

SELF-IGNORANCE

Eric Schwitzgebel

I.

Philosophers tend to be 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 regard 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 ordinary, medium-sized matters (1985, 1988). Daniel 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 cannot in the same way be wrong, it seems, about the fact that I *think* it is 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 is sunny in some sense literally contains as a part the thought or belief that it is sunny.ⁱ Alex Byrne (this volume) argues that we typically ascribe beliefs by following a rule (“if p , believe that you believe that p ”) that is, he says, “strongly self-verifying”: Merely attempting to follow the rule renders the self-ascription true.

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 who portray themselves as at variance with this tradition 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 is 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 do not know our stream of consciousness or our own attitudes very well at all.

II.

First, consider currently ongoing conscious experience. Suppose you are looking directly at a sizeable red object in good light and normal conditions. You judge that you are 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 are feeling pain. How could you possibly be wrong about that either?

Well, in such cases I am 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 in simple judgments 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 on *exactly* the cases of seeing red and feeling pain, generalizing from there, thus implicitly treating those cases as typical.

Why not 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 is your house, say, from a particular point of view. But that is 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 cannot 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 it 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?ⁱⁱ

Now these are pretty substantial questions about your imagery experience. They are not 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 would 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 have talked to find that questions at this level of detail about their conscious experience of imagery are somewhat 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 folding tasks (these are tasks where you are asked to guess what something would look like when folded or unfolded), or tests of visual memory, or tests of visual creativity (Schwitzgebel 2002a, forthcoming-c).

Now of course people might still be quite accurate in their judgments about their visual imagery, even if the differences in their judgments do not 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 will grant me this: There is 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 emotional phenomenology 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 manifestation in 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 cannot, of course, force

a particular answer to these questions. I can only invite you to share my intuitive sense of uncertainty. And it does not 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, or merely the conceptual or theoretical matter of determining which aspects of a well known phenomenology are properly regarded as aspects of emotion.

How about visual sensory experience? Consider not your visual experience when you are looking directly at a canonical color but rather your visual experience, in an ordinary scene, of the region ten degrees or thirty degrees away from the center point. How clear is that region? How finely delineated is your experience of the shape and color of things that you are *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 one to two 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 thirty or fifty or one-hundred degrees of visual arc; still others find *shape* imprecise outside a narrow central area, but find *color* quite distinct even twenty or thirty degrees out. And furthermore, people's opinions about this are not stable over time. In the course of a conversation, they will shift from thinking one thing to thinking another. They change their minds. The phenomenal character of their visual experience is not securely known.ⁱⁱⁱ

We do not really know so much, then, I think, about our stream of consciousness experience, about the phenomenology always transpiring within us. We know *certain* things. I know, perhaps, that I am feeling hungry. But I do not know, as reliably or as well, how that hunger manifests experientially – exactly where, for example, I feel it (if it makes sense to give it an experiential location at all), and what other dimensions of phenomenality it may possess. I

know, perhaps, that I am, or just was, thinking about where to go for lunch, and maybe I know, too, whether I experienced that thought verbally or imagistically or in both ways simultaneously and maybe a few of the grossest contours of that thought. But if I experienced the thought verbally, I may not know, any more and perhaps less than I would in speaking aloud, in what words that thought was expressed (unless, perhaps, I work to create the words I self-attribute in the course of self-attributing them); I may not know whether those words were (or are) experienced as though actively spoken (“inner speech”) as opposed to passively received (“inner hearing”), or whether there was (or is) a motoric aspect to the verbal imagery – or maybe, indeed, whether there really was or is no inner verbalization at all and instead only a non-verbal (“imageless”? “unsymbolized”?) apprehension of that content. Nor, it seems, I am likely to know what the full conscious content of the thought was, or is, assuming that any verbalization only reflects a portion of it:^{iv} People’s reports about such matters are highly variable and unstable, at least before training (Hurlburt and Schwitzgebel 2007, forthcoming); and after training the stability might often be driven more by theory than by accurate apprehension of the target phenomena.^v Perhaps I typically know what my sensory experience is *of* – but I know little about its general form and structure – sometimes not even what sensory modality it occurs in (consider the feeling of being stared at, the entanglement of olfactory and gustatory input, the denial of echolocation [Schwitzgebel and Gordon 2000; Schwitzgebel forthcoming-c]); and to the extent I seem to know details about my sensory experience, typically that knowledge will be grounded in large part (and often either dubiously or vacuously) on my more secure knowledge of the corresponding details of the external world.^{vi} I probably typically know, broadly speaking, if I stop to think about one or another of them, the approximate thrust of my emotions, or my pains, or my somatic urges, or my degree of sleepiness, or the sense of an impending hiccup.

But such knowledge of the approximate gist of our ongoing or very recently past experience, assuming that we do indeed have such knowledge (and sometimes what used to seem obvious and undeniable becomes problematic on further investigation, e.g., color in the eyes-closed visual field [Schwitzgebel forthcoming-c]), 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 tiny island of (apparent) obviousness, that is, in what is mostly a sea of ignorance about our stream of experience. If someone knew so little about the outside world, it would seem the daftest blindness.

Furthermore, the island falls quickly undersea: You probably know the rough 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 is 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 a few rough things about your current experience, but try to extend your knowledge

back more than a few seconds, try to generalize, try to articulate a bit of detail, or try to discern even moderately large structural features of your experience, and soon you will err.^{vii}

III.

How about our self-knowledge of our attitudes? For some of our attitudes I am inclined toward a version of what is sometimes 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 do not 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 is 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”).^{viii}

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 am 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 fairly straightforward and 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.

Consider sexism.^{ix} 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 cannot 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 surprised than he would were a man to do so, even though his female students make insightful comments and submit excellent essays at the same rate as his male students. When Ralph is on the hiring committee for a new office manager, it will not 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 see no 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.^x

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 is no lack of authority about what one *believes*. Ralph believes that men and women are equally intelligent, you might suggest, he just tends not to *act* on that belief. But this seems to me an overly linguistic and intellectualist view of belief. Our beliefs manifest not just in what we say, but in what we do – they animate our limbs, not just our mouths – and they are also manifested in our spontaneous emotional reactions and our implicit assumptions. Now I think it is not quite right to call Ralph an out-and-out sexist who simply 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 how he thinks and acts, even if it is only a part. Ralph's attitude toward the intellectual equality of the sexes is what I would call an in-between state. His dispositions, his patterns of response, his habits of thought, are mixed up and inconsistent. It is 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.^{xi} 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 diverges from behavioral manifestation and spontaneous response. To believe in God, in the mainstream monotheistic sense, is in part to believe that there is 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 will do before God what they would not 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 write papers like this one. Yet I am off to work early, I come home late. I take family vacations and my mind is wandering in the philosopher's ether. I am 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? (*Do I live that way? How respectfully do I treat cashiers, students, secretaries? I doubt I really know. Ask them when I am not around.*)

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, either through a facile assumption of alignment between my avowals and my overall patterns of action and reaction or through empirical generalizations of doubtful accuracy, filtered through the distorting lens of self-flattery.

How about other features of my mentality? My personality traits, my moral character, the quality of my philosophical thinking, 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 inward and outward 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 does not tend to cast one in either a good or a bad light, and it is also pretty good when there are straightforward external measures. I know I am good at Scrabble. That is pretty narrow, concrete, and measurable. I know that I am more interested in business news than celebrity gossip. Just look at what parts of the newspaper I read.

Now of course there is 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 will 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 is okay to be talkative and it is okay not to be talkative, and talkativeness is a fairly straightforwardly observable trait; self-evaluations and peer evaluations

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 attributions like being flexible, creative, or lazy. In fact John and Robins (1993) found self-attributions and peer-attributions tending to correlate *negatively* for some such 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.

One of the most general evaluatively-loaded trait attributions is whether one is a morally good person. How well do we know this about ourselves? I would guess that there is 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 is probably just about as likely that you are actually a relatively good and admirable person as that your overall moral character is below average. It would be nice to have some empirical data on this. Unfortunately, both genuine moral self-opinion and real moral character are hard to measure. People are complex and wily.

I suspect that our habit – the habit of most of us, at least, and certainly me – is to assume that we are 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 have been trying to get out of this habit myself, and it is highly unpleasant. I have been trying to take an icy look at my moral behavior, applying simple objective standards, and I cannot say that I have shown up as well as I hoped. At work, I tend to

carry less than the average load of committee duties, suggesting that I am a shirker; too frequently I forget about meetings with students or even qualifying exams, suggesting that I am self-absorbed; I seem to make more requests and ask for more exceptions than average from editors and conference organizers, suggesting that I am 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.^{xii} Attending to simple objective measures like hours in committee meetings, number of forgotten appointments, 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 would rather not share that here. Unfortunately, it looks no better. I will tell you one whimsical objective measure I have concocted – whimsical, but I do take it *somewhat* seriously. I call it the automotive jerk-sucker ratio. Suppose there is a line of cars slowed down to make a left turn or to exit the freeway. They are not stopped. Their lane is just slower than your lane, because it is 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 are not turning or exiting. Getting over early and tolerating the jerks is the sucker option. Suppose there are forty-eight cars waiting and two who cut in at the last moment. If you are one of those two who cuts in, you are in the 96th

percentile for jerks. If there are eighty-five cars waiting and fifteen who cut in, and you cut in, you are 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 have better reasons to hurry than all the others waiting or that I am a skilled enough driver to cut in at the last second without negative consequences. And maybe for some people that is true; but in my own case I worry that that would just be defensive rationalization. Of course, I do not regard this little test as a valid measure of jerkhood across the board: There may be little if any relationship between one's driving behavior and how one treats one's students or 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 misestimate their intellectual or philosophical abilities? Psychologists have repeatedly found that North Americans and Western Europeans – especially men – tend to rate themselves as significantly more intelligent than their peers.^{xiii} Psychologists have also repeatedly found only modest correlations between self-rated intelligence and intelligence as measured by the IQ tests (with self-ratings typically accounting for only about 1%-9% of the variance in test scores).^{xiv} Of course, IQ tests might be poor measures of intelligence: That would explain the poor correlation with self-report. It would also nicely explain the overestimation tendency if people tended self-flatteringly to regard *real* intelligence to be revealed by just the sorts of things they think themselves good at – whether that be mathematics, verbal fluency, business acumen, or diagnosing automotive troubles. (“Jeez, he doesn’t even know what a slipping belt sounds like? What an idiot!”) Evidence also suggests that correlations are often quite modest between self-rated ability and measured performance,

compared to peers, in various intellectual sub-domains such as grammatical ability, spatial ability, and logical reasoning – though to my knowledge philosophical ability in particular has not been tested.^{xv} Based both on general psychological evidence and on personal experience, I will hazard this guess about you, the reader (and myself too): Your opinion of how your intelligence and philosophical ability compares to the intelligence and philosophical ability of your classmates (if you are a student) or colleagues (if you are a professor) is at best a weak indicator of your actual intelligence and ability. Probably your peers know you better. But don't bother asking; they won't tell you the truth.

V.

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 (see Haybron 2008) – about such matters I doubt we have much 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 is 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;^{xvi} and so also, I suspect, is blasé confidence in answering questions about one's attitudes and stream of experience. It is mainly *depressed* people, Taylor and Brown argue, who have a realistic self-image and an adequate appreciation of their limitations. That is 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.
- Bayne, Timothy and Michelle Montague, eds., (forthcoming). *Cognitive Phenomenology*. Oxford.
- 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 (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT.
- Dretske, Fred (1995). *Naturalizing the Mind*. Cambridge, MA: MIT.
- Evans, Gareth (1982). *The Varieties of Reference*. Oxford: Oxford.

- Ford, Jason (2008). "Attention and the new skeptics." *Journal of Consciousness Studies*, 15 (3): 59-86.
- Furnham, A. (2001). "Self-estimates of intelligence: Culture and gender differences in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, 31: 1381-1405.
- 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.
- Haybron, Daniel (2008). *The Pursuit of Unhappiness*. Oxford: Oxford.
- 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.
- Kusch, Martin (1999). *Psychological Knowledge*. London: Routledge.
- 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.
- Paulhus, Delroy L., Daria C. Lysy, and Michelle S. M. Yik (2008). “Self-report measures of intelligence: Are they useful as proxy IQ tests?” *Journal of Personality*, 66: 525-554.
- 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). “Introspection, what?” In D. Smithies and D. Stoljar, eds., *Introspection and Consciousness*. Oxford: Oxford.
- Schwitzgebel, Eric (forthcoming-c). *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’.” *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.
- Steinmayr, Ricarda, and Birgit Spinath (2009). “What explains boys’ stonger confidence in their intelligence?” *Sex Roles*, 61: 736-749.
- 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.
- Visser, Beth A., Michael C. Ashton, and Philip A. Vernon (2008). “What makes you think you’re so smart? Measured abilities, personality, and sex differences in relation to self-estimates of multiple intelligences.” *Journal of Individual Differences*, 29: 35-44.

ⁱ The characterization is a bit simplified, but Burge (1988, 1996), Heil (1988), Dretske (1995), and Shoemaker (1995) have said things along roughly these lines.

ⁱⁱ Jason Ford (in print: Ford 2008) and others (in conversation) have suggested that earlier versions of this exercise (Schwitzgebel 2002a) involve illegitimate questions about the periphery of experience – questions only appropriate to more detailed, central experience. I believe that this objection is misguided: While it might be illegitimate to ask exactly what specific colors and shapes inhabit the periphery (if the periphery is indistinct), it does not seem to me in the same way illegitimate to ask *whether* colors and shapes in the periphery are clear or indistinct.

However, I have heard this objection so often now that I worry there might be something to it

that I stupidly or stubbornly refuse to see. Consequently, I now conclude with several questions that do not seem to concern the periphery.

ⁱⁱⁱ The arguments in the last two paragraphs are adapted from Schwitzgebel 2008, forthcoming-c.

^{iv} Consider incomplete or unspecific or misspoken inner verbalizations: “Why can’t I remember about the...”, “John! [in an exasperated tone]”, “That’s not an unfair deal” [bitterly, with a single negative rather than a double negative intended]. In such cases, the verbal content does not fully reflect the apparently experienced thought content. So also, I am inclined to think, in many other cases where the mismatch or incompleteness is less obvious.

^v As in the debate about “imageless thought” in the early 20th century and its contemporary descendant, the debate about cognitive phenomenology (Kusch 1999; Bayne and Montague, eds., forthcoming).

^{vi} See Schwitzgebel forthcoming-b for a more detailed discussion of the processes driving introspective judgment.

^{vii} This section rides quickly through material covered in more depth in Schwitzgebel forthcoming-c.

^{viii} For transparency views of self-knowledge of attitudes, see, e.g., Evans 1982; Moran 2001; Byrne 2005, this volume; Gordon 2007; against transparency see Gertler forthcoming.

(Incidentally, I wish that Byrne, in this volume, had suggested that we tend to self-ascribe desires on the basis of a rule like “If X is good, believe that you want X”, rather than casting the antecedent in terms of desirability. It seems to me that we much more often think about whether things would be good, and self-ascribe desire partly on that basis, than we think about whether things would be desirable. Such an account might suggest less privilege than an account in terms of desirability, but I regard that as a feature, not a bug.)

^{ix} This example is adapted from Schwitzgebel forthcoming-a.

^x For empirically informed discussion of professional philosophers' apparent ignorance of their own sexism, see Haslanger (2008) and Saul (forthcoming).

^{xi} I develop this idea further in Schwitzgebel 2002b, forthcoming-a.

^{xii} 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 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.)

^{xiii} Findings on Asians are mixed. See Furnham 2001 for a review. Some recent studies are Visser, Ashton, and Vernon 2008; Steinmayr and Spinath 2009.

^{xiv} See e.g., Paulus, Lysy, and Yik 1998; Furnham 2001; Visser, Ashton, and Vernon 2008; Vazire 2010.

^{xv} See, e.g., Kruger and Dunning 1999; Shynkaruk and Thompson 2006; Visser, Ashton, and Vernon 2008.

^{xvi} More recently, see McKay and Dennett 2009; for caveats see Kwan et al. 2008; Moore and Healy 2008.