Self-Unconsciousness

Eric Schwitzgebel
Department of Philosophy
University of California at Riverside
Riverside, CA 92521-0201

eschwitz at domain- ucr.edu

July 28, 2010

Self-Unconsciousness

i.

Philosophers tend to be pretty impressed by human self-knowledge. Descartes (1641/1984) thought our knowledge of our own stream of experience was the secure and indubitable foundation upon which to build our knowledge of the rest of the world. Hume – who was capable of being skeptical about almost anything – said that the only existences we can be certain of are our own sensory and imagistic experiences (1739/1978, p. 212). Perhaps the most prominent writer on self-knowledge in contemporary philosophy is Sydney Shoemaker. The central aim of much of his work has been to show that certain sorts of error are impossible (1963, 1988, 1994). David Chalmers has likewise attempted to show that, for a suitably constrained class of beliefs about one's own consciousness, error is impossible (2003, sec. 4.1). Even philosophers most of the community thinks of as pessimistic about self-knowledge of consciousness seem to me, really, to be fairly optimistic. Paul Churchland, famous for his disdain of ordinary people's knowledge about the mind, compares the accuracy of introspection to the accuracy of sense perception – pretty good, presumably, about medium-sized, nearby matters (1985, 1988). Dan Dennett, often cited as a pessimist about introspective report, actually says that we can come close to infallibility when charitably interpreted (2002).

The above references concern knowledge of the stream of conscious experience, but philosophers have also tended to be impressed with our self-knowledge of our attitudes, such as our beliefs and desires. Consider this: Although I can be wrong about its being sunny outside, I can't in the same way be wrong, it seems, about the fact that I *think* it's sunny outside. Some

philosophers have argued that this accuracy is due to the operation of a fairly simple and straightforward self-detection mechanism that takes our attitudes as inputs and produces beliefs about those attitudes as outputs, a mechanism so simple that it rarely errs (e.g., Nichols and Stich 2003; Goldman 2006). Others have argued that our attitudes, at least some of them, can contain each other in a self-fulfilling way, so that my thought or belief that I think that it's sunny in some sense literally contains as a part the thought or belief that it's sunny. (That characterization is a bit simplified, but Burge [1988, 1996], Heil [1988], Dretske [1995], and Shoemaker [1995] have said things along roughly these lines.)

From Descartes to the present, the philosophical literature on self-knowledge of consciousness and attitudes has focused, with a few exceptions, on statements of or attempted explanations of the fact that we know ourselves remarkably well. Even those philosophers portrayed as skeptical have mostly been exercised to concoct bizarre or pathological scenarios designed to show that although our self-knowledge about our attitudes or current conscious experience may be excellent it's not *wholly* infallible (e.g., Armstrong 1963; Churchland 1988). The debate, that is, has been between the infallibilists and the not-quite-infallibilists. I, however, am inclined to think we don't know our stream of consciousness or our own attitudes very well at all.

ii.

Let's consider currently ongoing conscious experience first. (The next few paragraphs are adapted, I should say, from more detailed presentations in Schwitzgebel 2002a, 2008, forthcoming-b.) Suppose you're looking directly at a sizeable red object in good light and

normal conditions. You judge that you're having the visual experience of red. How could you possibly be wrong about that? Or suppose someone has just dropped a 60 pound barbell on your toe. You judge that you're feeling pain. How could you possibly be wrong about that either?

Well, in such cases I'm inclined to think it *is* highly unlikely that one would go wrong. But the question is this: How *representative* are such cases? Does the apparent difficulty of going wrong about color and pain experiences in canonical conditions reflect the *general* security of our judgments about our ongoing stream of conscious experience, or are those cases exceptional, best cases? Optimists about our self-knowledge of our conscious experience tend to focus, with stunning frequency, on *exactly* the cases of seeing red and feeling pain, generalizing from there, thus implicitly treating those cases as typical.

Let's start somewhere else for a change. Close your eyes and form a visual image. (Go ahead and do it now if you want.) Imagine, for example, the front of your house as viewed from the street. Assuming that you can in fact form such imagery, consider this: How well do you know, right now, that imagery experience? You know, I assume, *that* you have an image, and you know some aspects of its content – that it's your house, say, from a particular point of view. But that's not really to say very much yet about your imagery experience. Consider these further questions:

How much of the scene can you vividly visualize at once? Can you keep the image of the chimney vividly in mind at the same time you vividly imagine your front door? Or does the image of the chimney fade as you start to think about the door? How much detail does your image have? How stable is it? Supposing you can't visually imagine the entire front of your house in rich detail all at once, what happens to the aspects of the image that are relatively less detailed? If the chimney is still experienced as part of your imagery when your image-making

energies are focused on the front door, how exactly is it experienced? Does it have determinate shape, determinate color? In general, do the objects in your image have color before you think to assign color to them, or do some of the colors remain indeterminate, at least for a while? If there is indeterminacy of color, how is that indeterminacy experienced? As gray? Does your visual image have depth in the same way your sensory visual experience does, or is your imagery somehow flatter, more sketch-like or picture-like? Is it located in subjective space? Does is seem in some way as though the image is in your head, or in front of your forehead, or before your eyes? Or does it seem wrong to say that the image is experienced as though *located* anywhere at all? How much is your visual imagery like the experience of seeing a picture, or having phosphenes, or afterimages, or dreams, or daydreams?

Now these are pretty substantial questions about your imagery experience. They aren't piddling details, but questions about major-to-middle-sized features of the visual imagery that is presumably currently ongoing in you right now. If I asked you questions at that level of detail about an ordinary external object near to hand, you'd have no trouble at all – about a book, say. How stable is it? Does it flash in and out of existence? Does its cover have a durable color? What happens to its shape (its real shape, not its "apparent shape") when you open it up, spin it around, look at the underside? These questions present no difficulty. And yet most of the people I've talked to find that questions at this level of detail about their conscious experience of imagery are considerably more difficult. In fact, I think people often simply *get it wrong* when they think about their imagery experience. One kind of evidence for this is the failure of psychologists, in more than 100 years of research, to find any real relationship between people's self-reports about their imagery and their performance on cognitive tasks that would presumably be facilitated by imagery. For example, some people say they have imagery as vivid and detailed

as ordinary vision, or even more so. Others claim to have no imagery at all. And yet there is no consistently detectable performance difference between self-described high- and low-imagery people on psychological tests like mental rotation tasks, or mental unfolding tasks (these are tasks where you're asked to guess what something would look like when unfolded), or tests of visual memory, or tests of visual creativity (Schwitzgebel 2002, forthcoming-b).

Now of course people might still be quite accurate in their judgments about their visual imagery, even if the differences in their judgments don't correspond to any sort of performance differences in behavioral tests. Maybe phenomenological differences are irrelevant to behavioral performance, or at least performance on the sorts of tasks that psychologists have so far concocted. But if you share my intuitive sense that it feels somehow difficult to introspect your imagery – if you share my insecurity about your self-knowledge of your own ongoing conscious experience of sustaining a visual image – then maybe you'll grant me this: There's no special, remarkable perfection in our knowledge of such things, no elite epistemic status. We probably know normal, outward objects better, in fact.

How about emotional experience? Reflect on your own ongoing emotional experience at this moment. Do you even have any? (If not, try to generate some.) Now let me ask: Is it completely obvious to you what the character of that experience is? Does introspection reveal it to you as clearly as visual observation reveals the presence of the text before your eyes? Can you discern the gross and fine features of your emotions as easily and confidently as you can discern the gross and fine features of the desk at which you are sitting? Can you trace its spatiality (or nonspatiality), its viscerality or cognitiveness, its involvement with conscious imagery, thought, proprioception, or whatever, as sharply and infallibly as you can discern the shape, texture, color, and relative position of your desktop? I can't, of course, force a particular

answer to these questions. I can only invite you to share my intuitive sense of uncertainty. And it doesn't seem to me that the problem here is merely linguistic, merely a matter of finding the right *words* to describe an experience known in precise detail.

How about visual sensory experience? Consider not your visual experience when you're looking directly at a canonical color but rather your visual experience of the region 10 degrees or 30 degrees *away* from the center point. How clear is it? How finely delineated is your experience of the shape and color of things that you're *not* looking directly at? People give very different answers to this question – some say they experience distinct shape and color only in a very narrow, rapidly moving foveal area, about 1-2 degrees of arc (about the size of your thumbnail held at arm's length); others claim to experience shape and color with high precision in 30 or 50 or 100 degrees of visual arc; still others find *shape* imprecise outside a narrow central area, but find *color* quite distinct even 20 or 30 degrees out. And what's more, people's opinions about this aren't stable over time. In the course of a conversation, they'll shift from thinking one thing to thinking another. They change their minds. The phenomenal character of their visual experience is not securely known.

We don't really know so much, then, I think, about our stream of consciousness experience, about the phenomenology always transpiring within us. We know *certain* things. Our knowledge tends to be good about just one aspect of our stream of experience, in fact, I think: what I'll call the *gist* of the experience.

Now I'm not sure how to characterize what I'm calling "gist" here, if it's even a well-formed concept at all. Here's my best stab: It's the main topic and comment that an experience contributes to the stream of conscious thought. Compare the gist of a photograph or the gist of a lecture. The photograph, say, is of a man and a woman dining together at a fancy restaurant –

that's its gist (though maybe no short string of words could capture the gist of a photograph exactly). Gist leaves out two sorts of things. It leaves out details, such as exactly what's on the table, the color of the tablecloth, the shape of the wineglasses (unless those things are somehow central to the photograph or the viewer's interpretation of it). And it also leaves out major structural features that are not part of the topic and comment, such as whether the photograph is black and white or colored, clear or fuzzy, from a near perspective or farther away. Similarly the gist of a lecture is its main points, leaving out both detailed subpoints and major structural features such as the tone and gender of the voice and the presence or absence of PowerPoint slides.

Here are some examples of what I take to be the gist of various experiences, that is, the main topic and comment that the experiences contribute to the stream of thought. For the experience of hunger, the gist is simply that one is hungry (and perhaps how intensely); it's not, for example, *where* in the body the hunger is felt or *how* it is felt, even though these are major structural features of the hunger experience. For imagery and inner speech, gist is similar to gist in photographs and lectures: the main, summary content of the image or thought, and not, for example, how richly detailed the imagery is or whether the thought transpires in imagined words. (If at the center of your thinking is the impression that you are having richly detailed imagery, the gist would be *that* you have having richly detailed imagery – the content of your introspective judgment about your imagery – which may or may not co-occur with an actual wealth of detail in your imagery.) For an emotional experience, the gist is the main target and rough label of the emotion, for example that I'm feeling frustrated by slowness of the traffic I'm stuck in.

I submit that the main thing, maybe the only thing, that people dependably know across the board about their stream of experience is its gist. Intuitively – when I introspect with appropriate epistemic caution – I feel that I am, and others are, on much more secure footing when reflecting on the gist of experiences than when reflecting on things other than gist – including major non-gist aspects of experience, such as structural features of the experience that don't contribute a topic and comment to the flow of conscious thought. The empirical evidence I have about the inaccuracy of reports about the stream of experience touches many such structural aspects of our reports – the degree of detail in peripheral vision, the coloration or not of dreams, the vividness of one's imagery, the echoic qualities of auditory experience, the bodily aspects of emotion, the doubling or not of objects farther or nearer than the visual focal point, the richness of experience outside of attention (Schwitzgebel forthcoming-b) – but my skeptical evidence never seems to touch gist. Perhaps that's because self-knowledge of gist is crucial to self-regulation of the stream of thought and thus has a kind of functional value that knowledge of structural features of consciousness does not have.

Let me emphasize that the main thrust here is meant to be skeptical: We have poor self-knowledge of everything *but* gist. And gist is so basic, so obvious, that to know that is not yet to know very much: If all I securely know about my current visual imagery is that it's my house as viewed from the front, really I know very little about my imagery. Knowledge of gist, assuming that we do indeed know it, is just knowledge of the very most basic stuff that ought to just hit one over the head unless the most utterly radical skepticism about self-knowledge is true – a mere island of obviousness, that is, in what is mostly a sea of ignorance about our stream of experience.

Furthermore, the island falls quickly undersea: You probably know the gist of your current experience, and maybe of the last few seconds of experience, and of some selected and probably unrepresentative experiences from your more distant past. But what generally occupies your thoughts – what you *tend* to have near the center of your experience – about that I doubt you have much knowledge at all. I find Russ Hurlburt's work convincing on this point: A person might very frequently have angry thoughts about his children, as he reports when sampled at a random moments, and yet he might sincerely deny that it's so in the general case (Hurlburt and Heavey 2006, p. 6-7); commonly, people think that a random sampling of their mental lives will reveal lots of abstract or intellectual thought, or lots of thoughts about sex, and yet find upon actual sampling that they report virtually no such thoughts (e.g., Kane, forthcoming; Hurlburt and Heavey 2006, p. 141; nor do people appear to be very good, by Hurlburt's measures, in their generalizations about structural features of their experience, such as whether they experience lots of inner speech or lots of visual imagery: Kane, forthcoming; Hurlburt and Heavey 2006; Hurlburt and Schwitzgebel 2007; Hurlburt and Schwitzgebel forthcoming). You may know the gist of your current experience, but try to extend your knowledge back more than a few seconds, try to generalize, or try to discern even moderately large structural features of your experience, and soon you will err.

ii.

How about our self-knowledge of our attitudes? For some of our attitudes I'm inclined toward a version of what sometimes gets called a "transparency" view. The rough idea here is that if someone asks me something like "do you believe it will rain tomorrow?" I think about whether it

will rain. That is, despite the fact that the question is about what I *believe*, in answering it I don't think about what I believe, I think about external affairs – and then I express my judgment about those external affairs using, if it suits me, either self-attributive language ("Of course I don't think it will rain!") or objective language ("Of course it won't rain"), with the difference between these two sorts of expression grounded more in conversational pragmatics than in the presence or absence of an introspective act. This expressive procedure delivers accurate self-attributions if there's the right kind of hook-up between my judgment ("it won't rain") and my self-attributive expression of that judgment ("I don't think it will rain"). (For transparency views of self-knowledge of attitudes, see, e.g., Evans 1982; Moran 2001; Byrne 2005; Gordon 2007; against transparency see Gertler forthcoming.)

This sort of procedure works fine, I think, for fairly trivial attitudes or attitudes that connect fairly narrowly to our actions – attitudes like my preference for vanilla ice cream over chocolate when I'm asked on a particular occasion or my general belief that it doesn't rain much in California in April. The vanilla preference and the rain belief don't tangle much with my broad values or self-conception, and their connections to my behavior are pretty limited – an evening's ice cream consumption, my springtime habits in picnic planning and umbrella carrying.

But those aren't the attitudes I care about most – or at least they're not the ones most critical to my self-knowledge in the morally-loaded sense of "self-knowledge", in the sense of the Delphic oracle's recommendation to "know thyself". The oracle was presumably not concerned about whether people knew their attitudes toward the April weather. To the extent the injunction to know oneself pertains to self-knowledge of attitudes, it must be attitudes like your

central values and your general background assumptions about the world and about other people. And about such matters, I believe (I think I believe!) our self-knowledge is rather poor.

Let's start with sexism. (This example is adapted from Schwitzgebel forthcoming-a.) Many men in academia sincerely profess that men and women are equally intelligent. Ralph, a philosophy professor let's suppose, is one such man. He is prepared to argue coherently, authentically, and vehemently for equality of intelligence and has argued the point repeatedly in the past. And yet Ralph is systematically sexist in his spontaneous reactions, judgments, and unguarded behavior. When he gazes out on class the first day of each term, he can't help but think that some students look brighter than others – and to him, the women rarely look bright. When a woman makes an insightful comment or submits an excellent essay, he feels more surprise than he would were man to do so, even though his female students make insightful comments and submit excellent essays at the same rate as the men. When Ralph is on the hiring committee for a new office manager, it won't seem to him that the women are the most intellectually capable, even if they are; or if he does become convinced of the intelligence of a female applicant, it will have taken more evidence than if the applicant had been male. And so on. Ralph may know this about himself, or he may not. I don't see any reason to think that Ralph would have any special authority in such matters, compared to other people who have observed substantial portions of his relevant behavior. In fact, he may be disadvantaged by a desire not to see himself as sexist and by the more general desire to see himself as someone whose actions reflect his espoused principles. (For empirically informed discussion of professional philosophers' apparent ignorance of their own sexism, see Haslanger [2008] and Saul [forthcoming].)

Now you might want to say that in a case like Ralph's – and let's assume that Ralph is *not* aware of the pervasive sexism in his behavior – there's no lack of authority about what one *believes*. Ralph believes that men and women are equally intelligent, you might suggest, he just doesn't tend to *act* on that belief. But this seems to me an overly linguistic and intellectualist view of belief. Our beliefs guide us not just in what we say, but in what we do – they animate our limbs, not just our mouths – and they're also manifested in our spontaneous emotional reactions and our implicit assumptions. Now I don't think it's quite right to call Ralph simply an out-and-out sexist who straightforwardly believes that women are intellectually inferior. What Ralph says and how he reasons in his most abstract and most thoughtful moments is an important part of what he does, even it it's only a part. Ralph's attitude toward the intellectual equality of the sexes is what I'd call an in-between state. His dispositions, his patterns of response, his habits of thought, are mixed up and inconsistent. It's neither quite right to say that he believes in the intellectual equality of the sexes nor quite right to say that he fails to believe that. But he has no specially privileged self-knowledge of that fact.

Many people profess to believe in God and Heaven. Here again, I think we have a case where sincere linguistic avowal often comes apart from behavioral manifestation and spontaneous response. To believe in God, in the mainstream monotheistic sense, is in part to believe that there's an omniscient agent who is always observing you, with the power to reward you with eternal bliss or condemn you to eternal torment. Many people who sincerely verbally espouse the existence of such a God fail to act and react in their daily lives as though such a God exists: They'll do before God what they wouldn't do before any neighbor, even the most forgiving one; and only *human* eyes and *human* condemnation will give them the pinch of fear and remorse. Such people are, I think, like Ralph the sexist. But rarely do they realize that they

are. If you take yourself to believe in such a God, and if your behavior is less than saintly, you should be terrified about the state of your faith.

I say I value family over work. When I stop and think about it, it seems to me vastly more important to be a good father than to give talks like this one. Yet I'm off to work early, I come home late. I take family vacations and my mind is off in the philosopher's ether. I'm more elated by my rising prestige than by my son's successes in school. My wife rightly scolds me: Do I really believe that family is more important? Or: I sincerely say that those lower than me in social status deserve my respect; but do I really believe this, if I don't live that way?

If my attitudes – my beliefs and my values, especially – are not so much what I sincerely avow when the question is put to me explicitly but rather what is reflected in my overall patterns of action and reaction, in my implicit assumptions, my spontaneous inclinations, then although I may have pretty good knowledge of the simple and trivial, or the relatively narrow and concrete – what I think of April's weather – the attitudes that are most morally central to my life, the ones crucial to my self-image, I tend to know only poorly.

iii.

How about other features of my mentality? My personality traits, my moral character, the quality of my philosophical work, my overall intelligence?

My own view is that traits of this sort are structurally very similar to attitudes. Personality trait attributions, skill attributions, and attitude attributions can all be seen as shorthand ways of talking about patterns of action and reaction (Schwitzgebel 2002b). And our degree of self-knowledge is roughly similar: Our self-knowledge is pretty good about narrow and

concrete matters, especially when an attribution is normatively neutral in the sense that it doesn't tend to cast one in either a good or a bad light, and it's also pretty good when there are straightforward external measures. I know I'm good at Scrabble. That's pretty narrow, concrete, and measurable. I know that I'm more interested in business news than celebrity gossip. Just look at what parts of the newspaper I read.

Now of course there's a whole industry in psychology based on the self-report of personality traits. It often works by asking people broad or medium-sized questions about their traits or attitudes – asking them, for example, whether they enjoy chatting with people or whether they are assertive – and then looks for patterns in the answers. If you generally say yes to questions like that, you'll score as an extravert. There is some stability in people's answers to such questions over time, and some relationship between how people rate themselves in such matters and how their friends rate them. Correlations to outward behavior, though, tend to be at best moderate, and self-evaluations and peer-evaluations tend to break apart when the trait in question is difficult to directly observe and evaluatively loaded (John and Robins 1993; Gosling et al. 1998; Vazire 2010). It's okay to be talkative and it's okay not to be talkative, and talkativeness is a fairly straightforwardly observable trait; self-evaluations and peer evaluations on that fairly narrow and neutral measure tend to line up, and in at least one study (Vazire 2010) both measures were moderately correlated with experimentally observed talking frequency. Self-ratings, peer ratings, and actual behavior tend to align much more poorly, though, for evaluative and relatively fuzzy attributions like being flexible, creative, or lazy. In fact Oliver John and Richard Robins found trends for self-attributions and peer-attributions to correlate *negatively* for some traits: People whose peers judged them to be (relatively) ignorant, undependable, stupid, unfair, or lazy were actually a bit *less* likely to describe themselves as

(relatively) ignorant, undependable, stupid, unfair, or lazy than were people whose peers did *not* attribute them those vices.

Perhaps the most general evaluatively loaded trait attribution is simply whether one is a morally good person. How well do we know this about ourselves? I'd guess that there's approximately a *zero* correlation between people's actual moral character and their opinions about their moral character. Plenty of angels (but not all) think rather poorly of themselves, and plenty of jerks (but not all) think they're just dandy. If you think pretty well of yourself, it's probably just about as likely that you're actually a relatively good and admirable person as that you're overall moral character is below average. It would be nice to have some empirical data on this. Unfortunately, there's not much so far. Both genuine moral self-opinion and real moral character are hard to measure. People are too wily.

However, I suspect that our habit – the habit of most of us, at least, and certainly me – is to assume that we're pretty decent people, above average overall in moral character (even if some of us are too modest to endorse that attitude explicitly), and then to defensively reinterpret and rationalize any counterevidence. I've been trying to get out of this habit myself, and it's highly unpleasant. I've been trying to take an icy look at my moral behavior, applying simple objective standards, and I can't say that I've shown up as well as I hoped. At work, I tend to carry less than the average load of committee duties, suggesting that I'm a shirker; too frequently I forget about meetings with students or even qualifying exams, suggesting I'm self-absorbed; I seem to make more requests and ask for more exceptions than average from editors and conference organizers, suggesting that I'm difficult and demanding. Now I tend to think of myself as a good department citizen, attentive to my students, and relatively easy-going. But so also, I suspect, do most professors, even those who lack such traits. For example, in one survey

study, Joshua Rust and I found that 66% of philosophy professor respondents estimated that they responded to 98% of the emails they receive from students (49% of respondents claimed to respond to 100% of student emails) – statistics which, when we have presented them to undergraduates, typically meet with incredulity and often outright laughter. And when Josh and I sent to our survey respondents some emails designed to look as if they were from undergraduates, those same philosophers who claimed at least 98% email responsiveness responded to just 64% of the emails. Philosophers who gave lower estimates of their email responsiveness responded to 57% of our emails; and overall, self-described responsiveness predicted 1.1% of the variance in measured responsiveness ($r = .11$; $p = .04$; Rust and Schwitzgebel in preparation; Schwitzgebel and Rust in preparation.) Attending to simple objective measures like hours in committee meetings, number of forgotten appointments, and number of special requests, and the number of undergraduates who appear to be frustrated with me (without, surely, just cause) might help serve as a check against my habitual self-deception. Of course, if I cherry-pick objective measures, I can find some that make me look good; but a self-flattering preference for some measures over others is exactly the sort of defensive rationalization that I seek to avoid.

I can carry this icy look over into my personal life, of course, but I don't feel like sharing that here. Unfortunately, it looks no better. I will tell you one whimsical objective measure I've concocted – whimsical, but I do take it *somewhat* seriously. I call it the jerk-sucker ratio. Suppose there's a line of cars slowed down to make a left turn or to exit the freeway. They're not stopped. Their lane is just slower than your lane, because it's crowded with cars planning to turn. The question is, how far along do you go before you change over into that lane? Cutting in at the last moment, of course, is the jerk option – it puts you in front of everyone else without

having waited your turn and furthermore it increases the risk of accident and slows down the cars behind you in your lane who aren't turning or exiting. Getting over early and tolerating the jerks is the sucker option. Suppose there are 48 cars waiting and two who cut in at the last moment. If you're one of those two who cuts in, you're in the 96th percentile for jerks. If there are 85 cars waiting and 15 who cut in, and you cut in, you're in the 85th percentile for jerks. (Of course, this measure breaks down as the ratio of cutters to waiters approaches 1:1.) I might think to myself that I've got better reasons to hurry than all the others waiting or that I'm a skilled enough driver to cut in at the last second without negative consequences. And maybe for some people that's true; but in my own case I worry that that would just be defensive rationalization. Of course, I don't take this little test as a valid measure of jerkhood across the board: There may be little if any relationship between your driving behavior and how you treat your students or your spouse. But this is the *kind* of thing, extended to more serious issues, that constitutes the objectively grounded, icy self-examination I have in mind.

Consider intelligence too, and skill in philosophy. What percentage of the people reading this article, do you think, substantially overestimate their intellectual or philosophical abilities? Might you be among them? (On overestimating intellectual skills see, e.g., Kruger and Dunning 1999; Shynkaruk and Thompson 2006.)

iv.

Self-knowledge? Of general features of our stream of conscious experience, of our morally most important attitudes, of our real values and our moral character, of our intelligence, of what really makes us happy and unhappy – about such matters I doubt we have much real knowledge at all.

We live in cocoons of ignorance, especially where our self-conception is at stake. The philosophical focus on how impressive our self-knowledge is gets the most important things backwards.

Maybe it's good that way. In a classic article, Shelley Taylor and Jonathon Brown (1988), reviewing a broad range of literature, suggest that positive illusions about oneself are the ordinary concomitant of mental health (more recently, see McKay and Dennett 2009; for caveats see Kwan et al. 2008; Moore and Healy 2008); and so also, I suspect, is blasé confidence in answering questions about one's attitudes and stream of experience. It's mainly *depressed* people, Taylor and Brown argue, who have a realistic self-image and an adequate appreciation of their limitations. That's a controversial conclusion, of course, but I start to feel the pull of it.

References:

Armstrong, David M. (1963). Is introspective knowledge incorrigible? *Philosophical Review*, 72: 417–432.

Burge, Tyler (1988). Individualism and self-knowledge. *Journal of Philosophy*, 85: 649–663.

Burge, Tyler (1996). Our entitlement to self-knowledge. *Proceedings of the Aristotelian Society*, 96: 91–116.

Byrne, Alex (2005). Introspection. *Philosophical Topics*, 33 (no. 1): 79–104.

Chalmers, David J. (2003). The content and epistemology of phenomenal belief. In Q. Smith and A. Jokic, eds., *Consciousness: New philosophical perspectives*. Oxford: Oxford.

Churchland, Paul M. (1985). Reduction, qualia, and the direct introspection of brain states. *Journal of Philosophy*, 82: 8-28.

Churchland, Paul M. (1988). *Matter and consciousness*, rev. ed. Cambridge, MA: MIT Press.

Dennett, Daniel C. (2002). How could I be wrong? How wrong could I be? *Journal of Consciousness Studies*, 9 (no. 5–6): 13–6.

Descartes (1641/1984). *Meditations on first philosophy*. In J. Cottingham, R. Stoothoff, and D. Murdoch (eds.), *The philosophical writings of Descartes*, vol. 2. Cambridge: Cambridge University Press.

Dretske, Fred (1995). *Naturalizing the mind*. Cambridge, MA: MIT.

Evans, Gareth (1982). *The varieties of reference*. Oxford: Oxford.

Gertler, Brie (forthcoming). Self-knowledge and the transparency of belief. In A. Hatzimoysis, ed., *Self-knowledge*. Oxford: Oxford.

Goldman, Alvin I. (2006). *Simulating minds*. Oxford: Oxford.

Gordon, Robert M. (2007). Ascent routines for propositional attitudes. *Synthese*, 159: 151–165.

Gosling, Samuel D., Oliver P. John, Kenneth H. Craik, and Richard W. Robins (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, 74: 1337-1349.

Haslanger, Sally (2008). Changing the ideology and culture of philosophy: Not by reason (alone). *Hypatia*, 28 (no. 2): 210-223.

Heil, John (1988). Privileged access. *Mind*, 97: 238–251.

Hume, David (1739/1978). *A treatise of human nature*, L.A. Selby-Bigge and P.H. Nidditch, eds. Oxford: Clarendon.

Hurlburt, Russell T., and Christopher L. Heavey (2006). *Exploring inner experience.* Amsterdam: John Benjamins.

Hurlburt, Russell T., and Eric Schwitzgebel (2007). *Describing inner experience? Proponent Meets Skeptic.* Cambridge, MA: MIT.

Hurlburt Russell T., and Eric Schwitzgebel (forthcoming). Presuppositions and background assumptions. *Journal of Consciousness Studies.*

John, Oliver P., and Richard W. Robins (1993). Determinants of inter-judge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61: 521-551.

Kane, Michael J. (forthcoming). Describing, debating, and discovering inner experience: Review of Hurlburt and Schwitzgebel (2007), "*Describing Inner Experience? Proponent Meets Skeptic*". *Journal of Consciousness Studies.*

Kwan, Virginia S. Y., Oliver P. John, Richard W. Robins, and Lu L. Kuang (2008). Conceptualizing and assessing self-enhancement bias: A componential approach. *Journal of Personality and Social Psychology*, 94, 1062-1077.

McKay, Ryan T., and Daniel C. Dennett (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, 32: 493-561.

Moran, Richard (2001). *Authority and estrangement*. Princeton: Princeton.

Moore, Don A., and Paul J. Healy (2008). The trouble with overconfidence. *Psychological Review*, 115: 502-517.

Nichols, Shaun, and Stephen P. Stich (2003). *Mindreading*. Oxford: Oxford.

Rust, Joshua and Eric Schwitzgebel (in preparation). Ethicists' and non-ethicists' responsiveness to undergraduate emails.

Saul, Jennifer (forthcoming). Unconscious influences and women in philosophy. In F. Jenkins and K. Hutchison, eds., *Women in philosophy* (Newcastle upon Tyne: Cambridge Scholars Publishing).

Schwitzgebel, Eric (2002a). How well do we know our own conscious experience? The case of visual imagery. *Journal of Consciousness Studies*, 9 (no. 5-6): 35-53.

Schwitzgebel, Eric (2002b). A phenomenal, dispositional account of belief. *Noûs*, 36: 249-275.

Schwitzgebel, Eric (2008). The unreliability of naive introspection. *Philosophical Review*, 117: 245-273.

Schwitzgebel, Eric (forthcoming-a). Acting contrary to our professed beliefs, or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly.*

Schwitzgebel, Eric (forthcoming-b). *Perplexities of consciousness.* Cambridge, MA: MIT Press.

Schwitzgebel, Eric, and Joshua Rust (in preparation). The self-reported moral behavior of ethics professors.

Shoemaker, Sydney (1963). *Self-knowledge and self-identity*. Ithaca, NY: Cornell.

Shoemaker, Sydney (1988). On knowing one's own mind. *Philosophical Perspectives*, 2: 183–209.

Shoemaker, Sydney (1994). Self-knowledge and "inner sense" [in three parts]. *Philosophy and Phenomenological Research*, 54: 249-314.

Shoemaker, Sydney (1995). Moore's paradox and self-knowledge. *Philosophical Studies*, 77: 211–228.

Shynkaruk, Jody M., and Valerie A. Thompson (1996). Confidence and accuracy in deductive reasoning. *Memory and Cognition*, 34: 619-632.

Taylor, Shelley E., and Jonathon D. Brown (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103: 193–210.

Vazire, Simine (2010). Who knows what about a person? The Self-Other Knowledge Asymmetry (SOKA) Model. *Journal of Personality and Social Psychology*, 98: 281-300.