Phenomenology Without Phenomenality: A Defense of a Modest Introspectionism and Other “Intuitions”

Abstract

Shoemaker (1996) and others have presented a priori arguments against the possibility of "self-blindness," or the inability of someone, otherwise intelligent and possessed of mental concepts, to introspect any of her concurrent attitude states. Ironically enough, this seems to be a position that Gopnik (1993) and Carruthers (2006, 2009a,b,c,d,e) have proposed as not only possible, but as the actual human condition generally! According to this "Objectivist" view, supposed introspection of one’s attitudes is not "direct," as it is in the case of sensory perception, but an "inference" in the way it is in the case of our beliefs about the attitudes of other people, moreover, an inference that has the advantage in our own case of only our own sensory data and memories, and of the context we are in; i.e., we are all substantially self-blind. After sorting out a number of methodological and verbal issues, I argue that the proffered a priori arguments against Objectivism don’t succeed, and that Gopnik and Carruthers are right to regard the issue as an empirical one. However, I think they seriously underestimate the difficulty of establishing Objectivism. In particular, I argue that it is unlikely there is an inferential procedure from the data of pure sensation, behavior and context to the relevant self-attributions that would be as spectacularly reliable as people manifestly seem to be. Moreover, there is a simpler model: the mind very likely consists of a panoply of sub-routines some of whose outputs are “tagged” for their having been so processed, rather in the way that software “documents” are on standard computers. Introspection plausibly consists in a person’s simply attending to distinctive constellations of these tags, even though they may lack phenomenal feels. This draws attention to an important independent fact, that much of phenomenology (or “what it’s like” to be in a certain state) may not be constituted by facts that are phenomenal, a fact that helps explain, pace Devitt (2006), the role of structural descriptions in determining linguistic intuitions and “semantic phenomenology, as in Lormand (1996) and Pitt (2004), as well as, along lines discussed by Peacocke (forthcoming), the role of “I” and “now” in our experience, despite the lack of phenomenal objects corresponding to them.
1. Background

1.1 Introspectionism and Objectivism About Self-knowledge

In his Royce Lectures at Brown in 1993, Sydney Shoemaker (1996) sought to defend “a moderate Cartesianism—a version of the view that it is of the essence of mind that each mind has a special access to its own contents” (p27). In these days of rampant ascription of unconscious mental processes, it may be risky to insist on such access being “of the essence of mind,” but I think Shoemaker is right to think that people do in fact seem to be able to know about a wide range of concurrent (propositional) attitudes, e.g., doubt, pretense, wonder, anger, envy, fear—in a special “introspective” way in which they don’t know about such attitudes in the case of others. I shall call this weaker claim, freed from Cartesian claims about essence, “Introspectionism.” An interesting question is how such self-knowledge is possible: what is it about the way we’re built that makes such knowledge of such complex facts about ourselves so apparently easy and “special”?

Shoemaker was intent on rejecting an “inner sense” conception of “introspection,” which treats introspection as a kind of sensory perception. He feared such a conception implied the possibility of what he called “self-blindness,” or the inability of someone, otherwise rational and having mental concepts, to introspect any of her concurrent mental states:

A self-blind creature would be one which has the conception of the various mental states, and can entertain the thought that it has this or that belief, desire, intention, etc., but which is unable to become aware of the truth of such a thought except in a third-person way.

–(1996:30-1)

He claimed not only that “from an evolutionary standpoint, widespread self-blindness is not a real possibility,” but that it was not even a genuine “logical possibility” (p226; see also Burge 1996).

Ironically enough, at about the very time that Shoemaker was defending special self-knowledge, the psychologist, Alison Gopnik (1993), was denying that people actually had any. At any rate, she defended what has come to be called the “theory theory,” or the proposal that people know about their own mental states by applying a theory to themselves as they might to anyone else:

The idea of intentionality is a theoretical construct, one we invent in our early lives to explain a wide variety of evidence about ourselves and others. –(1993:p2; see also p11)

There were, however, a number of problems with this formulation of the view. Firstly, the wording here combines semantic theses about “the idea” of intentionality and its origins with epistemic issues about how we apply it. Although the semantic issues are important, it is only the epistemic thesis that concerns Shoemaker. Secondly, Gopnik’s proposal lacked any account of first-person sensory knowledge, which she

---

2. I shall use “introspection” for the distinctive, comparatively more “direct” way we seem to have knowledge of our own states, whether or not it involves the kind of internal “inspection” of special phenomenal objects as seems to occur when one introspects a sensation (cf. Dretske 1995:39).

3. The view can be traced back to Ryle’s (1949) behaviorism and Sellars (1956/97)’s famous discussion of the “myth of Jones.” Gopnik discusses her own view further in Gopnick (1994) and in Gopnik and Meltzoff (1994). Burge (1996:p105fn10) claims to find the view in Hume and in Armstrong (1968) (but see §4.1 below for a different interpretation of the latter).
allows involves “first-person psychological experiences that may themselves be the consequence of genuine psychological perceptions,” such as “simple sensations,” but which she unfortunately only characterized as a “Cartesian buzz” (1993:10-11).

In a number of recent papers, Peter Carruthers (2006, 2009a,b,c,d,e) has sorted out these issues and focused discussion on defending what I will call “Objectivism” about self-knowledge. Relying on a view he developed in his (2000), he proposes an account of introspection of phenomenal states as essentially a re-deployment of modularized perceptual outputs; e.g., a representation that means [red] comes to mean [looks red] by virtue of its automatic redeployment by a person’s internalized theory of mind. However, he thinks that no such automatic story is true of the attitudes:

Although the mind-reading system has access to perceptual states, the proposal is that it lacks any access to the outputs of belief-forming and decision-making mechanisms that feed off those states. Hence, self-attributions of propositional attitude events like judging and deciding are always the result of a swift (and unconscious) process of self-interpretation.

Bu Carruthers is no behaviorist. Although behavior and external circumstances will provide part of the data on which this self-interpretation is based, it will partly be based on data we can introspect:

[I]t isn’t just the subject’s overt behavior and physical circumstances that provide the basis for the interpretation. Data about perceptions, visual and auditory imagery (including sentences rehearsed in “inner speech”), patterns of attention, and emotional feelings can all be grist for the self-interpretative mill. –(2009a:4)

Where “mind-reading” is understood to be this objective, explanatory process, Objectivism is, then, simply the view that

Our knowledge of our own attitudes results from turning our mind-reading capacities onto ourselves. –(2009a:1)

Thus, for Carruthers, although there is introspection of phenomenal states, there is no introspection of propositional attitudes (throughout, I shall understand “Objectivism” and “introspection” as so restricted).

For Gopnik, the illusion of introspection is the result of the phenomenon of “expertise,” whereby experts on some topic come to regard their often highly theoretic inferences as immediately perceptual (1993:10-2). Carruthers (2009c) also emphasizes that the inferences are swift and unconscious, but also adds

4. Terminology: some but not all writers on this issue distinguish introspection as “meta-cognition,” objective attribution as “mind-reading.” Hence Carruthers (2009a:3) himself calls his own view a “mind-reading is prior” view. This, however, doesn’t lend itself to succinct nominalization, and also raises issues about “priority” that are not my concern. So I prefer my “Objectivism.” In addition to Gopnik and Meltzoff (see fn 3), versions of the view have also been suggested by Gazzaniga (2000) and Wegner (2002), but, since Carruthers has offered the most recent, vigorous and systematic development of it, it is his treatment that I will exclusively discuss.

5. He (2009a:14) exempts “thinking that...”, which he takes to have so few commitments in terms of its relations to other states that it may well be treated on a par with phenomenal ones. I assume what he has in mind is mere “inner speech,” without any particular attitudinal commitment, a general topic I’ll address in due course.
the hypothesis that “a belief in the self-transparency of the mind is innate” (2009c:ms.p3): we are born disposed to believe that occurrent mental states are introspectible.

The main evidence that Gopnik (1993:§3) presented on behalf of her view was the fact that children who failed the “false belief” tasks in the case of other people also failed it in the case of their own prior beliefs. Much of this evidence has become less than compelling as a result of more recent, subtler tests of “false belief,” that seem to show that even 15-month-old infants are capable of ascribing it (e.g., Onishi and Baillgereron 2005), and it is difficult to test them for dissociation in ascription between self and others. Aware of these further developments, Carruthers bases his own defense of Objectivism instead on arguments from computational economy and evolutionary function (2009a:§5), criticism of data suggesting a dissociation between self-attribution and attribution to others (§§9-10), and on data about self-confabulation, in which people seem to be self-attributing attitudes that they can be shown not to have (§6).

In the present paper I want to raise a number of objections to this Objectivism, and defend at least the serious possibility of special introspective knowledge of many of our own present-tense occurrent attitudes. I do think that, pace Shoemaker and Burge, the issue cannot be settled a priori –I shall something briefly about the issue that concerns them, which turns out not to be, strictly speaking, the issue of Objectivism raised by Carruthers– and that Gopnik and Carruthers are quite right to force us to look at some of the sometimes surprising empirical evidence. But then, picking up on an issue that I think Shoemaker (1993) rightly raised against Gopnik, I want to press what seems to me a quite simple, but serious problem for Objectivism, viz., how a third-person inferential process could possibly be as spectacularly reliable as a significant set of self-attributions in a wide range of cases seem obviously to be (§3).

Of course, this seeming reliability may in fact turn not to be genuine; again, I don’t think the issue can be settled a priori. However, I will argue, not a shred of evidence has been adduced to think that we aren’t in fact as reliable about many states as we fully seem to be. But, in any case, it will be worth noticing that, the greater our reliability, the greater will be the challenge to Objectivism to explain it.

This latter challenge will, moreover, be increasingly serious if another, simpler theory could explain the greater reliability without such serious difficulties, and so I shall proposes the (I fear, idiotically) simple suggestion that the results of certain inner processes are “tagged” for their having been so processed, rather in the way that software “documents” are on standard computers (§4). I’ll argue that this affords a even more “moderate Cartesianism” than that proposed by Shoemaker, but it also a more modest, less problematic empirical proposal not only than the Objectivist one, but also than the “self-monitoring” proposals of Nichols and Stich (2003) and Goldman (2006) (although it is compatible with them). It also has the (to my mind) interesting consequence that, as I like to put it, there is more to phenomenology, or “what it’s like” for us, than phenomenal feels, a fact that seems to me to explain not only introspection of attitudes, but some of the “voice of competence” associated with linguistic intuitions and doubted by Devitt (2006); the “semantic phenomenology” recently discussed by Lormand (1996) and Pitt (2004); and the phenomenology of “the self” and “the present,” recently discussed by Peacocke (forthcoming).

6. By “a serious scientific possibility” I mean merely one deserving further research, not having been decisively ruled out by existing science, in the way that, say, ghosts have been. Needless to say, merely establishing such a possibility is not establishing its truth, which, in the present case for the forseeable future, seems to me far too ambitious a goal.
An important limitation on the discussion: I shall not be addressing the difficult issue of whether or how someone can introspect the content of an attitude state (see Wright, Smith and McDonald 1998 for many discussions). I shall be concerned almost entirely with the issue of whether and how someone can know which attitude it is they bear to a content, qua that attitude: e.g., whether it’s desire, dread, doubt, or pretense. Of course, the two issues can’t be entirely separated, and I’ll say (or assume) a few things on behalf of a “re-deployment” account in passing (e.g., §4.1), but the problems in sorting out the “which attitude” issue will take up space enough.

My defense of introspection is not unqualified. Although I think that we do have introspective knowledge of a great many of our attitudes, qua mental states, I by no means want to defend the claim that introspection is infallible. Unlike some, I think it’s at least a conceptual possibility that we have no attitudes at all, and that it’s a serious scientific possibility that there are no such things as the “consciousness” or “qualia” that people standardly take themselves to introspect (see Rey 1997). Nor am I defending any special epistemic role for introspection either in our knowledge of the world or of other people. I very much doubt introspective knowledge serves as the foundation of all knowledge, or even as the basis of our knowledge of other minds along the lines of the traditional “argument from analogy,” as a number of “simulation theorists” (Gordon 1986, Goldman 2006) have recently urged. All I’m defending here is the possibility of some substantial special introspective knowledge of some propositional attitudes states, specifically, most of the ones we ordinarily take ourselves to be able to introspect.

1.2 Methodological Remarks

As Carruthers (2009a:§3) acknowledges, Objectivism does seem intuitively pretty wild. It’s a cousin view to the preposterous behaviorist claim that someone knows her own mental states only though observation of her own behavior. The only difference is that the Objectivist includes sensations as data, which, we’ll see (§3), is a meager concession. Indeed, it suggests that our introspectible lives consist merely of sensations, and this seems to many of us a stunningly impoverished account of our inner lives.7

Carruthers, however, is right to insist that intuitions are far from decisive generally, and can be surprisingly unreliable regarding the mind (2009a:6-7). If the inference to self-attribution is swift and unconscious, it could easily seem special and immediate without being so. Indeed, introspection seems in general to be unreliable about the details of mental processes, e.g., language processing, perception, reasoning —and perhaps introspection itself. Insofar as Objectivism can be sustained as a serious empirical hypothesis (as I’ll argue in §2 that it can), at least some sort of evidence beyond mere introspection will be

---

7. This general view of our inner lives dominated much of the traditional “empiricist” conception such as one finds in Locke and Hume, and persists even in recent writings, see, e.g., Barsalou (1999). Wittgenstein (1967) nicely captured some of the phenomenological unease one might feel about it:

Is it hair-splitting to say: --joy, enjoyment, delight, are not sensations? --Let us at least ask ourselves: How much analogy is there between delight and what we call "sensation"? ... "I feel great joy" --Where?-- that sounds like nonsense. And yet one does say "I feel a joyful agitation in my breast".-- But why is joy not localized? Is it because it is distributed over the whole body? Even where the feeling that arouses joy is localized, joy is not: if for example we rejoice in the joy of a flower. ... 'Horrible fear': is it the sensations that are horrible? --(§§484-504)

Page 5 of 25
needed to decide the issue. Introspectionism can’t be established by introspection alone (which is not to say that introspection mightn’t be a source of some evidence).\footnote{8}  

1.3 (Attitudinal) Feelings vs. (Non-attitudinal) Sensations

Objectivism as an empirical hypothesis does, however, involve a specific commitment that may not be obvious at first glance. As we noted, Carruthers allows that people do introspect their sensations. But what precisely are these? A great many phenomenal states would seem on their face to involve the attitudes to which Carruthers claims people don’t have introspective access. Not to belabor the obvious, but pain involves aversion; hunger a desire to eat; thirst a desire to drink, fear a belief that danger is imminent, and so forth, to more sophisticated emotional states such as resentment, pride, exasperation. It is therefore crucial to understanding Carruthers’ proposal that we at least provisionally distinguish what I will call (non-attitudinal) sensations, like looks red, feels heavy, and mere “inner speech,” from (attitudinal) feelings, like pains, itches, thirst, and “inner assertion.”

I emphasize that this distinction is provisional, drawn only for the purpose of clarifying Carruthers’ claim. I simply am assuming for the sake of argument that he is entitled to distinguish modularized “perceptual” judgments that he thinks we can introspect from the more general attitudes that he thinks we can’t (see 2009a:45).\footnote{9} However, the issue could do with a lot more discussion. He himself has not been explicit about it in any of his recent articles –to the contrary, as we saw, he casually included “emotional feelings” in the data base for self-attributive inferences (2009a:pp4,16). I press these difficult cases as a way of drawing attention therefore to the extraordinarily strong claim to which Carruthers’ Objectivism is committed:

Attitude self-ascription is based on only on data about sensations, behavior and circumstances (what I’ll call “SBC data”).

It is this “self-blind” thesis that Introspectionists like myself find implausible.

1.4 Processes vs. Data

Carruthers’ is actually making two claims that need to be distinguished: (i) the claim that I’ve just set out, that the only data available for introspection is SBC data, but also (ii) that introspection is interpretative, i.e., involves applying the resources of a theory of mind as an explanation of that data. It is important to set this second claim aside, if only in view of the difficulties attending what counts as “interpretation,” “inference” and “explanation” in general. Quite apart from issues about introspection, it

---

8. “It is introspectible that p” certainly does not entail “It is introspectible that it is introspectible that p.” Carruthers (2009c:ms.p10) does, however, seem to me to overstate the point when he claims “the empirical data are pretty decisive in showing that the subjects themselves are unable to distinguish between confabulation and introspection. Hence subjects themselves can’t tell whether or not mental states are transparent” (2000C:ms. p10; see also 2009a:7). In support, he cites Gazzinaga’s (1995) observations of split-brain patients. However, this conclusion overlooks the fact that (i) data that can’t certify something by themselves might well do so in conjunction with other data, in this instance, with data about the sort of cases (unlike those of split-brains) in which people’s introspection are in fact reliable; and (ii) it’s not clear that any efforts were or ever have been made to determine whether people can distinguish such processes if they were carefully to try (in fairness, of course, as Carruthers (2009a:49) notes, it’s not obvious how to get subjects to do this). Although introspectivism can’t be established by introspection alone, it could turn out that people could be trained to introspect whether they’re introspecting (cf. the end of §4.1 below).

9. On the view I will be defending in the end (§4), the distinction may not be so clear –precisely as I think it introspectively doesn't appear to be.
is far from clear that the only way we deploy either mental or any other concepts, to ourselves or others, involves serious interpretation, inference or explanation. Not that such processes aren’t obviously crucial some of the time. But life is too short and matters too pressing for us to rely on them alone. I assume what evidence in studies of vision and language abundantly supports, that many of our applications of concepts involve merely some kind of simple pattern recognition: we are presented with a certain pattern of perceptual cues and immediately recognize (or mistake), say, a dog, a cat, a dangerous street or the Empire State Building. No serious inference, not even an unconscious one, seems to be involved—indeed, something may continue to look like a dog or a cat or dangerous long after we’ve inferred that the snap judgment was mistaken (see Pylyshyn 2003:??). The same could well be true of even the kind of Objectivist self-ascription Carruthers imagines: although sometimes it may be based on laborious inferences, sometimes it could probably be such a quick recognitional episode. It’s just that for Carruthers, unlike an introspectionist like myself, it is based only on SBC data. So the issue that divides us is not interpretation (or “interpretavism,” as he terms his own view), but simply the character of the data on which self-ascription can sometimes rely, whether recognitional or interpretive.

1.5 The Predictive Situation

Carruthers (2009a:§1) considers four models of self-attribution, his own Objectivism (which he calls “Mind-reading is prior”) and three Introspectionist rivals: a “two mechanisms” view defended by Nichols and Stich (2003); a “One Mechanism, Two Modes of Access” model that says was implicit in his own (1996); and a “MetaCognition is Prior” view defended by Goldman (2006). The differences between the three latter will not be important here—the modest proposal I will defend in §4 will be neutral between them—however it is important to note a general error that Carruthers seems to make about the predictive situation with regard to them and, it seems to me, to any reasonable Introspectionism. He claims that Objectivism is the only one of the models he considers that predicts that it should be possible to induce subjects to confabulate attributions of mental states to themselves by manipulating perceptual and behavioral cues in such a way as to provide misleading input to the self-interpretation process ... Likewise, the account predicts that there should be no such thing as awareness of one’s own propositional attitudes independently of any perceptually accessible cues that could provide a basis for self-interpretation... [T]he “mind-reading is prior” model is the only one of the four to make such predictions. –(2009a:3)

But it’s hard to see why this last is true. So far as I can see, all the rival views can allow that we can know about ourselves non-introspectively, and that in daily life our self-attributions may often be based not only on special introspection, but on all manner of other perceptual and theoretical information as well, as both Nichols and Stich (2003:163) and Goldman (2006:232,263-4) allow. Indeed, any Introspectionist view is committed to this happening, since any introspector presumably employs mental concepts that she applies to people generally, and so, a fortiori, to herself, perhaps without noticing that they might conflict with her introspections (many people believe they’re selfish because they think everyone is).

The only objection that Carruthers raises against such “dual mode” theories is that having such a second mode would seem to be superfluous:

Because [dual views] maintain that introspection for propositional attitudes exists, subjects should generally have no need of evidence of any kind when making self-attributions. –(2009a:14).
But it’s hard to see the force of this objection. In the first place, I see no reason to suppose that people restrict themselves merely to what they need, either in drink or in evidence. But, secondly, why wouldn’t it be a perfectly good idea to have both multiple and overlapping systems, especially given the merely superficial differences in merely one’s perspective on oneself and on someone else? Redundancy of information from different perspectives can fortify confidence (cf. Adler 2002), and it could serve us well sometimes to monitor ourselves “directly,” and sometimes also to see ourselves as others see us—even if one or the other way may sometimes trump the other. In any case, in the rough and tumble of ordinary life people obviously reach for any evidence that’s handy. All that is essential to Introspectionism is that it is possible for people sometimes to attribute attitudes to themselves in a way that involves more than SBC data, even if, contrary to one’s impression, that pure ability may be deployed haphazardly in ordinary life. Occasional failures in performance are hardly sufficient to show a lack of underlying competence.

With these clarifications of the predictive situation in mind, I turn now to a number of relatively a priori arguments that might appear to settle the issue, before considering to what seem to me to be the empirically significant facts.

2. The Irrelevance of (Relatively) A Priori Arguments

2.1 “Analytic” Self-attribution

There is, of course, a trivial way that everyone could agree that people have a kind of automatic self-knowledge of their own attitudes, viz., by so conceiving the relevant attitude concepts! This strategy is familiar in the history of analytic philosophy. A recent, representative example is perhaps Crispin Wright (1989), who writes that “the authority standardly granted to a subject's own beliefs, or expressed avowals, about his intentional states is a constitutive principle: something that is not a by-product of the nature of those states”–(p632)

Shoemaker himself argument proceeds by considering the peculiar alienation he thinks the self-blind would suffer with respect to their own attitudes, and writes:

What rationalizes the investigation [into one’s beliefs] are one’s higher-order beliefs about what one believes and has reason to believe. Creatures without introspective access to their beliefs and desires would lack this resource for rational revision of these beliefs and desires, and would fall short of normal human rationality. –Shoemaker (1996:240)
However. Shoemaker’s worries are actually beside the point of the Objectivism Gopnik and Carruthers aim to defend. As we’ve noted, Objectivism isn’t committed to self-ascriptions being *consciously* inferred in a third-personal way. Nor is it committed to them being *expressed* in third-personal terms. It can allow that self-ascriptions are standardly expressed, both in language and in the mind, with a special first-personal indexical that causes those self-ascriptions to play a distinctive, committed role in a person’s psychology.\(^{13}\) Just as the thought that *I* (and not merely GR) am about to be eaten by a bear causes me to flee, the thought that *I* am contradicting myself often (though hardly always) causes me to fret. That is, the Objectivist can distinguish issues about the importance of a first-person perspective on oneself from issues about *how that perspective is achieved*. It may well be that a serious critical reasoner has to be able to reflect upon and take responsibility for shaping her views in accord with an evaluation she makes of them, rendering “the first- and second-order perspectives...the same point of view” (Burge, p110). But none of this entails anything about their *etiology*; that she needs to have a *non-inferential* route to the attitudes themselves, that she couldn’t be relying entirely on swift and unconscious inferences from SBC data. I conclude that neither Shoemaker nor Burge succeed in refuting Objectivism as a possible psychology.

However, Shoemaker does elsewhere point to what seems to me the right worry about it as an account of our actual psychology. He was one of the commentators on Gopnik’s original article, and he rightly complained that Gopnik tells us basically nothing about what the immediate objects of "psychological experience" are that provide the basis for our theoretically grounded inference to intentional states; indeed, that she leaves it a mystery how these unspecified contents enable the child (a little Einstein already at 3-1/2) to construct the theory-theory and make self-ascriptions based on it. –Shoemaker (1993:79)

Now, Carruthers is more specific about experience than Gopnik, but he still seems to me to leave the inference a mystery, especially if one considers the full range of propositional attitudes that Objectivism purports to cover, to which we now turn.

### 3. The Poverty of Sensation

#### 3.1 The Range of Attitudes

A remarkable feature of much of the philosophical attention to the present topic is the oddly restricted diet of examples that are considered. Most of the discussions focus on introspection of "committal" attitudes, such as belief, judgment and decision, with occasional mentions of desire and intention,\(^{14}\) and this seems to me to have distracted attention from the extraordinary range of self-attributions about which Objectivism makes a claim. Doxastic cases in general are particularly distracting in this way, since, as Carruthers (2009a:5) notes many have observed, e.g., "I believe/judge/ expect/know p" are often not serious self-attributions, but simply a way of (perhaps with varying degrees of confidence) asserting "p" itself. But, of course, rather than this fact serving as an argument against the need of introspection, it should really serve as a reason to avoid using belief self-attributions as a good test case, as I shall. In any event, it is worth sampling briefly some of the

---

\(^{13}\) See Casteneda (19???) and Perry (19???). In my (1997:ch 11) I suggest a way to incorporate a first-person indexical into a person’s computational psychology.

\(^{14}\) Carruthers (2009a) pretty much restricts himself to belief, desire, judgment and decision (pp5,10-15). As we saw (§1), his statement of his proposal (2009a:4) focuses only on belief and decision, and he seems simply to presume (2009a:pp4,16) that "emotional feelings" are not relevantly attitudinal. It seems to me a conspicuous failing also of Dretske’s (995:ch 2) discussion of introspection that he discusses only thoughts about belief and perception.
A baffling claim to the contrary seems to be made by Hurlburt (2009), himself, who writes in his commentary on Carruthers (2009a):

“As a result of 30 years of carefully questioning subjects about their momentary experiences, my sense is that trained DES subjects...almost never directly apprehend an attitude. –(2009:30)

However, it’s very hard to reconcile this claim with all of his data that I’ve just cited.

very wide variety of kinds of states that people ordinarily appear to be able to introspect with what appears to be excellent reliability (the categories are not meant to be deep, exclusive or exhaustive):

**appetitive attitudes:** desire, hunger, thirst, itching, pain, lust, craving;

**emotional attitudes:** pride, joy, love, hatred; fascination, fear, disgust, dread, worry, envy, resentment, amazement, shame, embarrassment, exasperation, eager anticipation;

**committal attitudes:** judging, making up one’s mind; deciding to endorse, deciding to act; predicting; expecting; doubting;

**non-committal attitudes:** surprise; curiosity, wondering, imagining, supposing; joking; lying; pretending; being ironical; puzzling about; trying to solve a problem.

These all seem to be the sorts of states that people spontaneously introspect. Although there’s much to worry about in the methodology of Hurlburt’s (1993) “DES” protocols (“Descriptive Experience Sampling,” whereby a subject reports on her immediate conscious states on a hearing a randomly generated beep), the kinds of states reported include “feelings of love” and “recognition” that it was un reciprocated (p52); “wondering,” “puzzling,” (p76-8); “wishing,” “jealousy,” “frustration,” “anger” (pp98-9); “irritated,” “tolerant,” “self-critical,” “curious,” “envious,” “amazed,” “anxious,” “inadequate,” “bored,” “uneasy,” “indecisive,” “bitter,” “grieving” (pp128-9). Hurlburt and Schwitzgebel (2007:306-8) include "sadness/dread," "trying to remember" and "searching for" a word; "worry," "feeling exasperated," "yearning to go diving," "concern," "resentment." A particularly revealing incident is someone’s “puzzling about [a] concept being described.” Regarding it, Hurlburt writes:

> She could not say precisely how she knew she was puzzling except that it seemed as if she were approaching the concept first from one direction and then from another. This approaching/re-approaching was not verbal, imaginal or bodily; it seemed to be mental or cognitive, but could not be described more adequately. –Hurlburt (1993:p76).

At least on the face of it, people seem regularly to introspect a wide variety of propositional attitudes.15

15. A baffling claim to the contrary seems to be made by Hurlburt (2009), himself, who writes in his commentary on Carruthers (2009a):

As a result of 30 years of carefully questioning subjects about their momentary experiences, my sense is that trained DES subjects...almost never directly apprehend an attitude. –(2009:30)
3.2 Terrific Meditative Cases

How reliable are people about self-attributing these states? I know of no studies; but I also know of no serious doubts (which is perhaps partly why there have been no studies). Very likely, the reliability varies to some extent over the cases. On the one hand, we don’t find often people claiming to be, e.g., afraid/suspicious/doubtful/proud, and their not acting in ways consonant with those states (although this is imaginable); on the other, we are familiar with the repressed, the self-deceived, people in the grip of a bad psychological theory, or with an imperfect grasp of the relevant concepts, who sometimes seem to get their occurrent attitudes wrong. But, as one says, these latter are exceptions that prove the rule: the very fact that they seem to be unusual attests to the unusual reliability in most cases. I will therefore take for granted what I suspect most people take for granted, that there exists at least a substantial set of what I will call “Terrific” cases, in which people are virtually perfectly reliable. If future, unanticipated evidence shows that they’re not—that they regularly make mistakes about whether they’re wondering, doubting, fearing, etc., then, fine, so much the worse for the truth of my conclusions. All that really concerns me is how to explain whether whatever reliability we do display.

One kind of Terrific case that would be especially critical for distinguishing an Introspectionist from an Objectivist view would be ones in which behavioral and circumstantial data are minimal, say, someone sitting quietly in a non-descript room with her eyes closed for twenty minutes, what we might call "Meditative Cases" (not to be confused with states of "spiritual meditation," which are merely a subset). An Objectivist is committed to our self-knowledge (again, of our attitude. not merely its content) in such cases being based on sensations alone! Indeed, Objectivism should predict that people are less reliable in such meditative cases than they are in wide open-eyed, contextualized ones, since they are basing the same sorts of inference on significantly less data, indeed, without benefit of the sort of both behavioral and “contextual” data to which the Objectivist routinely appeals.

Carruthers doesn’t discuss any Meditative cases. Such cases do raise the question of why someone self-attributing, say, a pain or thirst in such a case shouldn't be regarded as introspecting an aversive or appetitive attitude, since it certainly would appear that the self-attribution would be proceeding in a fashion fundamentally different from their attribution of those states to someone else. Perhaps, as we already speculated, pain and thirst, what with their mix of sensations and automatically linked attitudes, might qualify as some sort of special borderline cases. But many attitudes don't seem "perceptual" even in this extended way. Desire, wonder, doubt, pretense, curiosity, for example, don't seem to be linked to any specific sensations. How do we proceed to self-attribute them apparently so reliably in meditative cases, as I presume we do?

3.3 Objectivist Explanations of Reliability

As I mentioned earlier, both Gopnik and Carruthers attempt to explain the reliability of self-attribution by appeal largely to the fact that we, well, spend an awful lot of time with ourselves, and so have abundant relevant data. As we saw, Carruthers allows that “people are, probably, excellent interpreters of themselves” (2009a:7; 2009c:13). Oddly, he adds in the same breath that “they are remarkably good interpreters of others,” as though there weren’t much of a difference. But the point is they seem spectacularly better in the
kinds of cases we are considering, particularly the meditative ones, a difference that must be due for Carruthers entirely to sensations, one’s story of oneself, and whatever inferences these alone can sustain.

Although Carruthers doesn't consider Meditative cases, we can perhaps guess at his strategy from a few ordinary cases he does discuss. For example, he writes:

As a result of an episode of...conscious activity, I might formulate and rehearse the assertoric utterance, "Polar bears are endangered." Under interpretation, this event will likely be heard as an assertion that polar bears are endangered. –(2009a:17)

But how is this self interpretation to proceed? A striking difference with respect to speech between the first- -and the other-person case is our assumption that others are in fact reliable introspectors of themselves: I can know whether and what you’re imagining or wondering pretty much only by having you tell me, and by assuming that you’re sincere and a reliable introspector. But how am I supposed to apply this assumption in my own case? Why do I not have remotely the kinds of problems disambiguating my own speech that I often have in understanding others?17 “Inner speech,” unlike the usual outer sort, does not standardly consist of any sort of speech act: it is not ordinarily produced to convince anyone of anything, or to achieve any particular effect; nor does it when it accompanies reading and writing (as it invariably does in my own case, where it seems largely “epi-phenomenal”). So it is not at all obvious quite how to adopt such a fully third-person attitude towards one’s inner life, much less that people standardly do –and still much less, that they do or could do so while being as reliable as they seem to be.

Carruthers offers another strategy in his discussion of one of Hurlburt’s “unsymbolized thinking” cases, where a subject without appropriate SBC data reported wondering whether her friend picking her would be driving his car or his truck. Carruthers allows that this would seem to be a problem for his view, but writes:

What cannot be ruled out, however, is that the thought in question is self-attributed because it makes the best sense of sensory activity that had been taking place just prior to the beep –for example, two memory images deriving from previous experience, in one of which the friend arrives in his car and in the other of which he arrives in his pickup truck. –(2009a:15)

But, again, it’s hard to see how this is really supposed to work. How could the mere images be enough for such an inference? Why not infer the hope, or the doubt, the fear, or just the memory or fantasies that he might come one way or the other? Surely these and other attitudes would make equally “good sense” of merely the images (n.b., in such an “unsymbolized” case, one couldn’t even rely on the syntax of wondering “whether...”). And consider all the other images, and sensations that were likely coursing through her mind: how is expecting the friend to come the best explanation in the light of all of them? What “laws” or even reasonable generalizations might she appeal to? That expectations (but not, e.g., hopes, doubts, fears, memories, fantasies) of an event regularly produce in her just such sensations and images of the event? Or that certain patterns of sensations or inner speech regularly manifest specific attitudes (cf. fn ?? above)? Maybe. But I wouldn’t count on it –and I doubt if people do (Lurz 2009:34 makes a similar point).

Why, indeed, infer anything at all? Trying to explain (within ordinary resources) why certain images have entered one’s meditative mind would seem the least plausible of Freudian projects. Unlike the patterns

17. Note that this problem should be especially acute for congenitally deaf “signers,” since, unlike the case of aural speech, which is “heard” indifferently to 1st or 3rd person perspective, signing appears quite different from the two perspectives, many of the relevant hand gestures visible to a third-person lying outside the visual field of the first.
in the behavior of other people, the course of meditative sensations—the stream of images, aches, tingles, tunes, snatches of speech and memory—seems virtually random. (Not so, perhaps, for much in the flow of thought; but it is precisely this the Objectivist claims we don’t experience!) In any case, short of the odd associative link or deliberate mental act, surely none of us have the faintest idea what would seriously explain much of any of it.

By way of contrast, consider the case of open-eyed vision. Vision theorists are similarly challenged to try to explain how a three-dimensional representation of a scene can be based reliably upon merely retinal stimulations. This is a non-trivial problem, which they have addressed with considerable ingenuity and detail, showing how certain computations plausibly realized in the brain can give rise in the contexts of human vision to perceptions of objects in space. For Objectivism to be plausible, an analogous theory would have to be provided to get us from sensations to self-attributions, and it is not at all easy to imagine how it would go. We would need theorems analogous to, e.g., those proved by Shimon Ullman showing how structure can be inferred from motion, and ultimately 3D forms from retinal intensity gradients (see Palmer 1999). But is it remotely plausible that there are such theorems, or even rough laws? A defender of Objectivism has to provide some reason to think so.18

In any case, while appeals to recent memories and bits of general self-knowledge are perfectly possible strategies that might work to account for some Terrific Cases, what needs to be shown is that they might reliably work for all of them. It is this general reliability claim that seems to me, if I may say so, staggeringly implausible; or, in any case, very difficult to show, especially across the extraordinarily wide range of cases—all attitudes!—to which Objectivism is supposed to apply. Perhaps Carruthers or other Objectivists can find ways to render it less implausible. But before effort is expended in that direction, it’s worth considering a more modest proposal.

4. A Modest “Tags” Model

4.1 Tags of the Output of Sub-Routines

Assume a computational-representational theory of thought, along the lines of Fodor (1975), Carruthers (2000) and Rey (1997). On such a theory, the mind presumably has many "sub-routines," some of which may count as "sub-systems" and even "modules," and maybe some are just little programs that are run now and then, rather like the varied programs on a computer. If so, many such sub-routines might simply "tag" their outputs, distinguishing the output of one such routine from that of another, just as a computer does, often tagging with a suffix not only the program that issued the output, but the date, time, user, and date of previous revisions. I presume in the brain they would be physical states that serve as indicators of various states and processes that enter into further computational processes, but also, of course, may also cause any number of non-computational and perhaps non-mental states and processes. After all, considered apart from these routines, the representations (data-structures, sentences in an LOT) consist of a vast swarm among which an intelligent mind surely had better keep careful track, and to which it might need to respond in any number of ways. Quite plausibly, some representations come with tags about the sense modality in which they originated, as well as, perhaps, the time, intensity and urgency of the specific state. Many memories, dream reports, "forbidden" thoughts and desires had better be so marked, and it's plausible there are deliberately "called up" sub-routines for such mental "actions" as wondering, pretending, imagining, supposing, as well as

---

18. Carruthers (2009a) proposes another source of reliability, self-fulfilling attributions, as might occur when agents “will feel obliged to act in ways that are consistent with the mental states that they have attributed to themselves” (2009a:7). This may be plausible for some doxastic and “committal” attitudes, but would pretty clearly not apply in non-committal, appetitive and emotional ones, such as wonder or worry (where indeed I might like to get over it!).
tags for non-deliberate sub-routines, as in worry, fear and desire. Many of these tagged outputs could be "globally broadcast," and, when one attends to them with mental concepts, readily occasion the appropriate concept (or near enough; I expect the concepts we ordinarily have available may be a crude lot), rendering them directly introspectible in a distinctive first-person way (cf. Perepelyplotchik 2009:37). But we could (come to) do this quite automatically, say, by pattern matching, without anything like the third-person inferences Carruthers proposes.

Of course, there needn't be a dedicated sub-routine for every distinct attitude. I presume no one yet has anything near an adequate idea of the variety of actual sub-routines, sub-systems or "modules" available to a normal mind. The ones I've mentioned seem plausible, but there's no general reason to suppose that our ordinary categories always carve our minds at its joints. As mentioned earlier, internal representations don't seem always to be well-tagged for memories vs. rational reconstructions. But, even in cases in which people seem perfectly reliable, it would be surprising if there were special routines for e.g., amazement, embarrassment and exasperation. These are likely the result of interactions of any number of sub-systems or sub-routines, each presenting their own tags, one's awareness of which state one is in depending on the particular constellation of tags and sensations that are available. Such constellations might well come to be recognized in learning a particular mental word along the lines of "recognitional concepts" discussed by Brian Loar (1996), but applied, of course, to data standardly available only to oneself.19

A question of course arises as to how these tags get their semantics. In his (1993), Goldman raised a problem for "functionalist" definitions of mental concepts: how could such complex relational states be so quickly and reliably introspected? Both there and in his (2006:259-72), he seems to want to tie the content of mental concepts to introspection, along lines we discussed in §2.1. But this is a problem more for semantics than a theory of introspection, and suffice it to say that theories of meaning needn't be so constrained. A tag such as "MOTOR" could have its meaning by merely by virtue of its role in indicating motor commands to other parts of the mind, whether or not that role itself is represented, much less surveyed either by the "learner" of the concept, or the user of it each time it is deployed. Moreover, the relevant contents of the tags (if not the contents of the representations so tagged) might be "non-conceptual," not entering into ordinary "central" conceptual reasoning, but serving only as input to it (cf. Nichols and Stich 2003:163). Ordinary lexical concepts of mental states might be triggered by configurations of tags and sensations without being "reducible" to them, just as the concept of dollar may be triggered by standard appearances to which they obviously can't be reduced.

It is important to bear in mind here the distinction discussed earlier (§1.4), between issues of interpretation vs. pattern-matching processes and the data to which either sort of process has access. It is not only possible, but strikes me off hand as quite likely, that some episodes of special introspection might well consist in trying desperately to interpret or otherwise explain some perhaps unusual constellation of tags (as well, of course, as SBC data) that happens to have come one’s way –“Why one earth do I feel quite so odd? What is this state I’m in? Depression? Anger? Anxiety? Or maybe something for which I haven’t yet come to deploy the right word or concept?” But of course the quick, terrifically reliable cases will presumably be ones involving quick pattern matching, in the case of tags as in the case of the quick visual recognition of dogs, cats, or, again, dollar bills.

19. As Goldman (2006:245) emphasizes, there may also need to be "intramental translations" from the codes of one sub-routine to the codes of another for normal conceptualized introspection. Tags, moreover, may presumably sometimes be lost, deleted or replaced; and, of course, much might be lost in such translation, especially when possibly crude or misguided concepts from one's theory of mind are deployed to interpret the constellations of tags and sensations.
A likely and illuminating instance of the kind of tagged states I have in mind is perhaps best provided by "efferent copies," whereby the mind is alerted of an endogenous motor command, and which are essential to our keeping track of when it's one’s own body, and not the world, that is moving. There's not merely a representation of the movement of a body part —"My right eye is moving"— but a representation of that motion that's marked for having been issued by the motor system itself, e.g., "MOTOR: my eyes are moving." Moreover, unlike typical afferent perceptual signals, these efferent copies don't seem to give rise to any phenomenal experiences. One "just knows," e.g., that one is moving one's eyes on the basis of these commands alone (cf. Goldman 206:264). Similarly, I want to suggest, are the representations produced by other processes so tagged for having been so produced, and one similarly "just knows" that one is, e.g., doubting, fearing, supposing, wondering, pretending, often without any apparent basis in sensation (and without, of course, any awareness of the tags as tags).

Indeed, what seems to me an important consequence of this view is that there's, so to say, more to "phenomenology" than mere "phenomenal experience," more to "what it's like" than merely the notorious "qualia" associated by many writers with standard sensations and sensory perceptions. It's easy to fasten our attention on these supposed qualia, and fail to notice the important representations that provide the content and structure to our experience, even though they themselves may have no qualitative correlate. Good examples are the geometrical representations that vision theory reveals inform how we visually experience objects (see Pylyshyn 2003:134), or the grammatical representations that (psycho)linguistics suggests constitute "how we hear" speech (see Fodor 1983:93); or the phenomenology of "the self" and "the present," which, as Peacocke (forthcoming) has rightly emphasized, involves unconscious representations of the subject and time of experience, without there being a phenomenal object in either case on which one can fasten one’s attention, or, to return to kinds of cases I’ve been discussing here, the efferent commands and other tags of various computational states that provide the basis for our introspective knowledge of our own attitudes. To reply to Wittgenstein's query (fn 8 above): in "horrible fear" it's the immediately apprehended whole state, tagged representations and all, not merely some constituent sensations, that is horrible.

It’s worth lingering for a moment on the case of linguistic intuitions, since Michael Devitt (2006, 20??) has recently proposed a view about them that is closely analogous to Carruthers’ view about self-ascription of attitudes. Devitt is intent on arguing against what he calls “the voice of competence” view that he finds presupposed in particularly Chomskyan linguistic methodology, which does, indeed, rely on speaker’s “linguistic intuitions” as the main source of data for hypotheses about the grammars represented in the speakers’ brains. Analogously to Carruthers’ claims about self-knowledge, Devitt thinks our “intuitions”

---

20. In my (2008) I suggested efferent copies as an analogy for attitudinal self-knowledge; the present proposal is an attempt to spell the analogy out. Peacocke (2008a,b) goes so far as to offer it as an account of mental acts, such as deciding, as well. However, I agree with many of Carruthers’ (2009b) reservations about this specific extension of efference copies, which, on the present account, are simply an instance of one kind of tag.

21. As my qualifications here indicate, I’m actually sceptical of the ways may philosophers think of qualia as some sort of entities distinct from our internal representations of them (see my 1997:ch 11). However, this issue is irrelevant to the issues presently at hand.

22. I might add that I am thus sympathetic to those, like Goldman (1993:24), Lormand (1996) and Pitt (2004), who claim that there is a distinctive phenomenology to cognition: cognitive states are a crucial bit of “what it’s like.” I see no reason, however to suppose that this is secondary to sensory experience (as in Lormand), or that it involves the detection of special properties (as in Pitt).

23. Peacocke nicely explains how this accounts for, e.g., Hume’s despair in finding any item in his experience that corresponds to his “self”, and Wittgenstein’s (1921:#5.641) otherwise enigmatic remark that the self is “not part of the world” –without drawing their extravagant metaphysical conclusions.
Chris Peacocke has reminded me also of the perception of at least Western Classical music, of, e.g., the seventh as leading to the tonic; a tritone needing resolution, or simply the structural units plausibly indicated by Shanker (18??) and Jackendoff and Leh Dahl (19??), all without peculiar qualia corresponding to the leading or the needing or the structural units. Correlative to my claim that there’s more to phenomenology than phenomenality might be the claim that there’s more to introspection than, so to say, meets the introspective ear. If you want to understand even your own introspectible conscious experience, you’d better not rely on introspections alone, but pay close attention to the researches of linguists and perceptual psychologists!

What of introspection itself? Is it a form of the "inner perception"? Here, too, I have no serious theory, but am content to rest on the obvious evidence that there is some such process by which people consciously attend to some of their own mental states, my only contribution being that this process involves a sensitivity to some of the tags attached to the introspectible contents. A full theory here would need to distinguish "phenomenal" attention of the sort we can fasten on (supposed) qualia from a more "vacant" attention that one directs on merely oneself, without any experience of an "object" of one’s attention, as in the case of many attitudes, and, as I’ve said, structural descriptions of language and vision, and of indexicalities such as "I" and "now." It would also need to explain how, given a tagged representation, T(σ), or a constellation of such representations, can provide a basis for automatic self-ascription not only of an attitude type –hope or dread– but that attitude type with the content, [σ], expressed by the representation σ. As many have noted, deploying [σ] in an attitude ascription, A[σ], may just be a re-deployment of that very content as it occurs in the attitude itself. But, of course, it’s a special re-deployment that, arguably, changes something about how [σ] is being conceived: if I introspect the fear that Zeus is mad, my attention may well be not only on my fear, but on the content, qua content, [Zeus is angry], not on Zeus or anger themselves. For Frege (1892/1952), for example, the reference of expressions embedded in attitudes was, “obliquely,” their customary sense. It is not obvious how this special re-deployment comes about. Presumably it’s a consequence of having mental concepts and a “theory of mind” in which they explanatorily figure. However, this still wouldn’t make their application to oneself inferential in any third-person way. In one’s own case the appropriate concepts, replete with their “oblique” references, may simply be triggered by the tags.

But, for lack of a specific introspective mechanism, other breakdowns are likely to be spotty. It could happen, but probably fairly rarely for the standard attitudes people reliably self-ascribe (schizophrenics who experience their own thoughts as “inserted” by others may be a case in point; see Frith 1992). At worst, people might sometimes simply be inattentive, or deploy too crude a set of mental concepts. So, apart from defects of attention or the lack of mental concepts, it would be surprising if there were general introspective

24. Chris Peacocke has reminded me also of the perception of at least Western Classical music, of, e.g., the seventh as leading to the tonic; a tritone needing resolution, or simply the structural units plausibly indicated by Shanker (18??) and Jackendoff and Leh Dahl (19??), all without peculiar qualia corresponding to the leading or the needing or the structural units. Correlative to my claim that there’s more to phenomenology than phenomenality might be the claim that there’s more to introspection than, so to say, meets the introspective ear. If you want to understand even your own introspectible conscious experience, you’d better not rely on introspections alone, but pay close attention to the researches of linguists and perceptual psychologists!

blindness, since there would be no separate organ for attending to tags that could breakdown (which is perhaps part of what worried Shoemaker).

However, as already emphasized, there's no reason whatever to suppose people are infallible about their sensitivity to these tags, or even that the tags themselves or resulting self-attributions are always reliable. As mentioned earlier, people seem to be unreliable about distinguishing genuine memories from rational reconstructions of the past; and vivid mental images can sometimes be mistaken for actual visual experiences.26 And people are notoriously prey to self-deceptions, thinking they think and feel what they think others expect them to think and feel; or just don't have time or motivation for careful introspection. Indeed, come to think of it, there are plenty of reasons to ignore and even suppress introspection in the normal course of things, just as we often suppress the complex varying visual appearances of a landscape or the sea. An exquisitely Proustian sensitivity to the exasperating ups and downs of one’s inner life is a pretty sure way to fail in one’s outer one (Proust, after all, had trouble getting out of bed!)

More interestingly, people's ability to attend to and discriminate subtle or complex constellations could be as variable and potentially unreliable as their ability to discriminate (the effects of) different wines, beers or styles of music or art. But, as these comparisons suggest, introspection can probably be cultivated, attention, so say, “trained” on one’s internal states. One often becomes more sensitive in this way to one’s own sensitivities, and so might become not only more reliable about one's actual states, but even, pace Carruthers (2009a:6-7), also about when one is genuinely introspecting: i.e., we might be able to come to introspect introspection after all!

3. Conclusion

Carruthers has quite usefully raised and defended an Objectivism about self-knowledge that many psychologists and philosophers, like Shoemaker and Burge, are likely to regard as so absurd as to be refutable a priori. I have argued that this latter response is mistaken, and that Carruthers has raised a possibility that can only be addressed empirically in the way that he admirably does. But, empirically, Objectivism has not nearly been established. The confabulations it predicts are predicted by any introspective model that also allows for self-attribution from a third-person perspective (as, trivially, any will), and it faces serious challenges to account for highly reliable self-attributions over a wide range of attitudes, particularly of feelings in addition to sensations, and particularly in Terrific Meditative cases, which seem to be every bit as reliable as open-eyed ones. By contrast, a modest “tags” model faces no such challenges, and promises to capture in a natural, independently plausible way both the data and the phenomenology surrounding introspection, in a way compatible with many traditional Introspectionist views. Of course, it could turn out, pace Shoemaker and Burge, that some or even all people are self-blind; but, for the reasons I’ve provided, it seems doubtful we all are; and, moreover, if we aren’t, then a very simple explanation of why we aren’t, and why we are as reliable as we seem to be, is readily available.

26. Carruthers discussed one kind of case that has seemed to undermine the notion of conscious agency, and deserves a bit more discussion. These are the surprising experiments of Brasil-Neto et al (1992), in which subjects think they're making a decision to move a certain finger, and that that's the cause of their finger moving, when in fact the cause is an independent electrical signal that is not plausibly causing any such decision (see also Libet (19??) and Wegner and Wheatley 1999).. It does appear in these cases that subjects are postulating a decision that at least wasn't the cause of their act, and may, per Carruthers, only be a post hoc interpretation of the act. However, it's crucial to note that one decidedly can't generalize from such cases to agency in general. It may well be that small acts, such as merely moving a finger are initiated in the motor cortex prior to one's "decision" to make them, and that what appears to the decision is perhaps no more than an endorsement of the impulse. But this can't be the general story for the production of non-basic acts that are the result of reasoning, as when one "pushes the button that indicates the right answer" to a question that presumably the motor cortex wouldn’t even understand.
References

Adler, J. (2002), *Belief’s Own Ethics*, Cambridge University Press


— (2009a), “How We Know Our Own Minds: the Relationship Between Mind-reading and Metacognition,” with commentaries by various authors (some of which are selected below) and Carruthers’ replies, “Mind-reading Underlies Metacognition,” *Behavioral and Brain Sciences*, 32(2), pp1-62


Casteneda, H. (19??), ??


Devitt, M. (20??), replies to critics, *Croatian Journal of Philosophy*


Hurlburt, R. (1993), Sampling Inner Experience in Disturbed Affect, New York: Plenum


Peacocke, C. (forthcoming), lectures on self-consciousness, delivered at Institut Jean Nicod, Paris, France, Oct 2010


— (1997), Contemporary Philosophy of Mind, Oxford: Blackwell


— (2008), “(Even Higher-Order) Intentionality Without Consciousness,” Revue Internationale de Philosophie,


— (1996), The First-Person Perspective and Other Essays NY: Cambridge University Press


Wittgenstein, L. (1921), Tractatus Logico-Philosophicus, Routledge Kegan-Paul


Wright, C., Smith, B. and MacDonald, C. (1998), Knowing Our Own Minds, Oxford University Press

************************************