

Short version, read at
Association for the Scientific Study of Consciousness 11
Las Vegas, June 24 1007

THE INTROSPECTIVE AVAILABILITY OF INTENTIONAL CONTENT

David Pitt
California State University, Los Angeles

I believe that there is a phenomenology – a “what it’s like” – of conscious occurrent thought. I believe this because I believe that *any* conscious state necessarily has phenomenal properties: any conscious state is conscious in some *way* or other, and these *ways* of being conscious are just phenomenal properties. I also believe that the phenomenology of a conscious occurrent thought is *individuating*: that is, in virtue of its having the phenomenal properties it has, it is a *thought* (as opposed to some other kind of mental state) with a specific *intentional content*. I believe this because I believe that we are able to identify our occurrent conscious thoughts as the thoughts they are, 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 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 mental 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 – not because they *represent* something – but because of their intrinsic nature. And I believe that the individuating phenomenology of conscious occurrent thought is *proprietary*: it is 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.

Amen.

I'll call the view that there is a proprietary, individuating phenomenology of occurrent conscious thought the *phenomenal intentionality thesis*. Many people find this thesis pretty hard to swallow. Well, I have tenure now; so they can all go to hell. No, wait! I mean, I'm a philosopher; so it's my job to try to convince you that the crazy things I say are true, that the crazy things you say are not, and that you would be crazy not to exchange your crazy views for mine. That's why I just summarized my arguments for the phenomenal intentionality thesis. And they're pretty convincing, if you ask me. The thesis really isn't crazy at all. You should believe it. Of course I still have problems – or what I would prefer to think of as *prima facie* problems that turn out on reflection either not to be problems at all, or to be problems that are not exclusively mine. This paper is about one such *prima facie* problem. I hope to provide at least one good reason for thinking that it might be *merely* *prima facie*.

Here's the worry. You might think that all of the phenomenal properties of a conscious occurrent state are, at least potentially (you might have to *attend* to it, really *hard*, for a *long* time, *over and over*), available to introspection. That is, you might think that there can be no *phenomenal* feature of a conscious experience which is *in principle* hidden from inner scrutiny. Conscious phenomenology *is* the introspectable: it is 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 (and report) 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 (and report) all the

introspectable features of occurrent conscious states. We may not be omniscient with respect to our minds; but, you might think, surely we *are* 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 conscious occurrent thought could be identified with its phenomenal character, when in fact we are pretty *lousy* at identifying the contents of our thoughts – because in particular we are 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 people (or maybe just other philosophers). Well, if the intentional content of the thought is cognitive-phenomenological, and it is compositional, then it is determined by the cognitive-phenomenal content of its constituent concepts. In particular, the content of the concept KNOWLEDGE is cognitive-phenomenological. 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. Am I not Lord of all I can inwardly, consciously, occurrently survey? But I cannot say what the content of the concept KNOWLEDGE is. Indeed, it doesn't look like *anyone* can. Likewise for all of the other concepts that confound our efforts at understanding: TRUTH, LOVE, JUSTICE, RIGHTNESS, ..., REPUBLICANISM

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: the phenomenal intentionality thesis must be false. That's my problem.

An easy response is to claim that it requires that concepts be *analyzable*, and that

(Didn't we all learn this 50 years ago?) there is no such thing as an analysis of a concept. The ostensibly problematic concepts are *primitive*. Indeed, *all* lexically simple concepts (i.e., concepts expressible by lexically simple expressions) are semantically primitive. Their introspective recalcitrance isn't philosophically interesting at all; it's just what one would 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 does not require knowing the contents of *their* constituents, since they don't have any. What is the content of the concept KNOWLEDGE? Just that: *knowledge*. There's nothing more to say.

I don't much like this response, myself. I'm a bad boy and didn't learn my Quinean lessons in graduate school. I rather like conceptual analyses, definitions, and their kin. (Primarily because I am 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 decompositional explanations of these facts are the most satisfying; and because I'm not afraid of apparent counterexamples like 'red' and 'wolf'.) So I'd prefer another way of addressing the objection. Here are a few possibilities.

It's not clear that *all* of the contents of consciousness must be *equally* available to introspection. 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 of the fine-grained details of conscious experience. There might be details 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 of 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 are not accessible, hence not reportable (as the unattended ticking of a clock, a song stuck in one's head, or the road one is driving on). Perhaps there are sectors of the field of our 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 isn't – or perhaps *can't be* – focusing on the whole thing at once.

I don't either of these suggestions terribly 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. If 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 justification, truth and belief.

Yet it also is hard to believe that the missing piece is missing because, though it is 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 presently unattended constituent?

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 we not 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, RIGHTNESS 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 "look at" a part of it. In this case there is 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 missing pieces be missing? It's not like we're *repressing* them, because they're so dangerous and 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's not a matter of making conscious what's unconscious.

Here's the strategy I'm partial 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. (Nobody knows what *X* is. Everybody knows that their shared word means something more than *justified true belief*, but nobody can say what that something else is. This is a very curious situation. But let's move on.) 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 Hell is other people," and your *words* might *mean that* you're thinking that you know that Hell is other people. 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 Hell is other people. 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 Hell is other people, then that's all you're consciously thinking. And if that's not sufficient for thinking that you *know* that Hell is other people, 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 cannot discover by introspection *is not there*. And if it's not there, then you're not consciously thinking it. If you believe that you can't think that you *know* that Hell is other people without having the concept *F*, and you cannot discover concept *F* in your conscious

experience, then you're not consciously thinking that you know that Hell is other people. 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*.)

You may be wondering: Have I not learned my *Burgeon* lessons either? Don't I know that one can possess concepts one doesn't completely grasp (ARTHRITIS, SOFA, and the like) – hence, a fortiori, that one can possess concepts that one doesn'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 are 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 Hell is other people; it's because you haven't internalized (made explicit to yourself) its content. (If *nobody* in the community knows the content of the concept KNOWLEDGE (everyone tries to, but everyone recognizes they don't know what *F* is), then perhaps nobody, including you, has it. That would also explain why you can't read its content off your conscious cognitive experience. But there are plenty of other concepts – ARTHRITIS and SOFA, for example – that could be used to make the same point, in a way that's relevant to the main concern of this paper.)

Well, since you asked: No, I haven't learned my Burgean lessons; and, no, 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 sites in defense of their allegiance to the basic tenets of externalism, the one in "Individualism and the Mental," 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 is 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 Bert) sincerely utters the sentence "My arthritis has spread to my thigh," we who are in the know about where one can and cannot (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 Bert has our concept ARTHRITIS (he can't have ARTHRITIS beliefs without it) in spite of his misunderstanding of it.

The intuitive power of Burge's thought experiment derives from the incontestable fact that it *is* – pretheoretically, innocently, commonsensically, overwhelmingly – *natural* to describe errors such as Bert's using the very words the ascriber does not completely understand, *without supposing that we are using them with anything but their ordinary meanings for us*. I think there is little doubt that this is correct. But the principle that what is intuitively natural in this way is *literally true* is not correct. There are too many cases where what is 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, *cannot* be construed as literally true, and where it is clearly *not* our intention that it be so construed. Here is just one.

Little Francine is taken to the zoo, having been told that, among other fun things, she will get to ride a big escalator. (It's the Bronx Zoo.) After passing a number of cages containing large animals, Francine asks, excitedly: "Which one is the escalator? When do I get to ride it?" It seems entirely natural to describe Francine 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, "Francine thinks elevators are animals! Ha, ha, ha! Isn't she *cute*?" (We routinely describe children's errors in is way: "When she was little, Lotta thought garter snakes were an article of clothing"; "As a child, Linus thought guerillas were apes"; "When she was a kid, Lucy thought dental bridges were built over dental rivers"; etc. (Okay, so now I've given *four* examples.)) But surely we don't think that Francine 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 such 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 Francie to be *literally* true of her.

But isn't it equally obvious that Bert doesn't know what arthritis is? And that, *therefore*, in describing his error we do not mean to be attributing to him the (by our lights) *contradictory*

belief that a disease that cannot occur in his thigh has spread to his thigh? So why isn't this a reason *not* take our description of him to be literally true of him? And this in spite of the fact that it is intuitively entirely natural to describe his error using the words 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 Bert 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.

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, TRUE and BELIEF, but those 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. *Say* what you will – make all the false claims to conceptual content you like. Saying it will not make it so. If you're really consciously thinking it (and it's not primitive) then you ought to be able to say what it is.

I think that it might, in fact, not often be the case that we are *saying* exactly what we are *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, *exactly*, do 'gamble', 'gaming', 'casino', 'sin', 'magic' and 'fabulous' (not to mention 'truth', 'justice' and 'Republicanism') mean *in English*? I think I can tell you what at least some of them mean *to me* – what concepts *I* use them to express. I'm not so

sure I could quote *Webster's* on any of them, however. And I'm not sure there's a difference, in my idiolect, between, say, 'atrocious', 'execrable' and 'rebarbative' – I think I probably just mean *very bad* by all of them.

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 are not sufficient to derail whatever 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 are by and large talking past each other. (Maybe this wouldn't be too hard for a *philosopher* to accept.)

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 *mean* (think) what we *say*, and don't *say* what we *mean* (think); nor do we often agree in what we think, even if we agree on what words we ought to use to express it. None of this bothers me much. 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 do not 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 do not 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 is *because* my

words as I use them have meanings that derive from the thoughts and concepts I use them to express that it is at best an idealization to speak of a unique, determinate, shared system of linguistic meaning.

I think this is a consequence of taking seriously the idea (which I do indeed take very seriously) that linguistic intentionality is derived from mental 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 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. There's no way to know with certainty that my interpretation is correct.

Perhaps there is an innately determined, shared system of *mental* content. (I'm eager to believe that there is.) But the *communication* of such content, which is necessary for the construction of objectively evaluable systems of *linguistic* meaning, is, as I said, 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) think 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.