the set of all worlds if **according to Fred's context,** the subject of c yawns at the time of c and otherwise the set of no worlds

53. More intricate example:

(a) $[\![\text{for}^{\text{Fred}}\Box_{\mathsf{v}}(\text{I wonder } \omega)]\!]^c = [\![\Box_{\mathsf{v}}(\text{I wonder } \omega)]\!]^{c'} = \{w : i_{c'} \subseteq [\![\text{I wonder } \omega]\!]^{c'}\}$
$= \{w : i_{c'} \subseteq \{w : q_{c'} \leq [\![\omega]\!]^{c'}\}\} = \{w : q_{c'} \leq [\![\omega]\!]^{c'}\} = \{w : q_c \leq [\![\omega]\!]^{c'}\}$

(b) Accordingly, 'for Fred, here's what it's like: I wonder whether goats eat cans' has as its semantic value against c the set of all worlds if **the**

> **subject of** c wonders whether goats eat cans and otherwise the set of no worlds (here we assume 'do goats eat cans' means the same against Fred's context as against ours)

This approach remains somewhat bad. It improves on the previous approach. It gets rid of the 'monstrosity' of 'for Fred, I now yawn'. And it does manage to key 'for Fred, here's what it's like: I now yawn' over to Fred's picture of the world. But it also fails to capture *either* Fred's situated sense of the world as centered on creature Fred *or* Fred's curiosity as bound together with his opinions. There is also the unanswered question of what 'Fred's context' is supposed to be.

Continuing to set that question aside, we will propose a simple fix which gets rid of the monsters begat by elegance *outside* the scope of the Veltman Box while also simultaneously binding all contextual parameters *inside* its scope. Namely, we treat 'for Fred' not as a sentential operator but rather as a sentential-*operator* operator. Specifically, we locate the semantic action of 'for Fred' in affecting a modification on the semantic action of *the Veltman Box*, which then goes to work on its sentential complement, rather than in affecting a modification on the sentence fronted by the Veltman Box. The details here involve some slightly intricate manoeuvring with the lambda-calculus, and we will skip over them for the final result:

54. $[\![[\text{for}^\sigma(\Box_v)]\varphi]\!]^c = \{w : i_{S(c, [\![\sigma]\!]^c)} \subseteq [\![\varphi]\!]^{S(c, [\![\sigma]\!]^c)}\}$

    Recalling that $c' = S(c, [\![\sigma]\!]^c)$:

    (a) Let $t_{c'} = t_c$

    (b) Otherwise, where $[\![\sigma]\!]^c = s^*$ and where c^* is the context s^* occupies at t_c (if any), then let $x_{c'} = x_{c^*}$

    Notational convention: in that case, for t_c understood, let c^{s^*} be c': for example, c^{Fred} is Fred's context at the understood time

    Consequences: $s_{c^{s^*}} = s^*$; $c^{s_c} = c$

55. Example:

- $[\![\mathsf{for}^{\mathrm{Fred}}(\mathrm{I\ yawn})]\!]^c$

  —undefined, because $\mathsf{for}^{\mathrm{Fred}}(\mathrm{I\ yawn})$ is not a sentence; and this, in turn, is because $\mathsf{for}^{\mathrm{Fred}}$ is not a sentential operator but a sentential-operator operator

56. More germane example:

   (a) $[\![[\mathsf{for}^{\mathrm{Fred}}(\square_{\mathsf{v}})](\mathrm{I\ yawn})]\!]^c = \{w : i_{c^{\mathrm{Fred}}} \subseteq [\![\mathrm{I\ yawn}]\!]^{c^{\mathrm{Fred}}}\}$
   $= \{w : i_{c^{\mathrm{Fred}}} \subseteq \{w : \langle w, t_c, s_{c^{\mathrm{Fred}}} \rangle \in [\![\mathrm{yawn}]\!]^{c^{\mathrm{Fred}}}\}\}$
   $= \{w : i_{c^{\mathrm{Fred}}} \subseteq$ the proposition that $s_{c^{\mathrm{Fred}}} = \mathrm{Fred\ yawns\ at\ } t_c\}$

   (b) Accordingly, 'for Fred, here's what it's like: I yawn' has as its semantic value against c the set of all worlds if **according to Fred's context** (at the time of c), Fred then yawns, and otherwise the set of no worlds

57. More intricate example:

   (a) $[\![[\mathsf{for}^{\mathrm{Fred}}(\square_{\mathsf{v}})](\mathrm{I\ wonder\ }\omega)]\!]^c = \{w : i_{c'} \subseteq [\![\mathrm{I\ wonder\ }\omega]\!]^{c'}\}$
   $= \{w : i_{c'} \subseteq \{w : q_{c'} \leq [\![\omega]\!]^{c'}\}\} = \{w : q_{c'} \leq [\![\omega]\!]^{c'}\}$
   $= \{w : q_{c^{\mathrm{Fred}}} \leq [\![\omega]\!]^{c^{\mathrm{Fred}}}\}$

   (b) Accordingly, against c, 'for Fred, here's what it's like: I wonder whether goats eat cans' has as its semantic value against c the set of all worlds if, at the time of c, **Fred** wonders whether goats eat cans and otherwise the set of no worlds

This iteration of the theory seems to make the appropriate predictions for the cases at hand. So let us consider its predictions in a broader scope.

We can recover the equivalence of 'here's what it's like for me: φ' with 'here's what it's like: φ', as follows:

58. $[\![[\mathsf{for}^{\mathrm{I}}(\square_{\mathsf{v}})]\varphi]\!]^c = \{w : i_{c([\![\mathrm{I}]\!]^c)} \subseteq [\![\varphi]\!]^{c([\![\mathrm{I}]\!]^c)}\} = \{w : i_{c^{s_c}} \subseteq [\![\varphi]\!]^{c^{s_c}}\}$
   $= \{w : i_c \subseteq [\![\varphi]\!]^c\} = [\![\square_{\mathsf{v}}\varphi]\!]^c$

The argument uses the thesis that $c^{s_c} = c$: the context associated with the subject of context c is just c itself.

The support condition of 'for Fred, here's what it's like: φ' is this:

59. $\lfloor[\text{for}^{\text{Fred}}(\square_v)]\varphi\rfloor = \{c : i_c \subseteq [\![[\text{for}^{\text{Fred}}(\square_v)]\varphi]\!]^c\}$

   $= \{c : i_c \subseteq \{w : i_{c^{\text{Fred}}} \subseteq [\![\varphi]\!]^{c^{\text{Fred}}}\}\} = \{c : i_{c^{\text{Fred}}} \subseteq [\![\varphi]\!]^{c^{\text{Fred}}}\}$

Here, it is the turn of the *context* variable to bind vacuously. That is to say that the support condition revealed includes either *every* context, if Fred's context does in fact support φ, or *no* context, if not:

60. (a) If $c^{\text{Fred}} \in \lfloor\varphi\rfloor$, $\lfloor[\text{for}^{\text{Fred}}(\square_v)]\varphi\rfloor = C$

   (b) If $c^{\text{Fred}} \notin \lfloor\varphi\rfloor$, $\lfloor[\text{for}^{\text{Fred}}(\square_v)]\varphi\rfloor = \varnothing$

This suggests that, at some level, there can be no reasonable doubt about what it is like for Fred: that, at worst, ascertainment of what it is like for Fred is like ascertainment of the status of a difficult mathematical claim: perhaps a calculational challenge, but ultimately not a matter about which there might be uncertainty. That prediction might seem to be entirely outrageous. But as we will see, properly understood, it actually turns out to be precisely in line with the *Verstehen* tradition.

Before getting to this, let us push forward with logical investigations. The support condition for φ is, of course, this:

(9) $\lfloor\varphi\rfloor = \{c : i_c \subseteq [\![\varphi]\!]^c\}$

Now, in case (60a), $\lfloor\varphi\rfloor \subseteq \lfloor[\text{for}^{\text{Fred}}(\square_v)]\varphi\rfloor$, whereas in case (60b), $\lfloor\varphi\rfloor \subseteq \lfloor\neg[\text{for}^{\text{Fred}}(\square_v)]\varphi\rfloor$. In general, however, the converses will not hold. Indeed, for that matter, in (60a), $\lfloor\psi\rfloor \subseteq \lfloor[\text{for}^{\text{Fred}}(\square_v)]\varphi\rfloor$; while in (60b), $\lfloor\psi\rfloor \subseteq \lfloor\neg[\text{for}^{\text{Fred}}(\square_v)]\varphi\rfloor$. So assuming contingent φ and ψ:

61. (a) When (60a), $\varphi \nvdash [\text{for}^{\text{Fred}}(\square_v)]\varphi$

   - Indeed $\psi \nvdash [\text{for}^{\text{Fred}}(\square_v)]\varphi$

   (b) While when (60b), $\varphi \nvdash \neg[\text{for}^{\text{Fred}}(\square_v)]\varphi$

   - Indeed $\psi \nvdash \neg[\text{for}^{\text{Fred}}(\square_v)]\varphi$

These are not especially 'useful' entailments. There is certainly nothing 'formal' about them, after all. They seem to appear and disappear entirely capriciously as Fred's epistemic position changes, paying no heed to generalities of meaning. Does this suggest that our apparatus sheds too little light on the phenomenon, and should be chucked out?

A dot of light is visible up ahead. If we allow the parameter in the 'for' operator to float freely, we observe the following—regardless of whatever Fred's situation pertaining to φ may be:

62. $\lfloor [\text{for}^{\sigma}(\square_{\mathsf{v}})]\varphi \rfloor = \{c : i_c \subseteq [\![[\text{for}^{\sigma}(\square_{\mathsf{v}})]\varphi]\!]^c\}$

$\quad = \{c : i_c \subseteq \{w : i_{c^{([\![\sigma]\!]^c)}} \subseteq [\![\varphi]\!]^{c^{([\![\sigma]\!]^c)}}\}\} = \{c : i_{c^{([\![\sigma]\!]^c)}} \subseteq [\![\varphi]\!]^{c^{([\![\sigma]\!]^c)}}\}$

$\quad = \{c : i_{\varsigma} \subseteq [\![\varphi]\!]^{\varsigma}\}$

Writing ς for $c^{([\![\sigma]\!]^c)}$.

And because, varying σ freely, we observe no general pattern of inclusion between the support conditions in (9) and (62), we observe also the following patterns of nonentailment:

63.   (a) $\varphi \nVdash \nVDash [\text{for}^{\sigma}(\square_{\mathsf{v}})]\varphi$

      (b) $\varphi \nVdash \nVDash \neg[\text{for}^{\sigma}(\square_{\mathsf{v}})]\varphi$

This suggests that the capriciousness observed in (61) results from excessive specificity in the meaning of σ. Perhaps the sort of bluriness of meaning associated with supervaluations is generally a significant part of our understanding of others. But again, more below.

Even though schematizing σ gets rid of undesirable entailments here, we do still preserve the desirable entailments discussed in the previous subsection. Appealing to (62),

64.   (a) $\lfloor [\text{for}^{\sigma}(\square_{\mathsf{v}})]\square_{\mathsf{v}}\varphi \rfloor = \{c : i_{\varsigma} \subseteq [\![\square_{\mathsf{v}}\varphi]\!]^{\varsigma}\} = \{c : i_{\varsigma} \subseteq \{w : i_{\varsigma} \subseteq [\![\varphi]\!]^{\varsigma}\}\}$

          $= \{c : i_{\varsigma} \subseteq [\![\varphi]\!]^{\varsigma}\} = \lfloor [\text{for}^{\sigma}(\square_{\mathsf{v}})]\varphi \rfloor$

      (b) $\lfloor [\text{for}^{\sigma}(\square_{\mathsf{v}})]\neg\varphi \rfloor = \{c : i_{\varsigma} \subseteq \overline{[\![\varphi]\!]^{\varsigma}}\}$

          $\subsetneq \overline{\{c : i_{\varsigma} \subseteq [\![\varphi]\!]^{\varsigma}\}} = \overline{\lfloor [\text{for}^{\sigma}(\square_{\mathsf{v}})]\varphi \rfloor}$

          —the inequality holds generally for contingent φ

59

(c) $\lfloor\neg[\text{for}^\sigma(\Box_v)]\varphi\rfloor = \{c : i_c \subseteq \llbracket\neg[\text{for}^\sigma(\Box_v)]\varphi\rrbracket^c\}$

$\quad = \{c : i_c \subseteq \overline{\{w : i_\varsigma \subseteq \llbracket\varphi\rrbracket^\varsigma\}}\} = \{c : i_\varsigma \not\subseteq \llbracket\varphi\rrbracket^\varsigma\} = \overline{\{c : i_\varsigma \subseteq \llbracket\varphi\rrbracket^\varsigma\}}$

$\quad = \overline{\lfloor[\text{for}^\sigma(\Box_v)]\varphi\rfloor}$

65. So we do observe such patterns of entailment and nonentailment as

(a) $[\text{for}^\sigma(\Box_v)]\varphi \dashv\vdash [\text{for}^\sigma(\Box_v)]\Box_v\varphi$

(b) $[\text{for}^\sigma(\Box_v)]\neg\varphi \not\dashv\vdash [\text{for}^\sigma(\Box_v)]\neg\Box_v\varphi$

(c) $[\text{for}^\sigma(\Box_v)]\neg\Box_v\varphi \dashv\vdash [\text{for}^\sigma(\Box_v)]\neg\Box_v\Box_v\varphi$

So the unadorned Veltman Box behaves just the same way within and without the scope of $[\text{for}^\sigma(\Box_v)]$: consciousness is equally as transparent abroad as at home.

In application to our sentences R and G, here is how the apparatus rolls out. Recall the strong support conditions of these sentences:

(29) $\lfloor\lfloor G\rfloor\rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c\rangle \in \text{GREENNESS}\}\}\cap\{c : \text{in } c, \text{ attention is targeted 'qua' green}\}$

(30) $\lfloor\lfloor R\rfloor\rfloor = \{c : i_c \subseteq \{w : \langle w, t_c, a_c\rangle \in \text{REDNESS}\}\}\cap\{c : \text{in } c, \text{ attention is targeted 'qua' red}\}$

Accordingly, assuming that 'for-shifters' interact straightforwardly with presupposition, these yield the following strong support conditions:

66. $\lfloor\lfloor[\text{for}^\sigma(\Box_v)]G\rfloor\rfloor$

$\quad = \{c : i_\varsigma \subseteq \{w : \langle w, t_\varsigma = t_c, a_\varsigma\rangle \in \text{GREENNESS}\}\}$

$\quad \cap \{c : \text{in } \varsigma, \text{ attention is targeted 'qua' green}\}$

67. $\lfloor\lfloor[\text{for}^\sigma(\Box_v)]R\rfloor\rfloor$

$\quad = \{c : i_\varsigma \subseteq \{w : \langle w, t_\varsigma = t_c, a_\varsigma\rangle \in \text{REDNESS}\}\}$

$\quad \cap \{c : \text{in } \varsigma, \text{ attention is targeted 'qua' red}\}$

Let us see how these interact with norms of sentence acceptance.

Suppose that I have narrowed down a restricted understanding of what the value ς might be: it is somewhere among the contexts in the set $\mathcal{K}$, but I don't know which one it is. In that case, should I accept $[\text{for}^\sigma(\Box_v)]G$? I should do so in

c just if I am confident $c \in \lfloor\lfloor[\text{for}^{\sigma}(\square_{\text{v}})]G\rfloor\rfloor$. My problem is that my ignorance of the value of ς ramifies to ignorance of which set of contexts $\lfloor\lfloor[\text{for}^{\sigma}(\square_{\text{v}})]G\rfloor\rfloor$ is. It is either C or $\varnothing$: if the former, I should accept $[\text{for}^{\sigma}(\square_{\text{v}})]G$, of course; if the latter, I should of course reject it. I could confidently accept $[\text{for}^{\sigma}(\square_{\text{v}})]G$ if for all $k \in \mathcal{K}$, $i_k \subseteq \{w : \langle w, t_k, a_k \rangle \in \text{GREENNESS}\}$ and in k, attention is targeted 'qua' green. And I could confidently accept $\neg[\text{for}^{\sigma}(\square_{\text{v}})]G$ if for *no* $k \in \mathcal{K}$ are both conditions met. But in the event of a mix, my best bet is to neither accept nor reject.

While I straddle the fence, however, I remain in a bad position. Because every reasonable context accepts whichever of $[\text{for}^{\sigma}(\square_{\text{v}})]G$ and $\neg[\text{for}^{\sigma}(\square_{\text{v}})]G$ is true, so long as I fail to do so, my context stands accused of performance error. There is rational pressure to make up my mind about what it is like for the other of a sort that there is not about humdrum matters of fact. If I just so happen to get evidence settling the issue of how many beans are in this jar—well, now I have it. But assuming there is no instrumental difficulty resulting from my ignorance, I was equally well off without that evidence. Having information is of no intrinsic rational significance; understanding what it is like for the other is of intrinsic rational significance; therefore understanding what it is like for the other is not having information. Information is instrumental to this understanding: only through getting it can I single out the context in a way required for the understanding. But it is the understanding rather than the information that matters.

How do I get off the fence? Obviously by getting information about the subject at issue: by finding out whether they see something green, for a start. This suggests that singling out a context is a matter of determining the features of the subject of the context.

Let us now develop a story about the structure of singling out Fred's context more or less determinately. When I have opinions about Fred now, I attribute properties to him as he is now. In c, my information about the world as a whole is the set $i_c \subseteq W$. Given a time t_c and an individual j^* (as it happens, Fred), we may extract a set F of ordered triples $i_c \times \{t_c\} \times \{j^*\}$: the set containing $\langle w, t_c, j^* \rangle$ for each $w \in i_c$. Sets of world-time-individual triples are 'stage-level' properties: ways for something to be for at least a moment. The set F then represents my information

61

'about Fred as he is now'; in particular, it determines which predicates I apply to Fred as he is now. For example, if $[\![\text{yawns}]\!]^c = Y \subseteq W \times T \times J$, then I (now) accept 'Fred yawns' just if $F \subseteq Y$. (Parallel remarks go for 'at-shifted' predications.) F, then, presumably determines which context I think c^{Fred} might be.

There is without a doubt a function E from world-time-individual triples into contexts with the following property: $E(w^*, t^*, j^*)$ is j^*'s context in w^* at t^*, if any.

Assembling these devices, we may say that in c, one's range of uncertainty concerning which context is c^{Fred} is then this:

68. $C{\downarrow}(c, \text{Fred}) = \{c^* : (\exists w \in i_c)(c^* = E(w, t_c, \text{Fred}))\}$

$E(w, t, \text{Fred})$ is the context Fred is in at t if w is actual: it represents the state of consciousness I somehow 'find for Fred at t' under the hypothesis that w is actual. Because I am uncertain how things are and do not determinately single out a world w as actuality, I therefore do not single out determinately a single state of consciousness as the one I find for Fred at t. Rather, there is a range of contexts representing the blurry guess at a state of consciousness as my best sense of what I find for Fred: that range is the set $C{\downarrow}(c, \text{Fred})$.

We may then establish something like a *diagonal support condition* for $[\text{for}^\sigma(\Box_\text{v})]\varphi$:

69. $/\!/[\text{for}^\sigma(\Box_\text{v})]\varphi/\!/ = \{c : (\forall c^* \in C{\downarrow}(c, [\![\sigma]\!]^c))(i_{c^*} \subseteq [\![\varphi]\!]^{c^*})\}$

That is to say, if all I know about Fred is that he is F, then that establishes a certain range of contexts, reachable through E, that from my point of view 'might' be Fred's. My 'best guess' at what it is like for Fred is then given, in effect, by supervaluating over all those contexts: if in all of them, it is part of what it is like that φ, then I should accept 'part of what it is like for Fred is that φ'; if in none of them—then I should reject; and if in just some, I should do neither. [***this is not compositional: fix]

And we may then recognize the entailments in (65) as tracking containment of diagonal support condition, even for distinctive values of φ:

70. (a) $/\!/[\text{for}^{\text{Fred}}(\Box_\text{v})]\varphi/\!/ = /\!/[\text{for}^{\text{Fred}}(\Box_\text{v})]\Box_\text{v}\varphi/\!/$

(b) $/\!/[\text{for}^{\text{Fred}}(\Box_v)]\neg\varphi/\!/ \subsetneq /\!/[\text{for}^{\text{Fred}}(\Box_v)]\neg\Box_v\varphi/\!/$

(c) $/\!/[\text{for}^{\text{Fred}}(\Box_v)]\neg\Box_v\varphi/\!/ = /\!/[\text{for}^{\text{Fred}}(\Box_v)]\neg\Box_v\Box_v\varphi/\!/$

So there is a structure to 'narrow' rational assessment of what it is like for the other to recapitulate the bare facts about consciousness as we recognize them at home.

Note moreover that if I have information about Fred to the effect that he is F, and the diagonal support condition for $[\text{for}^{\text{Fred}}(\Box_v)]\varphi$ contains those contexts with the information that Fred is F, then I thereby accept $[\text{for}^{\text{Fred}}(\Box_v)]\varphi$. In that sense, the following entailment is 'diagonal-valid':

71. Fred is $F \vdash [\text{for}^{\text{Fred}}(\Box_v)]\varphi$

There are, therefore, what some might regard as 'a priori entailments' from informational sentences to sentences about the conscious lives of others.

These entailments, of course, are undergirded by our grasp of E, the function from world-time-individual triples to contexts. The foregoing discussion suggests this grasp consists in our capacity for *empathy*. Accordingly, understanding what it is like for someone consists in empathy based on accurate information; and having some sense of what it is like consists in empathy based on some information or other. Let us see why that is.

When one is in a position to single out determinately which context is Fred's, in such a way that one can grasp the precise meaning of for^{Fred}, one then wants only for an adequate measure of something like computational capacity in order to know everything about what it is like for Fred.

I do not generally know very much about what it is like for Fred. This suggests either that I am lacking in the relevant computational capacities or that I cannot single out determinately which context is Fred's—or, perhaps in varying measures, both. I will give two arguments that in general lack of determinacy is more to blame than lack of computational capacity.

The first arguments are suggested by the 'me-shifted' case. One does in general know everything about what it is like. In light of the evident equivalence

between 'here's what it is like' and 'here's what it is like for me' and our explanation of this equivalence, one evidently does single out determinately which context is one's own. One also evidently does have exactly the computational capacities one needs to extract knowledge of what it is like from the identification of one's own context. One would think that comuptational capacities—being capacities—would have a generality that would typically apply well beyond the immediate first-person. One would also think that one's identity with oneself, coupled with one's distinctness from all others, would engender an asymmetry of identification. (After all, there is something in David Lewis's idea that identity is a form of acquaintance.)

Moreover, we do recognize the patterns of entailment and nonentailment in (65) but find those in (61) to be very alien. The former sugggests that we are in general pretty good with the relevant computations. The latter suggests we are very bad at the requisite singlings-out.

So, however things may go with special-purpose computational capacities that outstrip those generally in use in my own case, I seem not to be generally in a position to single out Fred's context determinately in such a way that I can even grasp the precise meaning of for$^{\text{Fred}}$.

Our hypothesis explains how I have perfect knowledge of my own context: empathy is a generalization of identity; I don't need to empathize with myself because I am here already; my empathizer is fed with perfect information about me in that form which I have by virtue of being me. This explains the relative challenges posed by the computing and the singling out: for matters we all have in common, such as the transparency of consciousness, everyone knows how to do the computation, because it is a baseline limit for what a conscious life could be. By contrast, objective information can, familiarly, be frustratingly hard to extract. And finally it puts the rational pressure to acquire the understanding in a more familiar cast: we broadly recognize the value of empathy as a part of love, and it seems quite likely that love is a distinctive form of reason—providing, so to speak, the 'matter' on which the 'formal' structures of practical and theoretical rationality grind away. After all, absent love, there would seem to be no point to

anything.

4.4   Assembly

Our analysandum in this section has been the sentence *Q*, as discussed at the outset of this paper: 'here's what it's like for Rebecca (in looking at a red thing): this is thus[red] and here's what it's like for Greg (in looking at a green thing): this is thus[green]'.

We argued that our *grasp* of *Q* consists in *empathizing* with Rebecca and Greg, on the basis of objective information about them; and that our *acceptance* of *Q* (in the sense of the diagonal support condition) consists in possessing objective information that, via empathy, engenders an appropriate empathetic state. Plausibly, we use this objective information to establish something like a rough template we wish ourselves to fit into. We then transform ourselves—bodily; presumably in large part neurally—so that we fit this template. Or at least that is our aim: we typically do not get anywhere near the actual condition of looking at a red thing or a green thing; some of us are 'more empathetic' than others, in the sense that we are just better at transforming ourselves so that we resemble the other bodily, and thereby come to resemble the other bodily more closely. Empathy is a *learned skill*, something we can get better at. What we have been through in the past—as with any other skill—goes into fixing certain boundaries. If we have never struggled to round up rent money, or watched the sun set over Mustique, or played music before a live audience, or 'whooshed upward' in a mass celebration, or looked at a red thing—we may then not be able to so comport ourselves such that we are in a position bodily such that we appropriately resemble those who have done these things.

Once one has done so, however, one enters a certain state of consciousness. Being in this state makes for accepting a certain class of sentences, along the lines of 'here's what it's like for Rebecca: 'this is thus[red]''. A sentence along these lines has expressive meaning. What it expresses is given by its diagonal support condition: namely, that one's information about Rebecca transforms through empathy into a range of contexts all of which (strongly) support 'this is thus[red]'.

Again, our sense of what support consists in is discussed in section 3.2: we understand support of a sentence with a certain meaning by way of empathizing with a mindset in which that sentence is accepted. So what it is like for one, when one's context fits that diagonal support condition, is this: one has a certain body of information about Rebecca's properties; one is, in a limited way, somehow 'as if' seeing red; and the latter is arrived at 'inferentially' from the former in such a way that one understands it as recognizably not too different from that state of consciousness Rebecca finds herself in.

In this story, information only intrudes at positions that are 'prior' to the attainment of a position that is in any way recognizably pertinent to consciousness as such. My information about Rebecca is solely information about her as a creature. Moving from this information to a sense recognizably about what it is like for her as such requires an *inference*; and the inferential form is available, as discussed at (71).

That is to say: the sentence Q, which the conceivability argument hoped to show *not* to be entailed by physical information, *is* in fact entailed by physical information.[8] Or at least it is entailed by information—and the physicalist thinks all information is physical information.

The dualist doesn't get to complain that this is question-begging, because the case for dualism presupposes, falsely, that Q has informational meaning. So really, while the entailment of Q by the physical is a nice consequence, it is not really the fundamental. The conceivability argument requires Q to have informational meaning. And Q does not have informational meaning. Rather, it expresses our empathy with the other.

An updating of the insights of the *Verstehen* tradition promised us a way out of

---

[8]This inference is a genuine entailment. It is not a mere 'inference to the best explanation'. In particular, it is *indefeasible*: assuming that my *information* about Rebecca was not arrived at in a way that allows nonmonotonic accretion, then if it entails that what it is like for her is a certain way, I could not learn otherwise.

But nor is it a straightforwardly informational entailment. The position denies that the inference moves from stronger to weaker information. Rather, the inference moves from an informational sentence to an expressive sentence. So the manifest difficulties faced by 'analytic materialism' are not available as objections to our view.

the conceivability argument against materialism. That argument assumed that our understanding of consciousness (as such) is not significantly different from our understanding of the nonconscious (as such)—contra to the *Verstehen* tradition. If we have not fully appreciated the insights of the *Verstehen* tradition because they have heretofore resisted formal treatment, I hope the discussion in this paper will diminish that source of resistance.