Carruthers and Constitutive Self-Knowledge

John C. Hill

Abstract: In his recent book, *The Opacity of Mind*, Peter Carruthers advances a skeptical theory of self-knowledge, integrating results from experimental psychology and cognitive science.[1] In this essay, I want to suggest that the situation is not quite as dire as Carruthers makes it out to be. I respond to Carruthers by advancing a constitutive theory of self-knowledge. I argue that self-knowledge, so understood, is not only compatible with the empirical research that Carruthers utilizes, but also helps to make sense of these results.

Introduction

In his recent book, *The Opacity of Mind*, Peter Carruthers advances a skeptical theory of self-knowledge, integrating results from experimental psychology and cognitive science. Carruthers argues that our knowledge of ourselves is always mediated by a mindreading faculty that never directly perceives our attitudes. Indeed, recent work in experimental psychology and cognitive science challenges philosophical conceptions of self-knowledge through an emphasis on ways in which the sub-personal processes beyond our conscious awareness go awry.

In this essay, I want to suggest that the situation is not quite as dire as Carruthers suggests. Self-knowledge is tied to normative notions like commitment and agency in ways that other-knowledge is not. As a result, a more thorough understanding of ourselves, warts and all, can help us see where we make mistakes and how to improve on them. In this essay, I proceed as follows: first, I motivate a constitutivist theory of self-knowledge.[2] Next, I introduce and evaluate Carruthers proposed theory of self-knowledge, the interpretative sensory-access theory, and explore its skeptical conclusions. I

[1] Peter Carruthers, *The Opacity of the Mind* (Oxford: Oxford University Press, 2011).

[2] I adopt the term 'constitutivism' with some hesitation. In the literature on self-knowledge, constitutivism sometimes refers to a view of self-knowledge according to which it is constitutive of belief (or other propositional attitudes) that if one believes that p, then one believes that one believes that p. In other words, the nature of belief itself entails self-knowledge of beliefs. For examples of this view, see Sydney Shoemaker, "Self-Knowledge and 'Inner Sense'," *Philosophy and Phenomenological Research* 54 (1994): 249–314; Eric Schwitzgebel, "Knowing Your Own Beliefs," *Canadian Journal of Philosophy* 35 (2011): 41-62; and Declan Smithies, "A Simple Theory of Introspection," in *Introspection and Consciousness*, eds. Declan Smithies and Daniel Stoljar (Oxford: Oxford University Press, 2012), 259-93. Another candidate term, for instance, 'contructivism,' has anti-realist connotations that I would like avoid (see below). Consequently, I have settled on the term 'constitutivism', but keep in mind that the type of constitution at issue is self-constitution, rather than the constitution of belief itself.

argue that constitutivism, because it reconceives the nature of self-knowledge, is especially suited to incorporate such findings while preserving an important place for our ability to determine the state of our minds.

Constitutive Self-Knowledge

The constitutive theory of self-knowledge is a distinctive theory of the asymmetries that hold between one's relation to one's own mind and one's relation to the minds of others. The core claim of the constitutive account of self-knowledge is that knowledge of one's state of mind is a process of self-constitution, rather than self-discovery.

The constitutive account of self-knowledge can be described in opposition to an observational account of self-knowledge. According to an observational account of self-knowledge, when one ascribes an attitude to oneself, one reports an already given mental state which one observes via the mental faculty of introspection. Accordingly, observational accounts of self-knowledge explain one's first-person authority vis-à-vis one's own attitudes in terms of a privileged perspective from which one observes those attitudes. However, one's ascriptions of mental attitudes to others lack such special access, and consequently are less epistemically secure.

As many philosophers have argued, such observational theories of self-knowledge cannot account for the immediacy of our self-knowledge.[3] That is to say, when one ascribes an attitude to oneself, one typically does not attempt to gather evidence or try to observe what one thinks; instead, one merely ascribes that one is in such-and-such a mental state. If our practice of self-ascription is not wholly unwarranted, then the authority of our self-ascriptions must derive from elsewhere.

In contrast to observational accounts of self-knowledge, constitutive accounts of self-knowledge argue that one's first-person authority results from one's capacity to effectively constitute one's own attitudes. As Finkelstein puts it, "typically, what I say or think about my own mental state plays a constitutive role in determining what it is."[4] In other words, one has authority over one's mental life because one can make up his or her mind about some state of affairs. For instance, Susan knows that she believes her friend Amy is trustworthy, not because she has some special evidence which bears on her mental state, but because in considering the question of whether she believes Amy is trustworthy, she forms a novel attitude toward Amy's trustworthiness. In effect, she makes up her mind regarding Amy's trustworthiness.

According to the constitutive view, self-knowledge is the result of our capacity to commit ourselves to various attitudes. For instance, Bilgrami argues

---

[3] The constitutivism I endorse is inspired in part by Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton: Princeton University Press, 2001). Cf. Tyler Burge, "Individual and Self-Knowledge," *Journal of Philosophy* 85, no. 11 (1988), 649-63 and Akeel Bilgrami, *Self-Knowledge and Resentment* (Cambridge: Harvard University Press, 2006).

[4] David Finkelstein, *Expression and the Inner* (Cambridge: Harvard University Press, 2003), 28.

that "to desire something, to believe something, is to think that one ought to do or think various things."[5] When one ascribes an attitude to oneself or makes an avowal concerning one's state of mind, one undertakes a commitment to reason and deliberates in accordance with the normative requirements associated with adopting such an attitude. For instance, to constitute oneself as someone who desires social justice is to commit oneself to bringing about social justice. It is the capacity to commit oneself that explains the asymmetry between self-ascriptions and other-ascriptions. Importantly, one is responsible for one's commitments. In the constitutive view, one's authority is a result of one's responsibility for his or her state of mind.[6]

At this point, one might worry that the constitutivist theory entails that whatever one thinks about one's state of mind completely determines its character. Interpreted in this way, the constitutivist theory of self-knowledge is highly unrealistic. Finkelstein, for instance, complains, "constitutivism has the effect of misrepresenting the subject's responsibility for mental states about which he speaks with first-person authority . . . If my remarking that I had a headache made it the case that I did have a headache, then sympathy would not be appropriate in response to such a remark."[7] Or, consider Coliva, who writes regarding constitutivism, "psychological self-ascriptions such as 'I believe/desire/intend/wish/hope that P' do bring into existence the corresponding first-order mental states . . . On this model, there would be a sense in which it is literally true that we make up or create our minds."[8] In other words, constitutivism maintains that we have control over our mental lives that we seemingly do not. What is more, it seems as though constitutivism ends up committed to a stronger form of infallibility than observationalism. Considering that constitutivism is compatible with substantial fallibility, this is a poor result.

However, the process of self-interpretation that is involved in self-constitution is constrained. The act of self-interpretation is constrained by one's antecedent attitudes. Moran accurately locates the perceived problem in the language used to describe the process of self-interpretation:

> Verbs such as 'describe', 'interpret', and the like are fated to equivocate between a use that expresses one's genuine sense of how things are, with the same kind of commitment as belief, and a different, noncommittal use denoting an ordinary activity. Favoring verbs in this latter

[5] Bilgrami, *Self-Knowledge and Resentment*, 213.

[6] Cf Keith Frankish, *Mind and Supermind* (Cambridge: Cambridge University Press, 2004); Keith Frankish, "Dual Systems and Dual Attitudes," *Mind and Society* 11 (2012), 45-51; Peter Carruthers, *The Opacity of the Mind* (Oxford: Oxford University Press, 2011); and Peter Carruthers, "On Knowing Your Own Beliefs," in *New Essays on Belief: Structure, Constitution and Content*, ed Nikolaj Nottelmann (New York: Palgrave MacMillan, forthcoming).

[7] Finkelstein, 52.

[8] Annalisa Coliva, "One Variety of Self-Knowledge: Constitutivism or Constructivism," in *The Self and Self-Knowledge*, ed. Annalisa Coliva (Oxford: Oxford University Press, 20

> sense serves to dramatize the idea of self-transforming redescription, and to obscure its genuine basis.[9]

The upshot is this: there are two different senses of 'interpret' with different consequences for the process of self-constitution. The first sense allows for a great degree of freedom in process, by analogizing self-interpretation with an act of describing, unconstrained by cognitive considerations. The second sense constrains the process of self-constitution by connecting it with commitment. In self-interpreting, an individual forms a belief about himself or herself, sensitive to the same epistemic considerations as any other belief. For example, an individual who comes to see a cigarette, which was once formerly craved, as a harmful, addictive substance, adopts a new belief about its desirability and so creates a desire to avoid smoking. Such an individual does not merely describe smoking as unhealthy or undesirable, but forms a cognitive commitment not to smoke instead.

Carruthers's Skeptical Theory: Interpretative Sensory Access

Carruthers argues that the system by which one ascribes attitudes to oneself is the same system by which one ascribes attitudes to others. Following the extensive literature on so-called "theory of mind," Carruthers calls this system the mindreading faculty. Carruthers dubs his account of self-knowledge the "interpretative sensory-access theory" or ISA, because he contends that this system only has access to sensory or perceptual information about one's own bodily movements, one's physiological arousal, and one's inner speech. Carruthers' ISA theory has skeptical conclusions because it suggests that our access to our thoughts is always mediated by the mindreading faculty, which lacks direct access to our attitudes.

Allow me to describe the ISA theory in a bit more detail. Carruthers develops his theory in light of well-established empirical theories in the foundations of cognitive science. In particular, he assumes a global broadcast architecture of the human mind, proposed initially by Baars.[10] This means that Carruthers assumes the mind is differentiated into two classes of cognitive systems. There are sensory systems that receive sensory information and broadcast it to an array of consumer systems, which, according to Carruthers, include the mindreading faculty. In Carruthers' words, "human mental life consists of islands of conscious events surrounded by seas of unconscious processing."[11] As a result, there is no special epistemic relation grounding our first-person authority in relation to our own minds.

Carruthers claims his theory explains data gathered by experimental psychologists and neuroscientists on confabulation, or the process of filling in gaps in one's knowledge with false or unjustified judgments. Indeed,

---

[9] Moran, 53.

[10] Bernard Baars, *A Cognitive Theory of Consciousness* (Cambridge: Cambridge University Press, 1988).

[11] Carruthers, *The Opacity of Mind*, 50.

confabulation is a direct obstacle to self-knowledge as traditionally conceived, for it suggests that we lack first-person authority in relation to our states of mind. As Carruthers himself puts it, "[s]ince the [ISA] theory claims that our only access to our thoughts and thought processes is interpretative, relying on sensory, situational, and behavioral cues, there should be frequent instances where presence of misleading data of these sorts leads us to attribute attitudes to ourselves mistakenly."[12]

There is widespread literature on confabulation. Allow me to briefly present a particularly vivid example. Brasil-Neto et al. induced motor movement through direct stimulation of the motor cortex, bypassing decision-making systems of their subjects.[13] They made their subjects wear headphones, and then instructed their subjects to raise their index finger when they heard a clicking noise. However, they sometimes induced this finger movement artificially in their subjects without the presence of audio. Their subjects self-identified these finger movements as intentional decisions to move their fingers. The result of these considerations is that we are prone to misinterpret our actions and misidentify the reasons for which we act. Carruthers takes these examples to entail a strong form of skepticism about our access to our own minds. Indeed, they seem to undermine a constitutivist account of self-knowledge. But below, I will argue that if properly understood, these results can be incorporated into a constitutivist account of self-knowledge.

Responses to Carruthers

The constitutivist agrees with Carruthers that one's self-knowledge of one's state of mind is interpretive. Where Carruthers and the constitutivist disagree is in the implications of interpretation of first-person authority. Carruthers takes the interpretative nature of self-knowledge to undermine first-person authority, while the constitutivist takes the interpretative nature of self-knowledge to be essential for first-person authority.

The constitutivist also agrees with Carruthers that we can be influenced by mental states that work beneath the surface, but the constitutivist does not draw the same skeptical conclusions from this fact. It is important to distinguish mental states of which we have knowledge from mental states of which we are consciously aware.[14] I argue that we have knowledge of certain mental states without necessarily being consciously aware of them, and that this consciousness of our minds constitutes an important type of self-knowledge.

The distinction between knowledge and conscious awareness of states of mind can be articulated by means of an example. First, consider Cam, who is consciously angry at his brother, Matt. Cam experiences the phenomenal heat of anger; he feels a bodily reaction when his brother enters the room, and he is aware of considerations that justify his emotion toward his brother

[12] Ibid., 325.
[13] Joaquim Brasil-Neto et al., "Focal Transcranial Magnetic Stimulation and Response Bias in a Forced Choice Task," *Journal of Neurology, Neurosurgery, and Psychiatry* 55 (1992): 964-6.
[14] Cf. Finkelstein, *Expression and the Inner*, 20.

(for instance, his brother has done something to offend him). Second, consider Ned, who is not consciously angry at his brother. Instead, he may simply notice a tendency to think negative thoughts when his brother is around. Or perhaps their mutual friend tells Ned that his body language becomes aggressive and tense when he describes a meeting with his brother.

Ned considers his stance toward his brother and decides that he is angry at him. Although Ned knows of his anger toward his brother, he does not consciously experience it as anger. Instead, he might interpret his anger as a mere mood or inclination based on external facts about his current situation, for instance, the fact that he skipped breakfast. In other words, he lacks conscious self-knowledge of his state of mind. Nevertheless, his knowledge of his anger, once it is brought to light, opens the door to alternative ways of understanding it. For instance, Ned might come to recognize his behavior toward his brother as manifesting a non-conscious anger toward him, and consequently come to see his actions toward his brother as unwarranted.

Such a realization may not immediately change Ned's dispositions to act in relation to his brother (e.g., he might feel motivated to criticize his brother upon hearing his suggestions), but mere recognition of this aspect of his state of mind, though not directly accessible in some infallible, self-presenting sense, is still a crucial type of self-knowledge for Ned's process of self-constitution. Without consciousness of his anger, he would lack the reflective distance to commit himself to alternative ways of being.

It is not open to Carruthers to deny the existence of knowledge of our mental states in the sense described above. Indeed, Carruthers' own theory is a testament to how knowledge of the workings of our mind is possible, even if we cannot have conscious awareness of these workings. However, Carruthers might argue that the type of knowledge I am describing is unimportant. According to this line of thought, only consciously experienced mental states play a role in decision making.

Such an objection misses the importance of knowledge of our minds and their workings, along with the normative elements of commitment, in our practical reasoning. Speaking on the compatibility of observational and constitutive accounts of self-knowledge, Moran captures these considerations quite nicely:

> Neither perspective denies the truths of the other. The assertion from the Deliberative stance that "I am not bound by my empirical history" is not in any way a denial that the facts of my history are what they are. It does not deny either the truth of these claims or their relevance to the question at hand; but it does deny their completeness and, in a word, their decisiveness.[15]

In other words, Moran argues that knowledge of the empirical reality of our condition, of the ways in which we are prone to err when thinking about ourselves, is essential for coming to a reasoned stance about how we ought

[15] Moran, 163.

to proceed in our lives. Nevertheless, such empirical knowledge cannot fully determine how we ought to proceed. As Moran describes, it is incomplete. We have a responsibility to constitute ourselves, not simply to follow through on some projected trajectory. We are presented with an external world that requires us to make choices and perform actions. The option to simply submit to our habitual dispositions is not really an option. To put it another way, self-knowledge is deliberatively indispensable.

Note that constitutivism about self-knowledge, so conceived, neither requires transparency in the Cartesian sense advocated by Carruthers, nor does it deny interpretation. What is more, it acknowledges that self-knowledge is a genuine cognitive achievement. In other words, self-knowledge must be earned through constant revaluation of one's empirical conditions and one's commitments. But, as Moran points out, a more thorough knowledge of our circumstances need not undermine the core of the constitutivist position.

Conclusion

I have argued that Carruthers's skeptical conclusions about self-knowledge are unwarranted. The constitutivist account of self-knowledge for which I have argued recognizes that our knowledge of ourselves can be difficult to acquire, but it need not be any more inherently problematic than any other kind of knowledge. The essence of constitutivism is the thesis that self-knowledge is a process. Taking this view seriously allows us to appreciate not only the shortcomings, but also—and more importantly—the virtues of self-knowledge and its significance in our lives. ❖