INTROSPECTION, PHENOMENALITY AND THE AVAILABILITY OF INTENTIONAL CONTENT

David Pitt
California State University, Los Angeles

I.

I believe there's a phenomenology – a "what it's like" – of occurrent conscious thought. I believe this because I believe that *any* conscious state necessarily has phenomenal properties. Further, I believe that the phenomenology of occurrent conscious thought is *proprietary*: it's a *sui generis* sort of phenomenology, as unlike, say, auditory and visual phenomenology as they are unlike each other – a *cognitive* phenomenology. I believe this because I believe that the conscious occurrence of any of the more familiar sorts of phenomenal properties is neither necessary nor sufficient for the occurrence of conscious thought. I also believe that the phenomenal character of a conscious occurrent thought (type) is *distinctive*: that is, distinct thought types have distinct cognitive phenomenal properties. I believe this because I believe that distinct conscious occurrent thoughts are introspectively discriminable not only from other types of conscious states, but from each other. Finally, I believe that the phenomenal character of a conscious occurrent thought (type) is *individuative*: that is, in virtue of its having the phenomenal properties it has, it's a *thought* (as opposed to some other kind of mental state) with a specific *intentional content*. I believe this because I believe that we're able to identify our occurrent conscious thoughts as the thoughts they are, consciously, introspectively and non-inferentially, and that phenomenal properties are the only properties of occurrent conscious states that are so identifiable. In fact, I believe that the cognitive phenomenal character of an occurrent conscious thought *is* its intentional content. I believe this because I can (fear of psychologism turns out to be irrational), and because it provides for a simpler account of content: the

intentional content of a conscious thought is like the sensational content of a conscious pain – they are the states they are not because of their relational properties, but because of their intrinsic phenomenal nature.

Amen.[1]

I will call the view that there's a proprietary, distinctive and individuative phenomenology of occurrent conscious thought the *phenomenal intentionality of thought thesis* (PITT, for short). Not many people believe PITT. But skepticism about it seems to me more often based on prior theoretical commitment, or overreaching confidence in the explanatory resources of contemporary Naturalism (what Charles Siewert (this volume) calls "the tyrannizing anxieties and ambitions of mind-body metaphysics"), than on unbiased reflection upon our conscious mental lives, or careful evaluation of the arguments in its favor.

In this paper I consider three more reasoned lines of resistance to PITT, the first advanced by Joe Levine (this volume), the second taking its cue from extrospectionist views of self-knowledge, and the third stemming from concerns about introspective availability of intentional content. I argue that none of these challenges constitutes a serious threat to the thesis that the intentionality of thought is proprietarily phenomenally constituted.

I begin, however, with a new argument for cognitive phenomenology.

II.

For all *non*-cognitive kinds of mental states, sameness and difference *within* consciousness are entirely phenomenally constituted. The various modes of conscious *sensory*

[1] I provide more detailed arguments for these claims in Pitt 2004 (*PC*) and Pitt 2009.

experience, for example, are, *qua* conscious,[2] constituted by their proprietary kinds of phenomenology. Conscious *visual* experiences, *qua* conscious, share a particular kind of phenomenology that makes them *visual*, and distinguishes them from conscious experiences in all of the other modes. What it's like to have a conscious experience of yellow is the same, *qua* visual, as what it's like to have a conscious experience of green (or of any other visible property), *qua* visual, and it's the visual kind of phenomenology that makes them both, *qua* conscious, visual experiences. To be a conscious visual experience is to be conscious in the visual way – to have conscious visual phenomenology. Any conscious experience that has this kind of phenomenology is, necessarily, a conscious visual experience, and no conscious experience that lacked it could be a conscious visual experience. There's a proprietary visual mode of conscious experience, and it's phenomenally constituted.

Likewise, there's a proprietary kind of conscious *auditory* experience, and it's also phenomenally constituted. A conscious experience of the sound of thunder is, *qua* conscious, of the same general kind as a conscious experience of the sound of a c minor triad; and their sameness *qua* conscious *auditory* experiences is due to their shared auditory phenomenology.[3] Any conscious experience that has this kind of phenomenology is, necessarily, a conscious auditory experience, and no conscious experience that lacked it could be a conscious auditory experience. There's a proprietary auditory mode of conscious experience, and it's phenomenally

---

[2] That is, *as manifested in consciousness* – as opposed to, say, as caused by, or realized in, different sorts of brain states, or as occurring on a Thursday.

[3] I intend definite descriptions used in characterizing experiences to be rigid designators of particular kinds of experience, regardless of their causes ('the sound of thunder' in 'an experience of the sound of thunder' designates a particular kind of auditory experience, whether or not it is caused by the relevant atmospheric phenomenon).

constituted.

The same is true of all the other kinds of conscious sensory experience (olfactory, gustatory, tactile) we are capable of, as well. Each is, *qua* conscious,[4] constituted by its own general kind of phenomenology (olfactory, gustatory, tactile), and differs from all the others in virtue of its phenomenal kind. A conscious experience of the sound of thunder is different from a conscious experience of the smell of burning hair in part because of the intrinsic differences between auditory and olfactory phenomenology. Sameness within the various modes of conscious sensory experience, and the differences between them, are phenomenally constituted.

Further, differences within the various modes of conscious sensory experience are also phenomenal differences. A conscious experience of the smell of burning hair is of a kind different from a conscious experience of the smell of fresh basil, in virtue of their differing olfactory phenomenologies: what it's like to smell burning hair is different from what it's like to smell fresh basil. They differ as kinds of conscious olfactory experiences because of their distinctive phenomenologies. A conscious experience of the taste of sugar is different from a conscious experience of the taste of salt in virtue of the difference in their distinctive gustatory phenomenal properties. It's the difference between sweet and salty phenomenologies that makes them different types of gustatory experiences.

Finally, the phenomenology of a conscious experience makes it the kind of conscious experience it is. Differences in kinds of phenomenology between and within the various modes of conscious sensory experience make them different kinds of conscious experiences; but what individuates a conscious experience, *qua* conscious, is also its phenomenal character. A

[4] I will try to stop saying this (though I will continue to mean it).

conscious experience of the feel of an unshaved chin is different from a conscious experience of the feel of polished marble in virtue of their differing tactile phenomenologies. But it's also the case that a conscious experience of the feel of an unshaved chin is the particular kind of experience it is because of its phenomenal character. Nothing that felt like *that* could be a conscious experience of the feel of polished marble, and, necessarily, any conscious experience that feels like that is a conscious experience of the feel of an unshaved chin. No conscious experience that lacked thundery auditory phenomenology could be a conscious experience of the sound of thunder, and any experience that has it is, necessarily, an experience of the sound of thunder. Likewise for the visual experience of green, the olfactory experience of the smell of burned hair, the gustatory experience of the taste of salt, etc.

Similar considerations could be adduced with respect to all of the further determinates of these determinable sensory experiences, as well as all of the other familiar kinds of conscious experience, e.g., somatic, proprioceptive, emotional, etc. They're all, *qua* conscious experiences, individuated and identified by their proprietary, distinctive and individuative phenomenologies. To *be*, in consciousness, is, for these kinds of experiences, to be *phenomenal*. A conscious experience can't occur unless some phenomenal property is instantiated, and *which* phenomenal property is instantiated determines which kind of conscious experience (up to maximal determinateness) has occurred. In short, in all of these cases consciousness *supervenes on* phenomenology: difference in consciousness entails phenomenal difference, and sameness in phenomenology entails sameness in consciousness.[5]

[5] Which is not to say that phenomenality entails consciousness. Given that phenomenality without consciousness is possible (as I believe it is; see Pitt MS1), to say that sameness of phenomenology entails sameness in consciousness is to say that *if* two experiences are phenomenally identical, then *if* they're conscious they're type-identical conscious experiences.

Now, that these principles of phenomenal individuation should be applicable to all kinds of conscious states (*qua* conscious) *except* conscious thoughts is, at the very least, improbable. Given that they apply across such a wide range of so radically different kinds of states of consciousness, surely the burden of proof falls on anyone who claims that conscious thinking is exempt. Why should it be so different?

Furthermore, if conscious thoughts don't have proprietary, distinctive and individuative phenomenologies, then they would have to be conscious in some way *other than* phenomenally. But what could that be? How could a state be conscious – i.e., be *manifest* in, or *appear* in, consciousness – without being conscious (or *appearing*) in some *way* or other? And what could such *ways* be if not phenomenal properties? Again, minimally, the burden of proof is on anyone who would claim that there can be consciousness without phenomenality.

The only attempt I know of to make such a case is Lormand 1996; but Lormand's efforts are unsuccessful. (My reasons for thinking so are given in *PC* (23-24).) Moreover, claiming that conscious thoughts are "access" conscious without being "phenomenally" conscious won't help here, since to be available for conscious use is not *per se* to be *in* consciousness. (See also pages {17-18}, below.) And to say that something is *in* consciousness is not *per se* to say that it's phenomenal. Even if it's necessarily the case that conscious states have phenomenal properties, it's *not* the case that consciousness and phenomenality are *identical* (that 'conscious' and 'phenomenal' have the same *meaning*.) They're not the same property. If they were, then phenomenality *sine* consciousness would be impossible (which it isn't), and Lormand's claim that some conscious states lack phenomenality would be *prima facie* incoherent (which it isn't. It's incorrect, and necessarily so; but it's not obviously contradictory). Further, since

consciousness is a *unitary* (non-determinable) property, if phenomenality were identical to it, then all conscious states would be phenomenally identical. But they're not: phenomenality is a *determinable*.[6] So it can't be the same property as consciousness. Hence, the claim that some thoughts are conscious is not, in this context, trivially question-begging. It's not *trivially* true that conscious thoughts have phenomenal properties, since it's not trivially true that conscious states in general have phenomenal properties. (Cf. *PC*: note 4.) And the claim that some thoughts are conscious is not just the claim that some thoughts are phenomenal in thin lexical disguise.

Thus, if conscious cognitive states (thoughts) constitute their own general kind – if they differ from all other kinds of conscious mental states – then they must enjoy their own proprietary sort of phenomenology. Pains are not tastes, sounds are not smells, visual experiences are not moods, in virtue of having different proprietary phenomenologies. Hence, if thoughts are not pains or tastes or sounds or smells or visual experiences or moods or ..., then they must have a proprietary mode of conscious existence – a proprietarily *cognitive* phenomenology. If they *are* different sorts of conscious states, then, *qua* conscious, they must be *phenomenally* different.

But conscious thoughts *can't be* identified with any other sort of conscious states (the most plausible candidate being conscious verbal imagery), since it's possible for any such to occur in the absence of thought. (Thinking is *not* the same as producing internal sentence tokens,

[6] I don't think phenomenal properties are determinates of consciousness, since I think phenomenality doesn't entail consciousness. Phenomenality is necessary, but not sufficient, for consciousness. (Thanks to Declan Smithies for pressing me to get clear on this.)

as anyone who has read Derrida should be able to tell you.[7]) Thoughts are states of a different kind from all others. Hence, there must be a proprietary phenomenology of *cognition* – a proprietarily cognitive way of appearing; a phenomenology that makes a state cognitive, as opposed to visual, auditory, olfactory, somatic, proprioceptive, etc.

And if there are different types of conscious thoughts, then each distinct type must have its own unique mode of conscious existence. Thus, the phenomenal properties of distinct thought-types must be sufficient to distinguish them from each other (as well as from all other kinds of conscious states), just as the phenomenal properties of different smells or sounds or color experiences must be sufficient to distinguish them. If a conscious thought *t* is to *be* a different thought from a conscious thought *t′*, then *t* and *t′* must have distinctive cognitive phenomenal characters.

Finally, since in general conscious states are the states they are in virtue of their proprietary and distinctive phenomenologies, the cognitive phenomenology of a conscious thought must be individuative as well. A conscious thought is a thought, and the thought that it is, in virtue of its distinctive cognitive phenomenology. Moreover, if thought types are individuated by their contents, then thought contents are cognitive phenomenal properties. Each thought that *p*, *q*, *r*, ..., where *p*, *q*, *r*, ... are different contents, has a proprietary, distinctive and individuative phenomenal character that constitutes its intentional content. (See Pitt 2009.) Not, as I argued in *PC*, because this is the only way we can *know* what they are, and discriminate them

[7] And, speaking of Frenchmen, in virtue of *what* do I and Jacques think the same thought when I inwardly utter "Paris is beautiful but boring" and he inwardly utters "Le Paris est beau mais ennuyeux"? Presumably it's our *meaning* the same thing (i.e., intending to express the same thought) by our utterances. But our utterances are different; so, meaning is not the same as uttering. *Qua* conscious, then, meaning and inner-uttering must be phenomenally different.

from each other in introspection (though I still think this argument is sound), but because, *qua* conscious, this is the only way for them to *be* what they are, and to be different from one another and all other kinds of conscious states. In consciousness, *esse est pareo*.

III.

In his contribution to this volume, Joe Levine maintains that the argument from self-knowledge in *PC* doesn't establish that there's a proprietary phenomenology of cognition, since the kinds of self-knowledge it's introduced to explain can be explained without it. He begins his critique by making a distinction between *implicit* and *explicit* self-knowledge of thought. Implicit self-knowledge

> is not the result of any explicit formulation or reflection. Rather, it's the knowledge that seems to come with the very thinking of the thought itself. ... All that's required is that one thinks in one's language of thought, mentalese. To implicitly know what one is thinking is just to think with understanding. [{11}]

Explicit self-knowledge, in contrast,

> is what we have when we explicitly formulate a meta-cognitive thought, such as "I believe that San Francisco is a beautiful city," [*id*.]

and is explicable in terms of

> the reliability of the relevant process yielding the higher-order sentence expressing the fact that one is thinking a certain content. [{8}]

He maintains that implicit and explicit self-knowledge, so construed, are all we need to explain self-knowledge of thought, and that neither requires a special phenomenology of cognition.

I think the distinction between implicit and explicit self-knowledge is very useful; but I don't think it can explain self-knowledge of thought if it's understood in Levine's terms.

For one thing, Levine's account engenders something of a dilemma. If representational content is *extrinsic*, then one can't have implicit knowledge of it simply by tokening a representation – any more one could have knowledge of the meaning of a sentence simply by inscribing it. Presumably it's the tokening of a representation in one's mind/brain that puts one in a relationship to it intimate enough to engender implicit knowledge of it. But if the content of the representation isn't tokened with it – if it's not intrinsic to the representation – then there can be no such intimate relationship with content. Hence, it would seem, in order for tokening to constitute implicit knowledge of content – i.e., in order for tokening to be thinking with understanding – content must be an intrinsic feature of representations. But this does not sit well with the sort of computational picture Levine is appealing to, on which contents are, typically, taken to be extrinsic.

On the other hand, if understanding a representation is knowing what its content is, then supposing that content is extrinsic leads to regress, for it entails that all knowledge of content is explicit, and explicit knowledge, on Levine's account, requires tokening of further representations. Knowledge of extrinsic content of a representation would be achieved through the tokening of a second-order representation that explicitly attributes content to it. But if one doesn't know, at least implicitly, what the content of the second-order representation is – if one doesn't think it with understanding – then one won't know the content of the first-order thought. One won't know what one has thought about it. However, *ex hypothesi*, content is extrinsic, and, hence, knowledge of it must be explicit. So knowing the content of the second-order representation requires tokening a third-order representation that explicitly attributes content to *it*. *Et cetera*. It seems the only way to avoid such a regress would land Levine back in the

intrinsicalist soup. He would owe us some sort of theory of intrinsic contents, which would be *prima facie* in tension with his computationalist outlook. And, given Levine's rejection of syntax and semantic phenomenology as candidate implicit *content* discriminators, it's hard to see what sorts of intrinsic properties he might enlist for the job.

Moreover, I don't think Levine's account can explain *conscious* implicit knowledge of thought content, since such knowledge requires the instantiation of properties sufficient to individuate content in consciousness; but, since (as argued above) conscious states are individuated by their phenomenal properties, such knowledge requires a distinctive phenomenology of content.

Mere occurrence of a mental state couldn't constitute conscious implicit self-knowledge of it unless the occurrence were itself conscious, and consciousness requires phenomenology. Even if there were some sense in which mere tokening of a mental representation whose content is that *p* counts as implicit knowledge that one is thinking that *p* – i.e., that the computational system "knows" which representations are being tokened – this in itself doesn't explain how *I* can implicitly know what *I'm* consciously thinking. You can't have implicit conscious knowledge of what you're thinking in virtue of tokening an unconscious mental representation. Levine's account doesn't seem to allow for there to be an *epistemic* difference between conscious and unconscious thinking. Any occurrence, conscious or not, of a mental representation counts as implicit knowledge of its content, and an occurrence of a relevant meta-representation of it counts as explicit knowledge of its content. So it seems that consciousness makes no difference to what I can know about what I'm thinking (unless Levine is advocating a higher-order theory of consciousness, on which thinking about a thought *makes it* conscious; which I don't think is the

case). But it does make a difference. There's a perfectly good sense in which I *don't* know what I'm thinking, believing, fearing, desiring, etc. if it's unconscious, and I *do* come to know it when it becomes conscious. Surely Freud wasn't wrong about *that*.

In addition, without characteristic phenomenal differences among occurrent conscious states, implicit self-knowledge couldn't be discriminative – that is, you couldn't be implicitly consciously aware *that* you're thinking, or of *what* you're thinking. Implicit knowledge of conscious experience requires implicit individuation of experiences, which, in consciousness, is purely phenomenal. One can't consciously implicitly know what one is experiencing unless the experience is implicitly discriminated in consciousness from all others. Hence, there must be a proprietary, distinctive and individuative phenomenology of conscious thoughts if one is to have implicit conscious knowledge of them.

Now, Levine seems to maintain that we simply *don't have* conscious implicit knowledge of the contents of our thoughts (cf. note 5). Indeed, he seems to concede that such knowledge would require a distinctive phenomenology of content, and, hence, that to suppose that we have it begs the question against opponents of cognitive phenomenology. And he suggests that all we have implicit conscious knowledge of is the *vehicles* of thought – sentences of mentalese.

But I don't think the claim that we're acquainted with the contents of our conscious thoughts begs the question. And discrimination of the vehicles of thoughts, as such, is not sufficient for distinguishing the thoughts themselves, given that thoughts are individuated by their contents.

In *PC* I characterized acquaintance with content as direct, non-inferential conscious knowledge of what we're consciously thinking. This does not, *per se*, presuppose

phenomenology. It does need an explanation, however, and I provided an argument that the only available one requires a proprietary phenomenology of content. Though some have tried to show that consciousness, and, presumably, conscious acquaintance, don't require phenomenology, I think this position is untenable. And I considered what I took to be the most promising non-phenomenal account, based on a reliabilist-computationalist theory of knowledge and belief (of the kind Levine appears to favor), and argued (as I do here) that such theories can't ground an explanation of direct access to conscious contents.

Claiming that it's question-begging to affirm direct introspective knowledge of what we're consciously thinking *because* it turns out that the only explanation for it appeals to cognitive phenomenology, is like claiming that it's question-begging to affirm direct introspective knowledge of what we're consciously feeling, because it turns out that the only explanation for it appeals to somatosensory phenomenology. It's plausible only if you've already decided that the phenomenology in question doesn't exist. It's not question-begging to maintain that what's required in order to explain a capacity we have in fact exists. And it does seem to me to be non-tendentiously, almost platitudinously true that we can have non-inferential conscious knowledge of the contents of our occurrent conscious thoughts – of what we're occurrently consciously thinking – and, hence, that providing an alternative explanation for it is a far better (though in the end doomed) strategy for resisting cognitive phenomenology than denying its existence. The connection between conscious acquaintance and phenomenology is very close. But this doesn't make it question-begging to assert it. It only makes it scandalous that it's been overlooked in the case of conscious thought.

Levine claims that all the phenomenal contrast that is required for such discrimination is a

contrast between non-semantic features of mental representations. Phenomenal contrast in the case of thinking different readings of an ambiguous natural-language sentence is just the contrast between tokening distinct sentences in mentalese which represent (unambiguously) the two readings. We tell our thoughts apart by distinguishing non-semantic features of their vehicles that *track*, but do not *evince* their contents. The contents of our thoughts are not individuated in experience, only their mentalese vehicles. There is a phenomenology of cognition, but it's not cognitive, it's a kind of linguistic phenomenology, where the relevant language is the language of thought.

I don't know if mentalese tokens have distinctive phenomenal properties. I rather doubt it, since they're supposedly subpersonal, computational entities. But, be that as it may, I don't see that Levine's account can explain knowledge of what we're thinking. We have direct conscious access to (unambiguous) mental representations of contents, which are individuated in consciousness by their phenomenal features. But these features are not the contents themselves; nor are they sufficient for determining what the contents are. So direct acquaintance with them can't explain first-person non-inferential knowledge of what one is consciously occurrently thinking.

As I understand it, the implicit-explicit distinction with respect to conscious knowledge is the distinction between acquaintance *as* knowledge – or "acquaintance-knowledge," as I'll call it – and knowledge *by* acquaintance; and I maintain that the mechanism by which beliefs about one's conscious experience are formed is in at least some cases not computational or automatic. Acquaintance-knowledge of a mental state consists simply in its conscious occurrence. (One can't be acquainted with (directly aware of) unconscious mental states.) In undergoing a

conscious experience, one has implicit knowledge of what one is experiencing (though not, *per se*, *that* one is experiencing it), and of what the experience is like. If I've tasted uni, then I know what uni tastes like (what it's like to taste uni), even if I didn't know it was uni I was tasting. Knowledge *by* acquaintance of a mental state, on the other hand, is knowing *that* one is in it, and requires application of concepts and formation of beliefs. If, after my first, innocent taste of uni, someone tells me that it was uni I ate, then I can know that I'm tasting (or have tasted) uni, and that I know what uni tastes like.

Moreover, neither sort of knowledge can be reduced to the other; and knowledge by acquaintance (knowledge-*that*) of conscious experiences presupposes acquaintance-knowledge (knowledge-*what*) of them.

When Mary leaves the Black and White Room, she comes to know what it's like to see red when she experiences it. In having the experience of red, she acquaintance-knows what seeing red is like. If she's re-imprisoned (poor thing) and her memory fades, then she'll no longer have this kind of knowledge – she'll no longer know what it's like to see red. Further, Mary's first experience of red would not constitute implicit knowledge if it weren't conscious, since if it weren't conscious it would make no difference to *her* (she couldn't be said to be acquainted with it). If her color experiences upon her release from the Black and White Room were unconscious, she would notice no relevant changes in her experience, and could not be said to have learned anything new. She still would not know what it's like to see red, or any of the other colors. And if her experience of red were not phenomenally different from her experiences of the other colors, her implicit knowledge could not be said to be of what it's like to see *red*, as opposed to another color.

Knowing what it's like can't be reduced to any form of knowing that. Knowing what it's like to see red is not knowing that seeing red is like *this* (or that *this* is what it's like to see red), where an instance of a phenomenal property is the referent (or perhaps a constituent[8]) of the concept THIS, since one may have an experience of red, and thus acquaintance-know what it's like to see red, without being able to categorize it conceptually – i.e., without being able to think that one is experiencing *red*, or that *this* is what it's like to experience *red*. This may be because one lacks the appropriate concept (which may be because it's not in the human repertoire), or because one doesn't know how to apply it to one's experience. As Nida-Rümelin has shown (Nida-Rümelin 1995), Mary could know what it's like to see red without knowing that it's red she is seeing, if she's trapped in (what I call) the Technicolor Vestibule – that place between the Black and White Room and the Wide Chromatic World, where there are colors but no familiar objects from which she could infer which color is which.

Of course, once she's out of the Technicolor Vestibule, Mary can also have explicit propositional knowledge of her psychological states: she can know *that* she's seeing red. But this presupposes acquaintance-knowledge of what it's like to see red. She can't introspectively know *that* she's seeing red if she doesn't know *what* seeing red is like. Knowledge by acquaintance of

---

[8] Some philosophers maintain that demonstrative 'phenomenal concepts' have sample experiences embedded in them, and are thus partly self-referential (see, e.g., Chalmers 2003). Before she has experienced red, Mary can't have this sort of concept, and so can't know that red looks like *this*. I prefer a view on which concepts and sensations are kept separate, so that the thoughts Mary might have in the technicolor vestibule (the brightly colored antechamber containing no recognizable objects that she is released into before getting out into the world (see Nida-Rümelin 1995)) *I wonder if red looks like this* (demonstrating a red patch) and *I wonder if red looks like this* (demonstrating a green patch) have the same conceptual content, but different truth conditions due to the different referents of 'this'. (I defend a general account of demonstrative and indexical concepts along these lines in Pitt MS2.)

conscious states presupposes acquaintance-knowledge of their distinctive phenomenologies. Mary's experience of red must be (implicitly) discriminated from other experiences; and she must apply concepts to it and form a belief about it in order to know that she's having an experience of red.

So the (revised) argument from self-knowledge is this: Immediate knowledge-that of conscious thought requires knowledge-what, and knowledge-what requires distinctive phenomenology. Knowledge-what consists in simply the conscious occurrence of the thought.[9] (This is what I called simple (non-epistemic) introspection in *PC*, though I now think it *is* epistemic in the sense that it constitutes a kind of knowledge. It's just not conceptual or doxastic.) One has implicit knowledge of the second-order thoughts whose occurrence constitutes explicit knowledge of the contents of first-order thoughts. I don't, and I don't think I ever did, "assimilate these two forms of self-knowledge," as Levine suggests – though initially I didn't think of implicit occurrence as a kind of self-*knowledge*. Immediate knowledge-that consists in beliefs about one's mental states formed on the basis of conscious acquaintance with them, which is just their conscious occurrence. One *recognizes* what one is thinking – just as one recognizes what one is hearing or smelling or seeing – and applies the relevant concepts and forms the relevant beliefs. The recognition is neither conceptual nor inferential, and the formation of the relevant beliefs, while of course conceptual, isn't inferential either.

No doubt Levine would still consider all of this question-begging, since he maintains that we can be "as it were, magically" (*OPT* {10}) aware of our occurrent conscious thoughts (i.e. (I

[9] Hence, my view is not committed to the regress Levine charges it with (*OPT*: {11-12}). One has implicit knowledge of the second-order thoughts whose occurrence constitutes explicit knowledge of the contents of first-order thoughts. I don't, and never did, "assimilate these two forms of self-knowledge" (*OPT*: {11}).

suppose), we're privy to the results of a computational process, but not to the process itself), without invoking "phenomenal appearance, as we have with sensory experience" (*id*.). A conscious thought occurs; a mechanism that can register which thought it is causes me to believe that it's that thought (tokens a mentalese sentence that expresses the fact that one is thinking it), and if the mechanism is reliable, the belief will count as knowledge. There's no work here for a proprietarily cognitive phenomenology to do.

But it's not the case that we *always* "as it were, magically" know what we're thinking or feeling – that the belief about our experience just pops into our head. We often *recognize* what we're thinking or feeling, *identify* it on the basis of its recognizable properties, and self-ascribe it. We make *voluntary judgments* about the contents of our consciousness on the basis of recognition of their distinctive phenomenologies. We're consciously aware, not just *that* we're in a particular conscious state, but *of* the state itself. Sometimes I come to have a belief about what I'm experiencing on the basis of attending to it and recognizing what it is. *This* is the kind of self-knowledge the argument in *PC* is concerned to explain. Maybe there's a reflex "I'm in pain!" that pops into my head when something hurts me. But I can also, so to speak, browse around in my conscious mind (selectively attend to the contents of my consciousness) and attend to things that are there (the song that's been in my head all day, the ringing in my ears, the thought that I'm condemned to be free). I may or may not form the thought that I'm in any of these states; but if I do, it seems that I can do it *voluntarily* – just as I might absent-mindedly (thoughtlessly) be looking at an orange flower, and then think to myself: "That's an orange flower." The seemingly automatic belief-forming mechanism story can't explain this.

The issue between me and Levine here is not whether or not there are conscious experiences, or whether or not we can have introspective knowledge of their occurrence and nature. We seem to agree on this. Our difference concerns, rather, *how* beliefs about experience are formed. Levine is claiming that they're *always* formed by a reliable, automatic belief-forming mechanism. My claim is that, whether or not there are beliefs about experience formed in this way, we can *also* voluntarily form beliefs about our conscious experiences on the basis of active introspection, and that this presupposes that we have some way of identifying and distinguishing them from each other, *qua* conscious. But the only properties of conscious experiences that can serve to distinguish them *qua* conscious are phenomenal properties (because these are the only intrinsic properties that conscious experiences as such can have). So, given that it's possible to gain self-knowledge of thought in this way, there must be a distinctive phenomenology for thoughts – a *cognitive* phenomenology. The activity of an automatic belief-forming mechanism can't, *qua* automatic, explain this sort of self-knowledge.[10]

Moreover, even in cases where a computational mechanism spontaneously informs me what state I'm in, by producing a thought about it in the appropriate way, there's *still* a need for distinctive cognitive phenomenology. For, unless I know what the mechanism has "said" – if, so to speak, the message that has magically appeared on the "belief-board" isn't legible, or I don't know what it means – I won't know what state I'm in. (Especially if the message is in *mentalese*,

[10] It would, I think, be very odd to suppose that all of our knowledge of our conscious occurrent *sensory* states is automatic, since this would render the phenomenology of such states irrelevant to our knowledge of them – we would come to know that, for example, we're hearing the dinner bell *in the absence of conscious auditory phenomenology*. It seems much more plausible that one recognizes the sound of the dinner bell, and on that basis comes to believe that the dinner bell has rung.

which no one can read.) Levine suggests that I rely on an (in the context) unduly "rich" – i.e., phenomenal, hence question-begging – conception of consciousness in the argument in *PC*. But the alternative is to suppose either that thoughts are never conscious at all, or that they're only ever *access* conscious. The first disjunct is a non-starter. On the second, for a thought to be conscious is just for it to be *available for use* in control of reasoning and behavior. But it's difficult to see how a thought could actually *be used* in *conscious* control of reasoning and behavior without the user being conscious of its content in a *non*-access sense (i.e., without being *acquainted with it*). And I can't see how *that* could be explained if all cognitive consciousness were access consciousness. Non-inferential introspective awareness of the contents of conscious states requires phenomenology. Even if the meta-cognitive belief is implicit – its mere occurrence on the belief-board constituting implicit knowledge of its occurrence and content, there's still a need for a distinctive cognitive phenomenology. As argued above, implicit conscious knowledge requires individuative phenomenology. Implicit knowledge that I'm consciously thinking that *p* (and not that *q*, *r*, *s*, ...), like implicit knowledge of any other sort of conscious mental state, requires a distinctive phenomenology. So the automatic belief-forming mechanism story doesn't really provide an alternative to the phenomenally-based account. It doesn't explain everything that needs to be explained.

IV.

A different line of resistance to PITT takes its cue from *representationalist*[11] views of self-knowledge. On more or less traditional views, self-knowledge of the qualitative content of

[11] Of the *reductive* variety. See, e.g., Byrne 2001, Dretske 1995, Harman 1990, Tye 1995, 2000, Lycan 1996. See Chalmers 2004 for the distinction between reductive and non-reductive representationalism.

perceptual experience is achieved through introspection, which is a kind of *inner sense*, directed at experience itself. The sensational content of perceptual experience is the qualitative properties (qualia) it instantiates, and one comes to know how one is experiencing through inner acquaintance with them. On representationalist views, on the other hand, self-knowledge of experiential content arises from a kind of extrospection: *deferred perception*. One comes to know what the content of one's perceptual experience is by focusing *outward*, on its objects and their properties. The qualitative content of perceptual experience is a kind of *representational* content, which is in part constituted by the external qualitative properties it represents. Knowing *what it's like* to see a clear sky at noon is not a matter of inspecting one's experience and finding a blue quale there, but of observing the sky and seeing that it looks blue. Given that the sky looks blue, you can conclude that you are representing it as such. In describing what it's like, one describes, not a property of one's experience, but a property of the object of one's experience – *the sky*. The blue is relevant to the characterization of the qualitative content of the experience because the experience *represents it*, not because it *instantiates it*: one's experience of the sky is no more blue than one's *thought* about it is.

(Motivation for such views comes from the alleged "transparency" of perceptual experience (when we attempt to inspect our experience, we find only extramental objects and their properties), from broadly physicalist scruples (qualia are mysterious, non-physical things whose relation to the brain is, at best, problematic: there's nothing blue in your brain when you are looking at the sky), and from certain epistemological worries.[12])

[12] Michael Tye (Tye 2000, 46) argues that "[t]o suppose that the qualities of which perceivers are directly aware in undergoing ordinary, everyday visual experiences are really qualities of the experiences would be to convict such experiences of massive error. That is just not credible. It seems totally implausible to hold that visual experience is systematically misleading in this way."

Thus, were one to argue that experiences must instantiate phenomenal properties because we come to know *how* we're experiencing by examining experience itself, and hence that the properties that constitute how we're experiencing must be there to be detected, the representationalist would counter that self-knowledge of experiential content can be otherwise explained. It's not based on inner sense, and there are no mysterious inner qualitative properties of experiences to be discovered.

Now, one might think that a similar strategy could be adopted in opposition to PITT. In *PC* I argued that, though it may not be readily apparent to some, there in fact *must* be a phenomenology of cognition, if a certain kind of introspective knowledge of content is to be accounted for. I maintained that if thoughts are to be distinguishable *to inner sense* from each other and from other kinds of mental states, then they must have their own proprietary, distinctive and individuative qualitative character. They must "show up" (appear) in consciousness in unique ways. But if the inner-sense view of knowledge of the contents of occurrent conscious states has been discredited, then the argument, which depends upon it, is fatally weakened: I haven't provided a reason to believe in cognitive phenomenology.

I think reductive representationalist theories of the qualitative content of perceptual experience founder on the twin hazards of dreams and hallucinations; though I won't argue for this here.[13] Rather, I want to try to show that extrospectionist theories of self-knowledge of

---

On the contrary, I would argue, if visual experience were *not* systematically misleading – if it didn't present itself as something it's not (viz., external reality) – it would be *useless*. Transparency is an illusion made necessary by the facts that what experience is supposed to represent is external to the mind, while experience itself is internal. Perceptual experience cannot present itself as what it *is* if it's to be a naively believable guide to what it's *not*.

[13] Proponents are reduced to talk of representation of uninstantiated universals and objects in non-actual possible worlds, neither of which seem to me to be representable in the right sort of

perceptual content and propositional attitudes in fact do not succeed in explaining how it is that we can know that we're experiencing, believing or desiring, and that in any case there's no plausible way to extend them to cover knowledge of intentional content.[14] Thus, what might seem like a good strategy for thwarting PITT doesn't get off the ground.

Extrospectionist theorizing about self-knowledge of propositional attitudes takes its cue from Evans, who (interpreting a remark of Wittgenstein's) writes (1982: 225):

> [I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward – upon the world. If someone asks me "Do you think there is going to be a third world war?," I must attend, in answering him, to precisely the same outward phenomena I would attend to if I were answering the question "Will there be a third world war?"

Evans's view has recently been developed by Richard Moran, in his book *Authority and Estrangement*.[15] Moran generalizes Evans's claim, and couches it explicitly in terms of transparency:

> With respect to the attitude of belief, the claim of transparency tells us that the first-person question "Do I believe P?" is "transparent" to, answered in the same way as, the outward-directed question as to the truth of P itself. [Moran 2001: 66]

Here, as in the case of perceptual experience, one determines the contents of one's mental states by looking *outward*. If you want to know if you believe that justice is a virtue, don't look into your mind, but consider *justice* and its relation to *virtue*.

---

way, or to have the right sort of properties. (What is *deferred perception* of such things?) I develop these considerations further in Pitt MS3.

[14] I'll use the phrase 'intentional content' to refer to the contents of cognitive/conceptual states such as thoughts, beliefs and desires. Some philosophers think that non-conceptual states have intentional content as well; but I won't be discussing such states in these terms here.

[15] I'm indebted in this section to Byrne's discussion of Moran's views in Byrne 2005, from which I've also taken the Evans and Moran quotations.

Now, Alex Byrne (2005) has argued persuasively that Evans-Moran-style views, on which self-knowledge of belief is achieved by a process akin to decision-making, can't be the complete story.[16] As Byrne points out, there are many cases in which, when asked what one believes, one *already knows* the answer, and, therefore, doesn't have to *figure it out*:

> Consider the question "Do I live in Cambridge, Massachusetts?" or "Do I believe that Moran is the author of *Authority and Estrangement*?" These questions can be answered transparently, by considering the relevant facts of location and authorship, but I do not need to make up my mind. On the contrary, it is already made up. [85]

Byrne concludes that transparency, *per se*, "does not show that knowledge of one's beliefs is in general a matter of making up one's mind" (*id.*). He then goes on to develop an extrospectionist account of self-knowledge which, he claims, avoids the Evans-Moran limitation and explains both *privileged* and *peculiar* access to one's own intentional states.[17]

On Byrne's view, one comes to know what one believes by applying to oneself (or at least *trying to* apply to oneself) the *transparent* epistemic rule BEL (95):

(BEL) If *p*, believe that you believe that *p*

In order to establish the truth of the antecedent, one considers whether or not *p*, where *p* is, typically, not a proposition about oneself or one's mental state. One looks outward to determine the status of *p*, and recognizing it to be true, applies the rule and believes that one believes that *p*.

But how is it that considering whether or not it's the case that *p*, where *p* concerns facts not about oneself but about a mind-independent world, can support attributions of mental states

---

[16] Martin (1998) has objected along similar lines. (See also Gertler 2003/8.)

[17] Beliefs about one's own mental states are *privileged* in that they're more likely to yield knowledge than beliefs about the mental states of others, and *peculiar* in that they're acquired in a way one couldn't acquire beliefs about the mental states of others.

to oneself? This is the "puzzle of transparency." As Byrne puts it (*id.*), it seems that "surely [BEL] is a *bad* rule: that *p* is the case does not even make it *likely* that one believes that it is the case." Here one seems to be in the same situation with respect to oneself as one is with respect to others. BEL, it would seem, is just as bad as BEL-3 (96):

(BEL-3) If *p*, believe that Fred believes that *p*

Determining the truth value of *p* won't help at all in coming to know what Fred believes.

Byrne claims that the solution to the puzzle of transparency lies in the fact that "[o]ne is only in a position to follow BEL ... when one has recognized that *p*. And recognizing that *p* is (inter alia) coming to *believe* that p" (*id.*).[18] That is, the only conditions under which BEL can be applied to yield self-knowledge are those in which the that-clause of its consequent is true: one must *recognize that* p, where recognizing that *p* entails *believing* that *p*. Hence, BEL is *self-verifying*. *p* may be a mind-independent fact, but that one recognizes that *p* is not; it's a psychological fact about oneself, and as such justifies a psychological conclusion. In making *cognitive contact* with the fact that *p*, one licenses the inference to an explicit self-attribution of a psychological state – in a way that making cognitive contact with *p* would not license attribution of a psychological state to someone else. Given that one is in the proper circumstances – the circumstances of recognizing that *p* – one is justified in applying the rule and inferring (the that-

[18] Simply *entertaining* the proposition that *p* is not sufficient, since one can think that *p* without believing it. Suppose someone says "George W Bush was the greatest American president." You, incredulous, think: *George W Bush was the greatest American president. I don't believe that*. (You probably had to think it to understand what was said in the first place. Clearly, however, you don't have to believe what someone says in order to understand it.) This isn't paradoxical. The thought: *p. I don't believe that p* is paradoxical only if it's assumed that the initial *p* is an (inner) expression of a belief. (Likewise, the *sentence* '*p*. I don't believe that *p*' isn't paradoxical. If '*p*, but/and I don't believe that *p*' is paradoxical, it's (I would argue) because 'but' or 'and' somehow implies that the first utterance of '*p*' is an *assertion*. Not all utterances of declarative sentences are assertions, however, though that might be the default assumption.)

clause of) its consequent.

But now it seems BEL is no longer *transparent*. For, in order to *apply* it, you have to *know that* you're in the proper circumstances. It's one thing to *be* in the proper circumstances – for there to be a justification for the application of BEL; but if you don't know that you are – if you don't know that you have such justification, then you have no reason, no motivation, for applying the rule. Byrne likens application of BEL when one recognizes that *p* to application of the rule of necessitation (*p* □*p*) "whenever one is in circumstances in which the rule applies – whenever, that is, one is confronted with a proof whose initial premises are axioms." (95) But just as you would have no reason to infer □*p* from *p* unless you knew that *p* was occurring in a proof whose initial premises are axioms (why not infer *q*, as you would if *p* appeared in a proof below a line on which *p* appeared, under assumption of *q*?), you would have no reason to infer that you believe that *p* from the recognition that *p* unless you knew that you recognized that *p*. In the absence of such knowledge, you'd have no more reason to apply BEL to yourself than you'd have to apply BEL-3 to Fred.[19] But knowing that you recognize that *p* is knowing that you're in a psychological state; and so BEL is not transparent.[20]

---

[19] Of course there can be a reason to apply a rule that one is not aware of. But such lack of awareness would prevent one from applying it. The fact that a crane is about to drop a Steinway D on the street right where I'll be in ten seconds is a reason for me to cross the street. But if I don't know this – if I don't know that I'm in a situation in which the rule 'If you're about to be crushed by a concert grand falling from a crane, cross the street. Quickly.' is applicable – in the absence of some other motivation, I won't cross the street; I won't apply the rule. I will, in a perfectly ordinary sense, *have no reason to*.

[20] Or, more precisely, BEL', which makes explicit the conditions one must be in in order for application of BEL to yield self-knowledge, is not transparent:

(BEL') If you recognize that *p*, believe that you believe that *p*

Moreover, given that recognizing that *p* is, as Byrne notes, *inter alia*, believing that *p*, knowing that one is in a position to apply BEL is *already* knowing that one believes that *p*. But that's what's supposed to be achieved by the application of BEL. Byrne claims that "the puzzle of transparency is solved by noting that BEL is self-verifying" (96). But the puzzle is not solved, since one is not really looking outward after all. Additionally, the explanatory value of the theory is lost, since application of BEL presupposes the knowledge it's supposed to generate: the theory is viciously circular.[21]

It might be objected that one need not recognize that one is in proper circumstances for application of BEL in order to apply it and come to know what one believes, because its application is *automatic*: whenever you're in the circumstances of recognizing that *p*, some mechanism that implements BEL is activated, and forthwith you believe that you believe that *p*. Simply being in the proper circumstances *is* sufficient to trigger the relevant mechanism; it's not necessary that you consciously recognize that you are, or self-consciously apply BEL, in order for application of BEL to yield self-knowledge. My objection relies on an *internalist* conception of justification, which is not inevitable.

But the sorts of cases Byrne is trying to explain are not automatic. He's concerned with a process in which one *considers* how things are, *applies* BEL, and *concludes* that one believes that

---

BEL itself is transparent, in that its antecedent is neutral (contains no psychological terms); but Byrne's solution to the puzzle of transparency entails that BEL cannot be transparently *applied*.

[21] Though I've focused on Byrne's account, I think these problems arise for any theory (including extrospectionist accounts of perceptual self-knowledge; more on this below) that relies on transparency. The general problem is that external facts are only relevant when they're cognized, and one must know the *way* in which they're cognized in order to draw any inference about one's psychological state. But knowledge of the way in which they're cognized is what the inference is supposed to yield.

p – a conscious, voluntary process of coming to know what one believes. (Note that the consequent of BEL is an *imperative* – an instruction to *do* something, and that this is something one may *try* to do and *succeed in* doing.) Surely this is one way to come to know what one believes. Whether or not there's a mechanism of the other kind, its operation cannot explain such a process.

Moreover, mechanizing the inference doesn't obviate the transparency and circularity problems. Given that the mechanism needs as input not just the content that *p*, which could be believed, doubted, hoped, etc., but, again, the *mode* in which it's cognized, the rule it implements would have to have the form of (something like) AUTO-BEL':

(AUTO-BEL') If that *p* is recognized, dump a token of 'I believe that *p*' into the belief box

the antecedent of which is not about the world – the fact that *p* – but about the psychological state of recognizing that *p*. Hence, even if the inference is automatic (and/or unconscious), the implemented rule isn't transparent: its antecedent refers to a psychological state.[22]

Further, whether the inference is voluntary or automatic, the conditions under which the rule is applicable must be represented somewhere in the system. If there are a number of such mechanisms, each putting out a different kind of attitude with potentially the same content (belief that *p*, fear that *p*, hope that *p*, ...), they'll each require information about the mode in which the content is cognized – whether it's built into the rule explicitly, as in AUTO-BEL', or encoded somewhere "upstream" in some kind of input sorting mechanism, or otherwise represented. Without this information, the system won't "know" which routine to run, any more than a conscious, voluntary user of the rule. So the automated account is just as circular as the

[22] I owe this point to Charles Siewert, in conversation.

voluntary one: it requires that in order to come to know what you believe, you must already know (be in possession of information concerning) what you believe.

It might be countered that the "knowing" of the subpersonal system and the knowing of the believer are sufficiently different to render the account non-circular. The state the automaton is in is not a psychological state – it's *not* the very state the believer comes to be in when the routine is run. But whether or not the state is properly called psychological, it must carry information in some form about content and attitude if it is to play a role in the causation of the second-order belief. And this is sufficient to render the account circular.

Analogous non-transparency and circularity problems afflict Byrne's (Fullerton) extension of his model to self-knowledge of what one *desires*. Byrne suggests that self-application of the transparent rule DES explains how one comes to know one's preferences:

(DES) If ing would be desirable, believe that you want to

This rule is supposed to be transparent because the judgment about the desirability of ing is a judgment about *ing*, and not about oneself.[23] However, as in the case of BEL, in order to have reason to *apply* DES (and not simply to *be in a position to* apply it), you'd need to know that you were in a position to apply it. This is, again, what the puzzle of transparency shows us is required. DES is as bad as DES-3:

(DES-3) If ing would be desirable, believe that Fred wants to

[23] Of course if it were the case that something is desirable if and only if one desires it, DES itself (not just its application conditions) would be circular, since it would be equivalent to DES':

(DES') If you desire to , believe that you desire to

This construal of desirability is controversial, however. Suppose it therefore incorrect for the purposes of this discussion.

You'd need to know that *you believe that* ing would be desirable in order to know that you want to , just as you'd need to know that Fred believes that ing would be desirable to conclude that he wants to . And this is a fact, not about ing, but about you, and what you believe. So DES is not a transparent rule.

Moreover, given the intimate connection that, according to Byrne, obtains between judging that ing would be desirable and wanting to , the account suffers from circularity as much as the account based on BEL. Byrne maintains that the relation between judging that something would be desirable and wanting to do it is, though not *necessary* (like the relation between recognizing and believing), close enough to make DES "strongly reliably self-verifying" (powerpoint slide 21). However, if this is so, then the fact that one must know that one believes that ing would be desirable in order to be justified in applying DES is sufficient to render the account circular. You can't conclude that you want to unless you know that you believe that ing would be desirable; and believing that ing would be desirable is sufficiently close to wanting to that knowing that the former is true is *already* knowing that the latter is true.

The same sorts of problems would confront an effort to apply this approach to self-knowledge that one was merely *thinking* – as opposed to believing, desiring, etc. Suppose I come upon you, frowning and staring off into space, and ask: "What's up? Why so glum?" and you answer, "I was just thinking about inner-sense and Byrne-style extrospectionist accounts of self-knowledge of belief." ("Poor you," I think.) How is it that you knew what you were doing – that you were *thinking* about something? Should we say that you applied the transparent epistemic rule TNK?

(TNK) If *p*, believe that you are thinking that *p*

But there's no more reason to suppose that TNK can be transparently and non-circularly applied than either BEL or DES. In order to *be in a position to* apply TNK, you must be thinking (merely entertaining the proposition) that *p*; but in order to have *reason* to apply TNK (to do what its consequent tells you to do), you must know that you're in that position. Hence, you must know that you're thinking that *p* in order to conclude that you're thinking that *p*.

Comparison of BEL and TNK is telling. Given that their antecedents are *the same*, it's imperative that one know what *kind* of cognitive contact one has made with the content of the antecedent. Otherwise, one wouldn't know which rule to apply, and which consequent to detach.

Further, TNK lacks a feature that gave BEL (and DES) whatever initial plausibility they may have had. For if one takes BEL's antecedent to record the result of an act of looking outward to determine *the way (non-mental) things are*, there would seem to be a fairly direct route to the conclusion that what the antecedent records is something one *believes*. Your answer to the self-posed question "How is it with the world?" expresses how you take the world to be; and how you take the world to be is what you believe. BEL formalizes this connection between answers to world-directed questions and knowledge of what one believes; so if you believe BEL[24] you can apply it to yourself and come to know what you believe. What's doing the work here is the close connection between how one takes things to *be* – what one takes to be *true* – and what one *believes*. But it's hard to see what features of the extra-mental world, and, hence, of one's stance with respect to it, could be used to ground knowledge of what one is (merely) thinking. Mere thinking is a *neutral* (non-committal) propositional attitude – if it's even a propositional attitude at all. In either case, in merely thinking, there's no question of the truth or

[24] Belief in BEL is what Dretske (1994) calls a *connecting belief*.

falsity or desirability of what one thinks, no way in which one is taking the world. There's no *stance* to correlate with properties of extramental reality; there's no objective correlate (the way things *are*, the way they *ought to be*, the way they're *not*) of a propositional attitude (belief, desire, disbelief) for the extrospectionist to exploit here.

It's no surprise, then, to find Dretske (the *éminence grise* of contemporary extrospectionism) despairing of finding an explanation of privileged knowledge that one is thinking what one thinks. Dretske doesn't deny that we can have privileged knowledge both of *what* we're thinking and *that* we are thinking it. But he argues that while the former has a straightforward explanation, the latter does not:

> We have privileged and exclusive access to our own thoughts through acquaintance with their propositional content. ... [but] there's nothing we're aware of in thinking that indicates we're thinking. We're aware of what we think ... but thought content is evidentially worthless." [MS]

(See also Dretske 2006.) And he can't see where else one might look to find a basis for an explanation:

> If our mode of contact with our own thoughts doesn't give us a reason to think we're having these thoughts, what does? I haven't found anything – at least nothing one has privileged access to .... [Id.]

I would like to think that these remarks evince some sensitivity to the problems just discussed for extrospectionist theories of self-knowledge of thought. (There's nothing *out there* for the extrospectionist to appeal to.) Though Dretske's stated reasons for his position are different. Thinking, he says, is not like being hungry:

> When you are hungry there's (often enough, anyway) something you are aware of, something you feel, that indicates you're hungry. ... When you think, though, there's nothing you're aware of, nothing you feel, that indicates that you're thinking. [Id.]

I, of course, would beg to differ with this assessment. There's an introspectively detectable proprietary phenomenology of occurrent, conscious thinking: it's the phenomenology of occurrent, conscious *thoughts*. One knows what one is thinking, and that one is thinking it, in the same sort of way one knows what one is hearing, and that one is hearing it. Hearing has a distinctive phenomenology, such that when you're consciously experiencing it, you can know what you're hearing, and that you're hearing it. And the same is true (or so I would maintain) for thinking.[25]

Hence, the extrospectionist's failure to explain self-knowledge of thinking, is due, I would argue, to not allowing for a distinctively *cognitive* kind of phenomenology. But I think the immediate cause of Dretske's impasse is the lack of an objective correlate of (mere) thinking that would allow the extrospectionist to extend his model.

Byrne (Forthcoming) has proposed a different account of self-knowledge of thinking. He claims that one can come to know that one is thinking that *p* by trying to follow the transparent epistemic rule THINK-THAT:

(THINK-THAT) If the inner voice says that *p* and *p*, believe that you are thinking that *p*[26]

Byrne argues that since there's no such thing as an inner voice (to admit that there is would be to

---

[25] There are of course problems lurking here. (For example, how does one know that one is *experiencing* something at all?) I think they have solutions (see below); but given what he says about hunger, Dretske seems to be as saddled with them as PITT.

[26] THINK-THAT combines BEL with THINK:

(THINK) If the inner voice speaks about *x*, believe that you are thinking about *x*

It's not clear to me why the clause 'and *p*' is needed here, since *p* need not be true, nor need one believe it, in order that one simply think that *p*.

accept that there's "a shadowy inner world of ... auditory images" (20)), THINK-THAT can't in fact be successfully followed. However, one may still *believe that* there's an inner voice (a "harmless delusion"), and that it has said that *p*; and Byrne claims that thus *trying* to follow THINK-THAT is "very likely" to yield knowledge that one is thinking that *p*.

Now, Byrne rejects THINK*

(THINK*) If you inwardly speak about *x*, believe that you are thinking about *x*

and would, presumably, also reject THINK'

(THINK') If you inwardly say that *p*, believe that you are thinking that *p*

on the grounds that they're not transparent, since their antecedents mention mental states. But I don't see how changing 'if you inwardly say that *p*' to 'if the inner voice says that *p*' solves this problem. Presumably, it's one's *own* inner voice that's at issue; and, even if there really is no such thing, THINK-THAT requires that one *believe that* there is , and, hence, is still ostensibly about the rule follower's mental state (the imaginary state of speaking inwardly). So, THINK-THAT is not transparent.

Further, application of THINK-THAT requires both that one take the inner utterance to be meaningful, and that one know what the inner voice has said. This is built into THINK-THAT (if the inner voice says *that p*), but if it's the (pretend) *utterance* that one defers to in coming to know what one thinks, Byrne isn't really entitled to this. One can't take an utterance, whether inner or outer, to be a mere string of vocables if it's to provide evidence of *thinking*. And one must assign a particular meaning to an utterance if it's to provide evidence of the *content* of the thinking. In short, the inner utterance must be *understood*.[27] But understanding the inner

[27] What if the inner voice says "You can take that stock to the bank!"? Shall I conclude that I'm thinking that I can take a farm animal to a river, or that I can count on shares of ownership in a

utterance is just thinking what it means, and so already involves thinking that *p*. Thus, given that one must in general know that applicability conditions hold, the account of self-knowledge of thinking based on THINK-THAT is viciously circular.

A further problem for extrospectionist theories of self-knowledge of intentional states is how to account for knowledge of *content* itself. In order to successfully apply a transparent epistemic rule, one must know not only what one's *attitude* toward the antecedent is, but what its content is as well.[28] Knowledge of *both* is required in order to justify self-ascription of the attitude of the consequent. If I don't know that it's *p* that I recognize or think (or ing that I find desirable), I won't be justified in concluding that it's *p* that I believe or think (or ing that I want to do). Knowledge of content is built in to BEL, DES, and THINK-THAT. But one might just as well wonder how one knows what the contents of one's thoughts, beliefs and desires are. BEL and DES are focused on the attitudinal components of their respective state types, as evidenced by the fact that their application conditions presuppose knowledge of propositional content, and are stated in terms of what's true or desirable. They don't provide any explanation of how it is that one knows the contents of one's intentional states.

So, presumably, we would need some other sort of rule to apply in order to explain how

---

company paying dividends? (Cf. note 4.)

[28] In fact I'm skeptical of immediate introspective knowledge of propositional *attitudes* such as belief and desire, since I think a functionalist account of believing is correct (and functional properties are not introspectable), and since (for familiar Freudian reasons) one can be wrong about what one believes or desires. It seems to me more plausible that one can have immediate introspective knowledge of *acceptance* (or affirmation – *seeming* to believe) or *attraction* (*seeming* to desire), and that these sorts of "quasi-attitudes" are *prima facie*, but defeasible, evidence for what one's proper attitudes are, and that they have characteristic phenomenologies. (Note that in *PC* I argued that there's a phenomenology of conscious *thought*, not of belief.)

one knows that one is recognizing or thinking that *p* (and not that *q*), or that one wants to (and not to), and we would need to specify the conditions under which such a rule would be applicable. Using BEL and DES as models, a rule of the form CONT seems a likely proposal

(CONT) If χ, then believe that the content of your intentional state is that *p*[29]

where χ is the non-psychological object one looks to, and the conditions under which one is to apply CONT specify the property of this object that indicates that it's the content of one's intentional state. But what should the object be, and what's the relevant property? In the case of BEL, the object is (as the use of a propositional variable indicates) a *proposition*, and the property that indicates that one believes it is *truth*: if you ask yourself how things are with the world, and you judge that *p* is a *fact* (a true proposition), then it's reasonable to conclude that you believe that *p*. (In the case of DES the object is ing and the property is desirability.) So it seems clear that we ought to replace χ with *p*: the object of a thought (as of a belief) is a proposition. But what of the conditions for application of CONT? What property is such that judging the proposition that *p* to have it would give one reason to infer that the content of one's intentional state is that *p*?

Given that propositions *are* intentional contents, it would seem that one would have to judge that it *is* the proposition that *p*. That is, one would simply have to identify it as the proposition it is. But this assumes that propositions have properties that we are able to detect, and on the basis of which we can discriminate them (just as, on the extrospectionist view, knowing the *sensory* content of one's perceptual experience assumes that one can discriminate its objects in terms of their properties). So, according to the extrospectionist, what sort of access

[29]Assuming, *arguendo*, that desiring to is desiring that one .

could we have to propositions and their properties?

> Dretske (MS) maintains that it's *acquaintance*:
>
> ... our point of contact with our own thoughts is through their content: what it is we're thinking. We have a privileged and exclusive access to our own thoughts through acquaintance with their propositional content.

Hence, to think (merely entertain) a thought is simply to be acquainted with its content, and acquaintance with a content entails awareness of its distinctiveness.

Now, one might well wonder how one could be *acquainted with* (what is presumably) an abstract object (as well as how one could be *exclusively* acquainted with it, given that it's public). Dretske dismisses such qualms – too quickly, in my view. (He says that being acquainted with a proposition is simply knowing the content of one's thought. But this is unhelpful, given that knowledge of content is explained in terms of acquaintance with a proposition.) If there's to be a plausible extrospectionist account of knowledge of content along these lines, we're owed some sort of explanation of our access to propositions and their properties. In the case of perception, we have our sense organs and their causal transactions with the physical world: we know how we're experiencing (representing) things by inference from their perceived properties, which we can distinguish with our selectively sensitive sensory apparatuses (or so the extrospectionist story goes). Should we say then that we have a special faculty that allows us to peer into Platonic Heaven, and to pick out a particular abstract object, distinguishable to our gaze from its co-inhabitants? A naturalist would of course reject such a view; and naturalism is the *raison-d'être* of representationalism.[30] So I'm not sure what Dretske has in mind here, or what sort of story an

[30] Naturalists are not, *per se*, nominalists. The program of "naturalizing content" is inaptly named. It's the *expression relation* between contents and brain states that's the target of this program, where contents are almost universally taken to be mind- and language-independent abstract objects (propositions). (See, e.g., Fodor 1990, note 6.) Equally, it's the *acquaintance*

extrospectionist could tell about acquaintance with propositions.[31]

My own view is that acquaintance with the intentional contents of one's thoughts works in the same way as acquaintance with the sensational contents of one's sensory experiences. Experiences in general instantiate and are individuated by phenomenal properties. One experiences a phenomenal property when it's instantiated in one's conscious experience, and one is acquainted with the phenomenal properties one consciously experiences.

Acquaintance with thought content works in the same way. Intentional contents are cognitive phenomenal types. To think that *p* is to token a phenomenal type that *is* the content that *p*. (See Pitt 2009.) One is thus acquainted with intentional contents when they're tokened in one's experience. Such acquaintance counts as a kind of knowledge. One knows *what* one is thinking in being consciously acquainted with it. Such acquaintance-knowledge can also be the basis for a kind of knowledge-*that* – knowledge *by* acquaintance. One can know *that* one is thinking that *p* in virtue of recognizing the token of the individuative cognitive phenomenal type that *p*, applying the appropriate concepts (*thought*, *that p*, etc.), and forming a belief about it.[32] Cognitive phenomenology can also be used to explain knowledge of *thinking*. One knows *which*

---

relation between contents and brain states that must be naturalized. No spooky extra-sensory powers of discernment, whether of abstract objects or immaterial qualia, are to be sanctioned. (Still, you might wonder why someone who countenances abstract objects would balk at other sorts of immaterials.)

[31] Here's a story that I think won't work. I know that I'm thinking that the sky is blue because I see that the sky is blue. Propositions are property complexes, and property instances can be detected in perception. But, first, how, on this account, is the thought that the sky is blue to be distinguished from the perceptual state of seeing that it is blue? And, second, not all thought is about what's perceivable. What of my thought that justice is a virtue? (In short, Empiricist theories of thought are inadequate.)

[32] Concepts that one may have to *acquire*. (Hence the developmental phenomena Dretske (MS) discusses.)

sort of psychological state one is in by being acquainted with a proprietary sort of phenomenology, and *that* one is thinking on the basis of such acquaintance-knowledge.

I close this discussion of extrospectionism by noting that non-transparency and circularity problems of the sorts discussed above affect extrospectionist theories of self-knowledge of *perceptual* (experiential) content, as well. For notice that the puzzle of transparency arises as much for such theories as it does for extrospectionist theories of knowledge of thought, belief and desire. PER is, ostensibly, as bad as BEL:

(PER) If *x* is *F*, then your experience is representing *x* as *F*

for, of course, that *x* is *F* constitutes no reason at all to believe that anyone is experiencing anything. And, as in the case of BEL, the presumptive remedy would be to specify conditions under which application of PER would yield knowledge of perceptual content. The obvious suggestion is that it be the case that it (perceptually) *seems to you* that *x* is *F*. But then the same problem arises in the application of PER as does for BEL. In order not merely to *be* in the circumstances that license application of PER, but to *have reason to apply* PER, one would need to know that *x* seems to one to be *F*. And, again, knowing how something seems is knowing something, not about an external object, but about one's experience of it. So PER is not transparent. Moreover, knowing that x seems to one to be *F* is already knowing that one perceptually represents it to be *F* (or, at least, is as close as finding ing desirable and wanting to are); so applying PER requires that one already have the knowledge its application is supposed to deliver.

I conclude, therefore, that extrospectionist theories of self-knowledge pose no serious threat to PITT.

Which is not to say that PITT is without problems. In the remainder of this paper I discuss one in particular that strikes me as, *prima facie* anyway, potentially serious.[33]

V.

Here's the worry. You might think that any phenomenal property of an occurrent conscious state is, at least potentially (you might have to attend to it, really hard, for a long time, repeatedly), available to the individual who's in it. That is, you might think that there can be no *phenomenal* feature of a conscious experience which is *in principle* introspectively undetectable. Conscious phenomenology *is* the introspectable: it's the *surface* of our minds; the only point of direct contact we have with ourselves; the very *substance* of consciousness. And just as you might think that we ought to be able to perceive all the perceivable features of the surfaces of external objects (the only points of direct perceptual contact we have with *them*), you might think it equally tautologous that we ought to be able to introspect all the introspectable features of occurrent conscious states. We might not be omniscient with respect to our minds; but we ought to be (something close to) omniscient (in principle) with respect to our *conscious* minds. And if you thought all of this, you might wonder how it could be that the intentional content of a occurrent conscious thought could be identified with its phenomenal character, when in fact we're pretty *lousy* at identifying the contents of our thoughts – because in particular we're pretty lousy at identifying the contents of their constituent *concepts*.

I take myself to be consciously thinking, for example, that I know that Hell is other

---

[33] Thanks to congregants Charles Siewert and Terry Horgan, as well as the infidel Joe Levine, for very helpful discussion of the issues in sections II through IV. I'm especially grateful to Joe Levine for the care he has taken in understanding, and challenging. my view.

people.[34] Well, if the intentional content of the thought is cognitive-phenomenal, and it's compositional, then it's determined by the cognitive-phenomenal content of its constituent concepts. In particular, the content of the concept KNOWLEDGE is cognitive-phenomenal. So, if I can know introspectively that I'm thinking that I know that Hell is other people, I ought to be able to know introspectively what the content of the concept KNOWLEDGE is. And if that content is just a kind of conscious phenomenology, all on the inner surface, so to speak, I ought to be able to determine introspectively what it is. But I can't say what the content of the concept KNOWLEDGE is. Indeed, maybe *no one* can. Likewise for all the other concepts that confound our efforts at understanding: TRUTH, JUSTICE, RIGHTNESS, LOVE, etc.

In sum, if there are conceptual contents that are introspectively unknowable, then it can't be that the contents of concepts, and, hence, of thoughts, are phenomenally constituted: PITT must be wrong.

An easy response to the problem is to claim that it requires that concepts be *analyzable*, and that (Didn't we learn this way back in 1951?) there's no such thing as an analysis of a concept. The ostensibly problematic concepts are *primitive*. Indeed, *all* lexical concepts are semantically primitive. Their introspective recalcitrance isn't philosophically interesting at all; it' just what you'd expect. An analysis of thoughts into their conceptual constituents is all the analysis there is: knowing the content of your thought that you know that Hell is other people requires knowing the content of its constituents (viz.: I; KNOW; [THAT]; HELL; IS; OTHER; and PEOPLE); but knowing the content of each of its constituents doesn't require knowing the contents of *their* constituents, since they don't have any.

---

[34] Maybe it's just other philosophers.

I don't much like this response, myself. I'm rather partial to conceptual analyses, definitions, and their kin. (Primarily because I'm compelled by intuitions like *'the present king of France is male' is analytic*, *'female sister' is redundant*, *'false knowledge' is contradictory*, *'male sibling' and 'brother' are synonymous*, etc.; because I think explanations of these facts in terms of the analysis of concepts are the most satisfying; and because I'm not afraid of apparent counterexamples like 'black' and 'wolf'.) So I'd prefer another way of addressing the objection. Here are a few possibilities.

In fact, it's not clear that *all* the contents of conscious experience must be *equally* available to the experiencer. Introspective knowledge requires introspective *attention*, and the capacities for conscious experience and introspective attention to it are distinct. Perhaps it is, then, not implausible that the *resolving power* of introspective attention might be insufficient to discern all the fine-grained details of conscious experience. There might be features of our conscious experience that we simply can't make out introspectively – just as there might be details of a visually represented scene that we can't quite make out, no matter how closely we look. Our *experience* is fully determinate in its details, but *reporting* all those details requires a fineness of discernment – of *attention* – that we just don't have.

Or maybe the *scope* of introspective attention is limited – it's like a kind of inner *foveating*, where (at least) the finer details of what's beyond the center of attention aren't accessible, hence not reportable (as the unattended ticking of a clock, the song stuck in one's head, the road one's driving on). Perhaps there are sectors of the field of conscious experience that are simply out of the neurologically determined range of the "inner fovea" (just as we can't see what's behind us merely by turning our eyes). So then it could be that when one can't say

what the complete content of one's concept is, it's because one is n't – or perhaps *can't be* – focusing on the whole thing at once.

I don't find either of these suggestions very plausible. For one thing, it doesn't seem that inability to discern beyond a certain fineness of grain, even if we have such a limitation, is the problem in the case of concepts like KNOWLEDGE and JUSTICE. There seem, rather, to be substantive pieces missing. Supposing that knowledge is *F* justified true belief, it seems unlikely that *F* is some fine detail of the concept JUSTIFIED TRUE BELIEF that we simply can't make out. It seems like something more substantive is missing – a constituent of the same level of grain as JUSTIFIED, TRUE and BELIEF.

Yet, it's also hard to believe that the missing piece is missing from our understanding because, though it's of the same scale and grain as the others, and equally conscious, we simply can't "foveate" on it, and so can't say what it is. Why can't we redirect our attention to a constituent presently unattended to? For sure, you can't see what's behind you *merely* by moving your eyes: their *in situ* range is limited. But you can resituate them. You can turn your head. And if you also turn your body, you can access all 360 degrees of the visual scene (not to mention looking up and down, underneath and above, etc.). Can't we similarly direct our inner attention all over the place, as it were scouring our conscious inner landscape? It seems odd – that is to say, unmotivated – that there should be places we can't look that just happen to be where we find ourselves at a loss with respect to KNOWLEDGE, TRUTH, JUSTICE, and the rest.

So let's try something else. Maybe in cases of incomplete accessibility one is *consciously* entertaining only *part* of the concept the relevant term expresses in one's language, the rest being

unconscious, and hence introspectively inaccessible. This is different from the previous strategy, according to which the whole concept is conscious, but you just can't discern a part of it. In this case there's nothing to consciously introspectively attend to. Part of the concept is, as it were, "submerged."

But this doesn't seem very plausible either. *Why* should the submerged pieces be submerged? It's not like we're *repressing* them, because they're so dangerous or sexy that we feel guilty for having them. It might be difficult to analyze a conscious concept – just as it might be difficult to become aware of all of the rich detail of a conscious experience of, say, hearing a mass by Josquin or tasting an old Bordeaux. But that's not because these experiences have deeply buried unconscious features that you'd need a psychoanalyst's help to disinter. Discernment is typically something that must be *acquired*. But acquiring discernment with respect to one's conscious experiences is not a matter of making conscious what was unconscious. Maybe it's a matter of becoming more sensitive to what's already there. Or maybe it's a matter of having a new and different experience in response to the same stimulus. In either case, it doesn't seem likely that it's a matter of making conscious what's unconscious.

Here's the strategy I'm partial (but not entirely committed) to. Suppose that there is, in English, this word 'knowledge', and that it has a socially (i.e., non-individually) determined meaning, and that this meaning is *F justified true belief*. Well, in that case, if you don't know what *F* is, and, hence, you don't know what 'knowledge' means, then *you can't have any KNOWLEDGE thoughts* – because *you don't have the concept KNOWLEDGE*. You might *say* "I'm thinking that I know that I'm condemned to be free," and your *words* might *mean that* you're thinking that you know that you're condemned to be free. But if knowledge is *F* justified

true belief, and, long and hard as you try, you simply can't say what *F* is, then you're *not* thinking that you *know* that you're condemned to be free. That might be what your *words* mean, but it's not what's in your head. What you're consciously thinking is what you can report; so if what you can report is just that you truly, justifiedly believe that you're condemned to be free, then that's all you're consciously thinking. And if that's not sufficient for thinking that you *know* that you're condemned to be free, then your self-attribution is wrong.

In general, if the intentional content of a conscious thought is phenomenally constituted, and if introspection has equal access to phenomenal properties of all conscious states, then what you couldn't discover by introspection *is not (consciously) there*. And if it's not there, then you're not consciously thinking it. If it's the case that one can't think that one *knows* that *p* without having the concept *F*, and you can't discover the concept *F* in your conscious experience, then you're not consciously thinking that you know that p. You're consciously thinking something else. (And if you can *never* consciously entertain concept *F* together with the concepts JUSTIFIED, TRUE and BELIEF, then you can never consciously think that you know *anything*.)

It might be objected that Burge has shown us that we can possess concepts we don't completely grasp (ARTHRITIS, SOFA, and the like) – hence, *a fortiori*, that we can possess concepts that we don't completely consciously grasp. You can very well be thinking that you know that Hell is other people even if you can't say what knowledge is, provided you're a member of linguistic community in which the words you utter to express your thought – in particular, in this case, 'know' – have a determinate meaning. Conceptual content is socially determined, and what's social is by definition not in any one individual head; so you shouldn't

expect to automatically have *introspective* access to the contents of your thoughts. If you've *internalized* the socially determined content (i.e., made it completely explicit to yourself), then you can access it introspectively. But if you haven't, then you can't. So if you can't say what the content of the concept KNOWLEDGE is, it's not because you don't *have* it, and, hence, that you're not thinking that you know that you're condemned to freedom; it is because you haven't internalized (made explicit to yourself) its content.

In response, I maintain that you *can't* possess concepts you don't completely grasp. Burge is a very subtle and resourceful philosopher, and has argued for his anti-individualist thesis in subtly different ways on different occasions. I don't have the space here to examine all of them; but it does seem clear to me that the most influential presentation of the thesis, the one that almost everyone cites in defense of their allegiance to the basic tenets of externalism – the one in "Individualism and the Mental" (cited 1,205 times to date, according to Google Scholar) – is fatally flawed. In brief, the problem is this. The lynchpin of the anti-individualist argument in that paper is the thesis that belief attributions it's intuitively *natural* to make to individuals, based on their sincere avowals, ought to be construed as literally true, all things equal. Hence, if someone (call him Art) sincerely utters the sentence 'my arthritis has spread to my thigh', we who are in the know about where one can and can't (as a matter of *conceptual* necessity) have arthritis will nonetheless describe him as believing that his *arthritis* has spread to his thigh. We describe his conceptual error using *our* words, with the contents they have in *our* linguistic community; and since this is the overwhelmingly *natural* way to describe his mistake, we take it to be literally true. But if it's literally true, then Art has our concept ARTHRITIS (he can't have

ARTHRITIS beliefs without it), in spite of his misunderstanding of it.[35]

The intuitive power of Burge's thought experiment derives from the observation that it's – pretheoretically, innocently, commonsensically, overwhelmingly – *natural* to describe errors such as Art's using the very words the ascribee doesn't completely understand, without supposing that we're using them with anything but their ordinary meanings for us. I think there's little doubt that this is correct. But the principle that what's intuitively natural in this way should be taken to be *literally true* is not. There are too many cases where what's powerfully intuitively the correct way to describe what someone believes, in full knowledge of his conceptual confusion, and without intending that our words have nonstandard meanings, *can't* be construed as literally true, and where it's clearly *not* our intention that it be so construed. Here's just one.

Little Simone is taken to the zoo, having been told that, among other fun things, she'll get to ride a big escalator. (It's the Bronx Zoo. (I seem to remember there being a long escalator up from the subway).) After passing a number of cages containing large animals, Simone asks, excitedly: "Which one's the escalator? When do I get to ride it?" It seems entirely natural to describe Simone as having mistakenly believed that escalators are animals. This is a perfectly intuitive way to describe the mistake she made. "Ha, ha, ha!" we say, "Simone thinks elevators are animals! Isn't she *cute*?" (We routinely describe children's errors in is way: "When she was little, Lotta thought garter snakes were something you wear!"; "As a child, Linus thought

[35] Burge's (1979: 91) parenthetical comment that he's "not convinced" that someone who believes that 'orangutan' is a word for a fruit drink and says "An orangutan is a fruit drink" shouldn't therefore be taken to *mean that* an orangutan is a fruit drink and *think that* an orangutan is a fruit drink is, I believe, strong evidence that the argument for anti-individualism in this paper is based on the principle that the intuitively natural way to describe someone's mental state, on the basis of the words they use (even mistakenly), is literally true. (Burge has since come to reconsider the emphasis he places on belief *ascriptions* in this paper (see Burge 2007), and has changed his mind about the orangutan example (personal conversation).)

guerillas were monkeys!") But surely we don't think that Simone was thinking that *escalators* are *animals*: "Really?" the tone-deaf philosopher asks, "You're saying that she thought that mechanical stairways for the vertical transport of pedestrians are biological entities of the *Kingdom Metazoa* ?" "Well, of course not," we reply. "She obviously doesn't know what an escalator is. She wasn't thinking *that* at all!" Well, if she wasn't thinking *that*, then what exactly was her *conceptual* error? We don't mean to be attributing to her either so sophisticated, or *conceptually incoherent* beliefs. Even though on a *literal* interpretation of what we say, this *is* what we're doing. We use our words with their literal meanings, but we don't intend what we say about Simone to be *literally* true of her.

But isn't it equally obvious that Art doesn't know what arthritis is? And that, *therefore,* in describing his error we don't mean to be attributing to him the (by our lights) *contradictory* belief that a disease that can't occur in his thigh has spread to his thigh? So why isn't this a reason *not* take our description to be literally true of him? And this in spite of the fact that it's intuitively entirely natural to describe his error using the very word he doesn't understand? It remains intuitively natural to use the very words the confused have used *even when it's perfectly clear that we don't think that what we say is literally true of them*. So the fact that it's intuitively natural to describe Art as believing that his arthritis has spread to his thigh tells us nothing about whether or not he has *our* concept ARTHRITIS. We certainly don't *mean to say* that.[36]

If a concept is a thing with constituents, and you don't possess all of the constituents, then you don't possess the concept. So if KNOWLEDGE has more constituents than JUSTIFIED,

[36] I develop a critique of Burgean anti-individualism at much greater length in Pitt MS4 (eventually finding the root of all content externalism to be the false identification of thinking with inner speech).

TRUE and BELIEF, but those three are all the constituents you possess, then you don't possess the concept KNOWLEDGE, and you can't think that you *know* anything. And if you can't think that you know anything, then it's not surprising that you can't introspectively know that you think you know anything. The content is not there to be introspected. If you were really consciously thinking it (and it's not primitive) then you ought to be able to say what it is.[37]

I think that it might, in fact, not often be the case that we're *saying* exactly what we're *consciously thinking*, *if* what we say is a matter of what the words we utter mean in the language we speak. Lexical understanding is a patchy and idiosyncratic affair. Do you *really* know what *all* of the words you use mean? (What, for example, is the precise difference in meaning between, say, 'atrocious', 'execrable' and 'rebarbative'?) I think I can tell you what at least some of the words I use mean *to me* – which of my concepts *I* use them to express. I'm not so sure I could quote *Webster's Third* on any of them, however. (By 'atrocious', 'execrable' and 'rebarbative' I think I probably just mean *very bad*.)

Maybe most of us don't have cognitive lives nearly as differentiated as our vocabularies would suggest. And maybe we don't have nearly as much in common with each other, cognitively, as our shared language would suggest. We get by in communication and interaction provided the differences in the contents of our thoughts aren't sufficient to derail whatever

---

[37] It *is* pretty weird that we can know that we don't know what knowledge is. If we don't have an analysis of knowledge, then how do we know when we get it wrong? What are we comparing candidate analyses to when we see that they come up short? Are our intuitions responsive to a complete but inaccessible (because unconscious?) analysis? (It's a paradox of analysis. If you can tell when you've gotten an analysis right, then you must already know what the analysandum means; in which case the analysis is pointless. On the other hand, if you can tell when you *haven't* gotten an analysis right, then, again, you must already know what the analysandum means; in which case it's a mystery why you can't get the analysis right.)

projects we may have going, or want to begin, with others. But each of us is in his own little semantic bubble, blissfully unaware that we're by and large talking past each other.[38]

The way past the objection I've considered thus has far-reaching consequences, some of which may seem unacceptable. We often, perhaps typically, don't *think* what we *say*, and don't *say* what we *think*; nor do we often agree in what we think, even if we agree on what words we ought to use to express it. I don't find any of this very disturbing. In fact, I'm inclined to think that *linguistic meaning* – conceived of as a fixed, mind-independent assignment of unique semantic values to expression types of a language, which all speakers get hooked up to in the same way in virtue of shared intentions to obey certain conventions, or in virtue of happening to have grown up in a place where by and large people tend to make the same kinds of noises – is a will-o'-the-wisp. This is not to say that my words as I use them don't have meanings for me, derived from the thoughts and concepts I use them to express. (Nor is it to say that the concepts I express by my words don't have constituent structures that are accessible to me in conscious introspection, *or* that you and I *never* think or say the same thing.) Indeed, it's *because* my words as I use them have meanings that derive from the thoughts and concepts I use them to express that it's at best an idealization to speak of a unique, determinate, shared system of linguistic meaning for a community of speakers.

I think this is a consequence of taking seriously the idea (which I take very seriously) that linguistic intentionality is derived from cognitive intentionality. If cognitive intentionality is phenomenal, and phenomenal properties are not intersubjectively accessible, then knowledge of shared intentional content (and, hence, shared linguistic content), becomes (though not

[38] Maybe this wouldn't be too hard for a philosopher to accept.

impossible) problematic. Though what you *say* you think provides me with evidence of what you do think, the possibility that the way I construe the evidence – the content I assign to your utterance – is not the content of your thought is a live one.

Perhaps there's an innately determined, shared system of *cognitive* contents. (I'm eager to believe that there is, just as there's a innately determined, shared system of *sensory* contents.) But the *communication* of such content, which is necessary for the construction of objectively evaluable systems of linguistic meaning, is problematic.

So, on the internalist theory of meaning I'm defending here, language is a much bigger mess than one might have hoped. Maybe any ostensible "language" (English, Spanish, Tagalog) is too big a mess to succumb to one set of theoretical generalizations. Indeed, some linguists (including The Linguist, I believe) have said that there really is no such thing as English, Spanish or Tagalog (from either a syntactic or semantic point of view), but only collections of more or less similar dialects – or perhaps even idiolects. For my own reasons, I think this is probably right. We ought not to assume that an individual's apparent competence with a given set of syntactic-phonological types is an indication that he has internalized the same unique, determinate system of linguistic meaning as anyone else who is apparently competent with that same set of syntactic-phonological types.

Ite Missa Est.

REFERENCES

Burge, T. 1979. "Individualism and the Mental," *Midwest Studies in Philosophy, Vol. IV*, P.A. French, T.E. Uehling and H.K. Wettstein, eds., Minnesota: University of Minnesota Press.

---------- 2007. "Postscript to Individualism and the Mental," in *Foundations of Mind*, Oxford: Clarendon Press.

Byrne, A. 2001. "Intentionalism Defended," *Philosophical Review* 110: 49-90.

---------- 2005. "Introspection," *Philosophical Topics* 33: 79-104.

---------- Forthcoming. "Knowing that I am thinking," in A. Hatzimoysis, ed., *Self-Knowledge*, Oxford University Press.

---------- Fullerton. "Knowing What I Want," talk with powerpoint (distributed), 39th Fullerton Philosophy Symposium, Cal State Fullerton, April 30, 2009.

Chalmers, D. 2003. "The Content and Epistemology of Phenomenal Belief," in Q. Smith and A. Jokic, eds., *Consciousness: New Philosophical Perspectives*, Oxford: Clarendon Press.

---------- 2004. "Representational Character of Experience," in B. Leiter, ed., *The Future for Philosophy*, Oxford: Oxford University Press.

Dretske, F. 1994. "Introspection," *Proceedings of the Aristotelian Society* 94: 263-278.

---------- 1995. *Naturalizing the Mind*, Cambridge, MA: The MIT Press.

---------- 2006. "Representation, Teleosemantics, and the Problem of Self-Knowledge," in G. Macdonald and D. Papineau, eds., *Teleosemantics*, Oxford: Clarendon Press.

---------- MS. "I Think I Think, Therefore, I am – I think," text of talk delivered at the 39th Fullerton Philosophy Symposium, Cal State Fullerton, April 30, 2009 (subsequently revised and renamed "Awareness and Authority: The Structure of Self-Knowledge").

Evans, G. 1982. *The Varieties of Reference*, Oxford: Oxford University Press.

Fodor, J.A. 1990. "A Theory of Content, II," in *A Theory of Content and Other Essays*, Cambridge, MA: The MIT Press.

Gertler, B. 2003/8. "Self Knowledge," *Stanford Internet Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/self-knowledge/.

Levine, J. MS. "What Is Cognitive Phenomenology, and Do We Have It?", presented at APA symposium on cognitive phenomenology, December 2010.

Lormand, E. 1996. "Nonphenomenal Consciousness," *Noûs* 30: 242-261.

Lycan, W. 1996. *Consciousness and Experience*, Cambridge, MA: The MIT Press.

Martin, M. 1998. "An Eye Directed Outward," in C. Wright, B. Smith and C. Macdonald, eds., *Knowing our Own Minds*, Oxford: Oxford University Press.

Moran, R. 2001. *Authority and Estrangement*, Princeton: Princeton University Press.

Nida-Rümelin, M. 1995. "What Mary Couldn't Know: Belief About Phenomenal States," in T. Metzinger, ed., *Conscious Experience*, Exeter: Imprint Academic.

Pitt, D. 2004. "The Phenomenology of Cognition, Or, *What Is It Like to Think That P*?", *Philosophy and Phenomenological Research* 69: 1-36.

-------- 2009. "Intentional Psychologism," *Philosophical Studies* 146: 117-138.

-------- MS1. "Unconscious Intentionality," http://www.calstatela.edu/faculty/dpitt/Unconscious Intentionality.pdf.

-------- MS2. "Demonstrative Thoughts" (to appear in U. Kriegel and T. Horgan, eds., *Phenomenal Intentionality*, Oxford University Press, forthcoming).

-------- MS3. "The Paraphenomenal Hypothesis," http://www.calstatela.edu/faculty/dpitt/PH.pdf.

-------- MS4. "The Burgean Intuitions," http://www.calstatela.edu/faculty/dpitt/ The%20Burgean%20Intuitions.pdf.

Tye, M. 2000. *Consciousness, Color, and Content*, Cambridge, MA: The MIT Press.

--------- 2009. *Consciousness Revisited*, Cambridge, MA: The MIT Press.