

## REPLY TO BOYLE, CARRUTHERS AND SCHWITZGEBEL

Pacific APA, Vancouver 2 April 2015

My book *Self-Knowledge for Humans* was the subject of an ‘author meets critics’ session at the Pacific APA in April 2015. Below is the text of my response to Matthew Boyle, Peter Carruthers and Eric Schwitzgebel. Carruthers’ contribution was published days after the session in the form of a review of my book for NDPR (<http://ndpr.nd.edu/news/57083-self-knowledge-for-humans/>). Schwitzgebel’s piece on the intrinsic value of self-knowledge is available here: <http://www.faculty.ucr.edu/~eschwitz/SchwitzAbs/IntrinsicSelfK.htm>. As far as I’m aware, an electronic version of Boyle’s commentary is not widely available at the present time.

---

One of the aims of *Self-Knowledge for Humans* is to make the case that philosophers of self-knowledge should spend much more time talking about what I call *substantial* self-knowledge. Carruthers agrees but worries that I understand this type of self-knowledge too narrowly. He says that I conceive of substantial self-knowledge as knowledge of one’s *individual* characteristics rather than ‘knowledge of oneself qua human being, or qua human being of some well-defined sub-category (*adult male*, for example, or *introvert*)’. He sees this as a serious omission and an opportunity lost because ‘many of the most important forms of self-knowledge.... derive from generalizations about human cognition and motivation discovered by psychologists and other cognitive scientists’. Substantial self-knowledge in this sense is as much about ‘knowing one’s own humanity’ as about knowing one’s own individual characteristics.

I’m a little puzzled by the idea of substantial self-knowledge as knowledge of oneself qua human being or some well-defined sub-category. Carruthers’ initial way of putting this, and the contrast with knowledge of one’s own individual characteristics, makes it sound as

though what is being proposed is a distinctive *subject-matter* for substantial self-knowledge. Whereas the individual is the focus of the substantial self-knowledge I talk about, what Carruthers is after is what might be called *generic* self-knowledge. But exactly what is that? If I know that I am prone to bouts of self-pity that is a piece of specific self-knowledge. If I know that human beings are prone to being influenced by implicit biases in their thinking this is presumably an example of generic knowledge. But is it *self-knowledge*? Only if know that and how the generalization implies to me. I might wonder, for example, which implicit biases are influential in *my* thinking and now we are back to specific self-knowledge. Indeed, if the generalizations of psychologists and cognitive scientists are expressive of generic self-knowledge then why not the generalizations of biologists and geneticists? Presumably these provide me with knowledge of myself qua human being but it seems they don't provide me with anything recognizable as self-knowledge in the philosophical or ordinary sense.

Perhaps this isn't what Carruthers intends. On a different reading, generic self-knowledge is not self-knowledge with a distinctive *subject-matter*, but self-knowledge with a distinctive *source*. For example, knowledge of my implicit biases is still individual and specific since the focus is on *my* implicit biases rather than anyone else's. However, in investigating and coming to know my implicit biases I would no doubt be more likely to arrive at the truth if I am aware in general terms of the influence exerted by implicit biases or peoples' reliance on implicit stereotypes. The resulting self-knowledge will still be individual in its subject-matter but will derive in part from generalizations about human cognition. This is substantial self-knowledge with a generic *source* rather than a generic *subject-matter*, and I take it for granted that much of our substantial self-knowledge will be generic in this sense. That is why so much of my book is about the differences between human beings and *homo philosophicus*. So in the end I'm not sure that Carruthers and I really disagree. It only looks

like that if one fails to distinguish clearly between two different senses in which substantial self-knowledge might be generic, the sources sense and the subject-matter sense.

Now for Schwitzgebel's discussion. Let me start by outlining my account of the value of self-knowledge. My question in chapter 15 is: 'what's so good about self-knowledge and bad about self-ignorance?'. My starting point is that substantial self-knowledge isn't intrinsically valuable, that is, valuable for its own sake. I take it that *if* self-knowledge is valuable it is because it promotes other goods whose value is not in question. One possibility is self-knowledge is valuable because it promotes well-being or contributes to the overall success of one's life. I call this a 'low road' explanation of the value of self-knowledge because it represents the value of self-knowledge as instrumental. In contrast, what I call 'high road' explanations explain the value of self-knowledge by reference to highly abstract, high-sounding ideals like authenticity and unity. I then go on to criticize a range of specific high-road explanations. For example, I argue that the value of self-knowledge doesn't derive from a strong connection with authenticity and that there are in any case questions about the value of authenticity. My conclusion is not that the high road accounts I discuss are totally useless but that when it comes to explaining the value of self-knowledge it's better to take the low road. This is how I end the chapter and, indeed, the book: 'when you are trying to explain the value of self-knowledge, don't be shy about stating the obvious: self-knowledge derives whatever value it has from the value of what it makes possible, and what it ultimately makes possible is for us to live well' (p. 227).

Schwitzgebel's presentation of my position gives the impression that high road arguments are attempts to establish the intrinsic value of self-knowledge. They aren't. Even if, contrary to what I argue in the book, self-knowledge is essential for authenticity the value of self-knowledge would in my terms still be 'extrinsic' rather than 'intrinsic' (see p. 211 middle paragraph); the value of self-knowledge would still derive from the value of the

authenticity it makes possible. It wouldn't be valuable *in itself*, without reference to anything else. But Schwitzgebel seems to think that self-knowledge *is* intrinsically valuable. This would mean that even if self-knowledge makes no difference to anything else we care about it would still be worth having and pursuing. I find this view bizarre.

Consider these scenarios. In scenario 1 self-knowledge literally makes no difference: having it has zero impact (positive or negative) on the quality or overall success of one's life. In scenario 2, self-knowledge does make a difference but its impact on one's life is negative. If you think that self-knowledge is intrinsically valuable then you must think that it makes no difference that self-knowledge makes no difference in scenario 1. When pressed to explain why you think that zero-impact self-knowledge is valuable, or what makes it valuable, there wouldn't be a whole lot you could say, beyond repeating that it is valuable in itself and that there is nothing else that makes it valuable. I don't believe this is either a sensible view of the value of self-knowledge or one that many people hold. Of course most people in our culture think that self-knowledge is worth having but that's because they think that it has a positive impact on the lives of those who have it, whether that positive impact is understood in high road or low road terms. And what if it turns out that we are wrong about that and that the net impact of self-knowledge is negative rather than positive, as in scenario 2? Suppose, like King Lear, we pay a high price for self-knowledge, and that having it makes us utterly miserable. Would we think 'Never mind, at least I know myself and that's good for its own sake'? I can't speak for anyone else but that wouldn't be my reaction. I think that the 'intrinsic value of self-knowledge' is a myth, and if we ever talk in these terms our words shouldn't be taken literally. Talking about the intrinsic value of self-knowledge is just a picturesque way of giving expression to our sense of its importance for other things we care about.

Schwitzgebel's first argument for the intrinsic value of self-knowledge is what he calls the Argument from Addition and Subtraction. Here's the subtraction version: imaginatively subtract from the world large portions of our self-knowledge while keeping everything else constant. Has something valuable gone missing? Yes, says Schwitzgebel, even if the loss of self-knowledge has no practical consequences. To me, this sounds not so much like an argument as a vivid restatement of the view that self-knowledge is intrinsically valuable. My suspicion is that to the extent that we do think that something valuable has gone missing that is because we don't take seriously the stipulation that loss of self-knowledge has zero impact. Imagine moving from knowing what I'm good at, or what makes me happy, to not knowing, in the sense that now my beliefs about what I'm good at or what makes me happy are now all false. How is that supposed to make no practical difference or have zero impact on the quality and overall success of my life? Asking whether we would still care about zero-impact self-ignorance is asking a question which we are not well equipped to grasp, let alone answer.

Schwitzgebel's second argument says that if knowledge of other people is intrinsically valuable then so is self-knowledge. But is knowledge of other people intrinsically valuable and would it follow that self-knowledge is intrinsically valuable? If pressed to explain why knowledge of other people matters I guess I'd want to talk about its contribution to such things as love and friendship. You need love and friendship to live well, and this means that you need to know other people to live well. If this makes knowledge of other people only extrinsically valuable that's fine by me. I take it that self-knowledge also matters for love and friendship, and that is one reason why it's hard to subtract self-knowledge from the world while leaving everything else constant. So I agree with Schwitzgebel's that there are interesting parallels between the value of self-knowledge and the value of knowledge of other people. The parallel is that the value of neither is 'intrinsic'.

Schwitzgebel's final argument says self-knowledge is tied to dispositions or capacities and that the relevant dispositions or capacities are intrinsically valuable. In what sense? In the sense that, as Schwitzgebel puts it, 'they are central to living a rich, meaningful life' and that 'if we subtract them away, we impoverish ourselves'. But this doesn't sound very different from the view of the value of self-knowledge which I defend: self-knowledge is valuable, if it is, because and to the extent that it enables us to live well, that is to live rich, meaningful and fulfilled lives. If this is what Schwitzgebel thinks then I don't know why he thinks he disagrees with me or insists on describing the value of self-knowledge as 'intrinsic'. The disagreement between our views seems terminological rather than real.

Unfortunately the same can't be said about the dispute between me and Boyle. There is far too much in Matt's discussion for me to respond to in detail today so I'll focus instead on some general themes, starting with my understanding of transparency. Boyle claims that what I call 'rationalism about self-knowledge' is a chimera constructed by fusing the transparency approach to a commitment that transparency theorists have no reason to accept. Is that really so? It's true, as Boyle points out, that the basic idea of transparency can be stated without any mention of rationality. The basic idea is simply that it is possible for me answer the question whether I hold a given attitude by answering a corresponding outward-directed question. However, as Moran observes, it isn't enough to point out that such a thing is possible. It also needs to be explained *how* it is possible to answer the question whether, say, I believe that P by answering the question whether P. What right, Moran asks, have I to think that 'my reflection on the reasons in favour of P (which is one subject-matter) has anything to do with the question of what my actual *belief* about P is (which is quite a different subject matter?)'.

This pertinent question can be and has been answered in many different ways but the rationalist is someone for whom reasons and rationality are the key. In Moran's words:

If the person were entitled to assume, or in some way even obligated to assume, that his considerations for or against believing P (the outward-directed question) actually determined in this case what his belief concerning P actually is (the inward-directed question) then he would be entitled to answer the question concerning his believing P or not by considerations of the reasons in favor of P' (2004: 457).

As I put it in my book, 'what makes this a *Rationalist* account of self-knowledge is that it takes your belief concerning P to be determined by your reasons and so to be knowable by reflection on your reasons' (p. 102). I don't believe that rationalism in this sense is a figment of my imagination. Rather, it is a position suggested by a straightforward reading of Moran, who at one point describes the simple Transparency condition as 'bringing us up to the region of something like a Transcendental assumption of Rational Thought' (2003: 406).

Understood in this way Rationalism about self-knowledge raises a whole bunch of questions which I'm certainly not the first to have asked and won't be the last. For example, what is the role in the transparency procedure of the assumption that my belief concerning P is determined by my reasons for or against believing P? What if this Rationality Assumption is false, and my belief is determined by something other than my reasons? How can the transparency procedure be adapted to explain self-knowledge of attitudes other than belief? If we are dealing with *homo philosophicus* then we won't have to worry about the possibility that his belief concerning P isn't determined by his reasons, or can't be known by reflecting on his reasons. Rationalism about self-knowledge is tailor-made for *homo philosophicus*, but is it tailor-made for us, given all the respects in which we aren't *homo philosophicus*?

On the last of these questions, I don't rule out the possibility that the transparency method is a viable source of self-knowledge for us. I don't reject the transparency method but argue instead that it's a relatively peripheral source of self-knowledge for humans and that the self-knowledge it makes available is inferential. The transparency method understood as

Moran understands it fails to account for many other varieties of self-knowledge, including varieties of self-knowledge which most of us care about the most. If the question is whether you love your children or your spouse it isn't obvious, to put it mildly, that the way to answer the question is to reflect on your reasons for loving your children or your spouse.

This brings us neatly to the issue of alienation. I don't have any problems with the idea that the transparency method only gives us inferential self-knowledge, and on this point I'm in complete agreement with non-rationalist proponents of transparency like Alex Byrne. Where does Boyle stand on this issue? His view is that I *must* be alienated from any attitude I know I have by inference. As he puts it, 'even if I endorse the attitudes I ascribe to myself by inference, there is a kind of alienation implied in the idea that I should need to discover them by inference'. Non-alienated self-knowledge would be self-knowledge acquired by using the transparency method. But I contend that the latter only gives us inferential self-knowledge so now we have an inconsistent triad: inferential self-knowledge is alienated, self-knowledge acquired by employing the transparency method is inferential, and the self-knowledge acquired by using the transparency method isn't alienated. This inconsistent triad brings out a peculiarity in rationalism about self-knowledge. On the one hand, proponents of this view are keen to place the transparency method at the centre of their account of self-knowledge. On the other hand, they claim that non-alienated self-knowledge is immediate. But how can non-alienated self-knowledge be immediate if the transparency method is its source?

One response to the inconsistent triad might to argue that use of the transparency method can make available non-inferential self-knowledge. But how can that be, given the role of the Rationality Assumption in justifying one's beliefs about one's own beliefs? If one's self-ascription is justified in part by this assumption then the resulting self-knowledge must be inferential. And if this isn't the role of the Rationality Assumption then what work is it doing? But even without going into these issues I think there are still reasons to doubt that

inferential self-knowledge must be alienated. Since it is controversial whether self-knowledge acquired by using the transparency method is inferential let's consider a different type of example in which there is no dispute that what is at issue is inferential self-knowledge. This is Krista Lawlor's case of Katherine who comes to know that she wants another child by reflecting on her feelings, emotions, fantasies, inner speech and other internal promptings. Her *route* to self-knowledge is inferential but that doesn't make her knowledge or the desires she comes to know by inference alienated. As I argue in the book, 'the fact that the desire is manifested in her emotions, dreams, and fantasies might actually *enhance* her sense of the desire as fully her own, an expression of who she is at this point in her life' and that 'if this is the case, then the fact that she comes to know her desire on the basis of these same emotions, dreams, and fantasies can hardly make her self-knowledge alienated' (p. 157). It just doesn't follow from the fact that self-knowledge is inferential that it is alienated.

I want to end with some general remarks about the philosophy of self-knowledge. As I see it, philosophical theories of self-knowledge vary along several different dimensions. To begin with there is the *easy/difficult* dimension. If you're at the easy end then you are impressed by how easy self-knowledge is. You don't deny sometimes self-knowledge doesn't come easily but you are more interested in accounting for the self-knowledge that does. If you are at the difficult end then you recognize that some self-knowledge is easy to come by but are more interested in the self-knowledge that isn't. Then there is the *metaphysics/epistemology* dimension. Metaphysical approaches concentrate on the metaphysics of belief and other such attitudes. They don't see the acquisition of self-knowledge as a matter of making a transition from, say, believing that you are wearing socks to knowing that you believe you're wearing socks. They think that beliefs are by their nature self-intimating and that if you believe you are wearing socks then you tacitly know that this is what you believe. Epistemological views, in contrast, deny such claims and see the problem of self-knowledge

in epistemological rather than metaphysical terms: the challenge is to explain how we come to know our own standing attitudes given that it isn't built into our attitudes that we know of them. Finally, there is the ideal/real dimension. Ideal theories have a highly idealized conception of the self-knower whereas real theories pride themselves on their realism about our capacities and limitations.

In these terms, philosophers such as Descartes and Shoemaker are metaphysical rather than epistemological in their orientation, and focus on easy rather than hard self-knowledge. I think the same is true of Boyle, and that is one reason why we disagree so fundamentally. I'm more interested in what makes self-knowledge difficult rather than in what makes it easy, and my focus is epistemological rather than metaphysical. I don't take it to be a basic and irreducible fact about someone who has, say, racist or sexist beliefs, that they also have tacit knowledge of this fact, or that their racist and sexist beliefs are constitutively self-intimating. Similarly, there are deep desires and fears we may not realize we have, and coming to know that we have them, like coming to recognize our own racist and sexist beliefs, represents a cognitive achievement which is far beyond the reach of many of us.

Once we start thinking in terms of attitudes we may not know we have, and of the numerous obstacles on the road to worthwhile self-knowledge, we are now thinking in 'real' rather than 'ideal' terms about self-knowledge. I regret that when I wrote *Self-Knowledge for Humans* I wasn't familiar with the work of Charles Mills but I think that some of the points he makes about the vices of 'ideal' moral theories also apply to 'ideal' theories of knowledge, including self-knowledge. As Mills puts it, 'what distinguishes ideal theory is the reliance on idealization to the exclusion, or at least marginalization, of the actual' and the attribution to humans agents as visualized in the theory of 'completely unrealistic capacities'. This is what I was getting at with the distinction between *homo sapiens* and *homo philosophicus*, though I didn't get into the ideological questions Mills discusses.

Obviously many ideal theories don't even recognize themselves as such, hence Boyle's suggestion that rationalism about self-knowledge is a figment of my imagination. No doubt those philosophers who I'm calling 'rationalists about self-knowledge' would flatly deny that they are concerned with anything other than self-knowledge for humans, but I believe they're wrong about that. You can't really philosophize about self-knowledge without making substantive if not very explicit assumptions about the cognitive capacities, achievements and limitations of the human knowers you are supposedly discussing. These assumptions need to be subjected to closer scrutiny than they have traditionally received, and I continue to find it very hard to recognize real human beings in the philosophical writing on self-knowledge to which I was responding in *Self-Knowledge for Humans*. My claim is that, as philosophers, we should be writing *for* and *about* real human beings, and that is what I'm trying to do in *Self-Knowledge for Humans*. The aim was to rattle some cages and I'm happy to see that I seem to have succeeded.