



---

A Century of Deflation and a Moment about Self-Knowledge

Author(s): Tyler Burge

Source: *Proceedings and Addresses of the American Philosophical Association*, Vol. 73, No. 2 (Nov., 1999), pp. 25-46

Published by: American Philosophical Association

Stable URL: <http://www.jstor.org/stable/3131086>

Accessed: 16-05-2018 23:07 UTC

## REFERENCES

Linked references are available on JSTOR for this article:

[http://www.jstor.org/stable/3131086?seq=1&cid=pdf-reference#references\\_tab\\_contents](http://www.jstor.org/stable/3131086?seq=1&cid=pdf-reference#references_tab_contents)

You may need to log in to JSTOR to access the linked references.

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

*American Philosophical Association* is collaborating with JSTOR to digitize, preserve and extend access to *Proceedings and Addresses of the American Philosophical Association*

---

# A CENTURY OF DEFLATION AND A MOMENT ABOUT SELF-KNOWLEDGE<sup>1</sup>

Tyler Burge, University of California Los Angeles

*Given as Presidential Address to the Pacific Division of the American Philosophical Association in Berkeley, April 1999.*

The person who notified me of my selection for this office tried to entice me into making this speech by saying that mine would be the last address of the century, indeed millennium, to this body. Maybe he thought I went in for apocalypses. Or maybe it was a set-up for impeachment. Or maybe he thought that I'd think I'd get the last word. I'm confident that President-Elect Professor Herman will not be speechless next century. Some of the rest of you may have something to say as well. Despite my sense that the occasion lacks quite the drama I was promised, I have succumbed to some temptation to hazard something synoptic.

Our subject got a good start in this century. There were some lasting ideas about language, and ferment over mathematical logic. But early on, philosophy fell into a negative mood that it has had a hard time shaking off. Self-nihilating movements have come in waves. Positivism, Wittgensteinianism, Ordinary Language Philosophy, Naturalism (as in Naturalized Epistemology), and many other isms have tried to show that philosophical problems are the result of some simple oversight. Quine's attack on positivism seemed to clear the air, and did in fact free the subject to develop in a fruitful way. But a self-destructive streak persists.

This streak resonates with a tendency, shared with movements outside of philosophy, to deflate older conceptions of the specialness of human beings—a specialness that philosophy had been wont to explore and extol. It is as if philosophy has recently measured itself and humanity, and found both gravely wanting.

I brought with me a transcript I recently found of an interview between a prospective graduate student who is seeking advice about a career in philosophy from a philosopher—Professor Carwittup. Although

the professor seems to have done nearly all the talking, one can make out a *virtual* dialog:

“Professor, I am thinking about going into philosophy as a way of finding meaning in my life.”

“Oh come off it! There isn’t any meaning. There is only choice among radical translation manuals. Translating the heathen is making up a system of claims we already agree with. For missionaries and philosophers, disagreement and insight are complementary illusions.”

“Sir professor, I want at least to know what *I* am talking about.”

“Reference/schmeference. Your word ‘mama’ refers to a momentary stage of a physical object on one interpretation, and to the number seventeen on another equally good interpretation. You might as well stop being sentimental about your mother.”

“Don’t philosophers pursue the truth? Isn’t truth referring to something and characterizing it as it is?”

“A benighted view: Calling something true is merely paying it a solidarity compliment. It is the compliment of removing quotation marks. A job in editing would better equip you to understand truth.”

“But science and philosophy make progress . . .”

“What an old-fashioned idea. So-called progress is just a series of power-driven paradigm shifts, or aimless redirections of the conversation. Perhaps you should consider becoming a talk-show host.

“You think at least mathematics is a stable science? Rationalist superstition! Mathematics just makes it up like the rest of us. It tells useful stories about fictional objects. Philosophy of math is literary criticism. An English department would train you for that.”

“Maybe I should go into ethics. I would like to understand my reasons for counting actions right or wrong.”

“Give it up! You would just be emoting in an evolutionarily inevitable way. There are schools of acting that would serve your purposes just as well.”

“But Professor, I can’t think of myself as an actor or some biologically programmed gush of emotion. Maybe I should work on the free will problem.”

“Even with close study of the brain, science seems not to have found the will. We now think the will may not exist. But if this result does not hold up, the free will problem and its solutions

---

are unverifiable, hence meaningless. And do not presume that you understand what I just said.”

“Philosophy of mind seems intriguing . . .”

“The mind? Are you kidding? You must mean an automaton, or a quivering blob of grey matter, overlaid with a functionally specifiable grid of causal relationships. You can stop worrying about going out of your mind at least. These are not the sorts of things you can go out of.”

“But there are conferences and magazine articles on consciousness. They seem kind of exciting . . .”

“No, consciousness is for engineers and sociologists to study. Consciousness is just a self-monitoring feedback mechanism, common to guided missiles and Asian nations. Philosophy does show us a new range of things we can be touchy-feely with.”

“You mean missiles and nations can feel things and I can get cozy with them? My friends in California have been saying this all along. I want to get in touch with my own feelings too. Aren’t there things philosophers call ‘qualitative feels’?”

“How quaint of you to ask. In us, pain is just a way of representing neural damage. In the Chinese nation, it is a report that the communist party has been questioned. Pain has no other properties than the news it provides. The pain is bad news, but it’s the damage that is bad. Don’t blame the messenger!

“And pleasure is no better than pain. Orgasm is the internal sentence ‘something pleasurable is going on down there.’ True or false? It doesn’t matter. Sex, too, is a text.

“There *may* be an opportunity for social activism in this theory of orgasm. The theory is a recipe for impotence. By espousing it you might help with the problem of overpopulation.”

“Well if I am going to be a social activist, I don’t want to be lost in the crowd. I want to be able to distinguish myself from others, so I can avoid an identity crisis.”

“The theory of personal identity casts doubt on identity crises. Persons are just fissions or fusions in a continuum of psychological states, impersonally specified and beamable to any receptive material in the universe. A politician could use the fissions. The prospects for voter registration are endless. Maybe politics is your field.”

“But Professor, on such a theory it seems hard to know who is alive and who is dead . . .”

“Philosophy frees us from caring seriously about our deaths. Since we are just functional patterns common to the Asian nations

and quivering blobs, we need not care much whether we live or die. What matters in personal identity is whether our patterns of quiver survive.”

“Sir, Herr Professor, how does one know such strange things?”

“Knowledge and justification have been naturalized: they are just reliable connections to the way things are. A thermostat has it all. Knowledge is going with the flow. Perhaps you should consider a job in heating and air-conditioning.”

“Isn’t there more to it than that? My friends are constantly upsetting me with their sceptical worries.”

“Scepticism is an obtuse failure to recognize standards of discussion appropriate to the context. If the sceptic would just learn some manners, he’d stop misleading the youth. Defusing scepticism is devising a system of etiquette, or keeping score in a game. Have you thought about being a water polo referee?”

“Professor, I wouldn’t be happy blowing whistles; I know myself that well . . .”

“Self-knowledge? Psychological experiments have now shown that we never get anything right about ourselves. But this is a good thing. Ignorance has been selected for. You can best cultivate collective ignorance by advising any number of Congressmen.”

“But those guys are so irrational . . .”

“We all are. We are gloriously irrational animals. Psychological experiments have also shown that our reasoning is entirely made up of fallacies. Like ignorance, irrationality has been selected for. We’d best nourish our irrationality, on pain of getting mowed down by beings crazier than we are. There are some self-help groups in Southern California that might help you become more irrational.

“So you still want to go into philosophy? You need not submit a writing sample. We base admission decisions on a personality test to determine whether you are sufficiently sympathetic with masochism. Philosophy is in the business of showing that it has no business being in business.”

The interview seems to be concluded there.

A few months ago, I was speaking with an accomplished philosopher who had been reading philosophers writing at the end of the nineteenth century. He marveled at the things they would say. They would blandly make the most ridiculous claims, he noted. He championed our enlightened age of precision and rigor, and conjectured that no one nowadays makes such absurd claims. I agree with his observations about the past. But I think that this satisfaction with the present doctrinal

---

state of the profession shows a certain historical blindness to the near. Our age has produced as many absurdities as any other. But a large bulk of ours are overkill in the service of self-immolation, quite a different matter from the wild fictions commonly told by revisionary metaphysicians. The excesses at the ends of the last three centuries were excesses of hubris in philosophy. Leibniz', Fichtel's, and F. H. Bradley's absurdities are the products of self-inflation. It is not just the individual philosopher's self-inflation; that is perennial. Philosophy, and with it humankind, were also seen in the warm glow of self-satisfaction.

Some of this inflation of humankind, which in philosophy commonly takes on some form of idealism, survives in our century in scientized guises. I have in mind various revivals of idealism, which present themselves as antirealism, instrumentalism, and various forms of *so-called* "realism" which aren't realism at all (for example, critical realism or realism with a sweet face). But the idealisms are not the excesses that are distinctive in our century. What is distinctive is the array of simplistic deflations of philosophy and humankind.

The social causes of these parallel deflations are not hard to find. The evolution of science has shown that philosophy can no longer see itself as the ultimate synthesis of everything. The critique of the special status of human beings has a broader array of cultural causes. The rise of psychoanalysis and neuroscience—and to go more basic, the advent of methods of mass destruction and a series of regimes bent on genocide—have made it hard to see human beings as special in any way worth extolling. It is true that they are capable of special evil—private as well as public. Unlike the animals, humans can be shallow, calculating monsters of serial betrayal and destructiveness. But these are not recent discoveries.

Philosophy is not the queen of the sciences. But it need not act as a docile slave. Philosophy *can* do science at a specially abstract level, and it can function as midwife. But it does too many things to fit in those boxes. Philosophy opens lines of inquiry, and occasionally provides knowledge, that no other subject is on to. Some of this knowledge is not, and never will be, systematized in any science. Some of it is an instance of part of what is special and valuable about human beings. A central task of philosophy is to articulate this specialness in ways that avoid the inflations of the past while firmly retaining a perspective on the wonder of humankind and its values, intellectual capacities, and institutions.

I have presumed on your patience and on the license associated with my happily temporary office. I have presented a scatter of unargued attitudes—not good philosophical practice. I have also come

close to joining the ranks of naysayers about philosophy, even in naysaying its naysaying. Before leaving the metaphilosophical mode, I want to say something good about our subject in this century. I believe that we have some things to be proud of.

One is the commitment to clarity and argument. For this we can thank the positivists, Quine, Church, Strawson, Grice, Davidson, Putnam, and Kripke, as well as Frege, Moore, and Russell. Clarity and commitment to argument do not produce creativity, but they aid it. Moreover, they enhance communication. They have nourished the growth of a vital philosophical community. This is an element of science that philosophy has succeeded in borrowing. Let it not go the way of scholasticism, herd journalism, or advertising.

There are substantive achievements as well as methodological ones. A better understanding of science and a concern to integrate philosophy with it is one. Scientific knowledge and modeling have checked metaphysical excess better than in previous centuries. A better understanding of language is another substantive achievement. Philosophy of science and philosophy of language have raised new philosophical issues, transforming and enriching old ones.

Another achievement is an offshoot of the denigration of traditional views of the specialness of humankind. The anti-descriptivist picture of reference, the anti-individualist view of mind, and the recognition of externalist aspects of epistemic warrant, have helped us see how our natures are determined by norms that reach beyond what we as individuals control. We can better understand the ways that rational beings depend on a universe that is not made up of structures of reason at all. These developments should help us understand our specialness, without falling into the idealisms and hyperintellectualizations that distorted early modern philosophy.

One route to understanding what is intellectually special about human beings is through types of knowledge and warrant distinctive of us. By understanding the relevant types of knowledge and warrant, we better understand the intellectual capacities that help ground the warrant. In earlier work I have explored a priori knowledge, knowledge through interlocution, and self-knowledge. Here I want to discuss some issues regarding self-knowledge.

It is empirically clear that higher nonhuman animals have propositional attitudes. They therefore have concepts. I believe that all higher animals have a concept of pain. A primary use of this concept is to recognize when they are in pain. We share such self-knowledge with higher animals. Yet we seem to have types of self-knowledge that are more distinctively ours. One distinctive aspect of our self-knowledge

---

---

concerns our first-person concept. The ape that knows it is in pain attributes the pain egocentrically. Some sort of mental file is associated with the egocentric element in the attribution, and is maintained over time. Higher animals commonly believe that they are in pain when they are.

But I think that they lack a *first-person* concept. This is a way in which some aspects of our self-attributions of pain, not to speak of more sophisticated self-attributions, are distinctive. I shall say more about this concept later.

A second way in which our self-knowledge is distinctive lies, I think, in our *understanding* of propositional attitudes. Higher animals have propositional attitudes. But although it is currently an empirically open question whether higher animals make rudimentary attributions of simple attitudes to other animals, I doubt whether any subhuman animals have concepts of propositional attitudes that include the full range of fundamental inferential connections—particularly connections to normative notions like *reason*—that ours do. Although animals have reasons, they do not understand them as such. I shall develop this point later.

As do several others, I hold that we have a nonperceptual, nonempirical self-knowledge that has a kind of authority. I have developed this view in the context of my anti-individualism about the mind. I have argued that many of our propositional attitudes are constitutively dependent on nonconceptual relations we bear to our physical and social environments.<sup>2</sup> The nonempirical, authoritative warrant for such self-attributions is partly independent of the content of the first-level propositional attitudes. The warrant is independent of environmental variations that could only be known empirically. I shall not defend this position here. I simply take it for granted.

I have also argued that the nature of the warrant has something to do with the fact that the content of the relevant second-level self-attributions cannot come apart from the content of the first-level propositional attitudes. Intuitively the knowledgeable self-attribution *I believe that water is a liquid* includes the content of the first-level belief that *water is a liquid*.<sup>3</sup> I have further argued that the nature of the warrant has something to do with the role that self-attributions play in critical reason.<sup>4</sup> Critical reason is reason that is capable of being (and sometimes is actually) guided by a conception and evaluation of reasons as such. Critical reason is a paradigmatically distinctive activity of human beings. Animals think and have reasons, but they do not think in the self-guiding way that involves the evaluation of reasons as such and the implementation of reason in accord with such evaluations. Thus critical reason necessarily has a second-order aspect. It can be theoretical. But it is also what is present in practical deliberation and in having a conception of

---



a life or life plan. It is part of what is essential to practical and intellectual responsibility.

The argument that the warrant for our self-attributions of certain propositional attitudes has to do with their role in critical reason turns on the existence of a certain immediacy in the rational applicability to first-level reasoning of reasons acknowledged in second-level self-evaluations. For example, if I am *prima facie* warranted in judging that my belief that A is guilty is a poorly grounded belief, then it follows immediately that I am *prima facie* warranted in changing the belief (the belief at the first level). It is rationally immediate that the warrant for a second-level evaluation judgment applies, at least *prima facie*, to the belief being evaluated. This immediacy is not present in evaluations of others' attitudes, or even in certain evaluations of oneself "from the outside." This is so only because there is an immediate rational connection, partly analogous to that of a logical connection, between the second-level judgment about the belief and the belief itself. I theorized that the immediate connection is one within a single point of view. I maintained that this multilevel point of view's being a system of critical reason helps explain the authoritative character of the warrant for self-attributions within it.<sup>5</sup>

I now want to return to what was for me an illuminating exchange with Christopher Peacocke here in the Bay Area four years ago. Peacocke sketched a different conception of the warrant in self-attributions. He agreed that critical reason essentially involves self-attributions that carry a nonempirical warrant. But he held that the source of the warrant lies not in the role of self-attribution in critical reason, but in a prior source. Critical reason is to be explained, on this view, in terms of an independently warranted capacity for self-attributions, together with certain other features.

Authoritative instances of self-knowledge comprise a range of cases.<sup>6</sup> There are *cogito* cases like *I am now thinking that presidential speeches are a peculiar phenomenon*, or *I hereby judge that examples need elaboration*. These self-attributions are either logically or performatively self-verifying. Their warrant clearly depends on understanding, not on any sort of perception. Neither the warrant nor the truth of the judgment depends on any causal relation between judgment and subject-matter. The subject-matter of the judgment and the judgment about it occur in the same act, and in some cases are one and the same.

The point that self-knowledge is not perceptual and has some special warranted status is obvious in these cases. But other types of self-knowledge—some knowledge of standing propositional attitudes, for

---

example—are also authoritative. Presumably these instances of self-knowledge depend on some causal relation between judgment and subject matter.<sup>7</sup> For example, my judgment that *I believe that California is south of Oregon* is about a standing state, which together with some triggering occurrence presumably helps cause the self-attribution.

Peacocke cites two other types of authoritative self-knowledge. In one, I have a conscious memory impression that Napoleon lost at Waterloo. I endorse this impression as indicating that Napoleon did lose at Waterloo. The memory impression and the endorsement cause my self-attributing judgment: *I believe that Napoleon lost at Waterloo*.

In the other type, the process leading to the self-attribution is one of making up one's mind.<sup>8</sup> When I am asked, "Do you intend to stay up late after the speech?" or "Do you believe that you can afford that purchase?" I can answer the question by putting into operation whatever decision-making or belief-forming procedure I have for deciding the answers to the first-level questions—whether to stay up late and whether I can afford the purchase.<sup>9</sup>

Peacocke offers accounts of the entitlements underlying the three types of non*cogito*-like self-attribution. He does not discuss the *cogito*-like cases. He holds that the entitlements mention the causal character of the belief-forming mechanisms. The entitlements are said to consist in the reliability of the mechanisms, and the lack of a certain sort of intervening layer of representation between judgment and subject matter. Let me explain his account in more detail.

With regard to self-attributions of standing attitudes without intervening conscious memory states, the attitude is reliably the "rational cause" of the self-attribution. Peacocke adds that if formation of such self-attributions were not caused by the relevant state, and did not accord with self-attributions based on conscious attitudes (like the conscious memory state), the thinker would abandon that way of forming self-attributions.<sup>10</sup>

In self-attributions "rationally caused" by a conscious memory and the endorsement of the memory, the entitlement consists in the self-attribution's having those causes and in the fact that those causes are sufficient for the truth of the self-attribution.

In the case of the mechanism in which the self-attribution is formed through making up one's mind, the entitlement is supposed to consist in the self-attribution's being "rationally caused" by the same states that cause the thinker to have the first-level state (for example, the intention to stay up late).

A key *explanandum* that I proposed for any account of authoritative self-knowledge is that authoritative self-attributions do not appear to be subject to *brute error*.<sup>11</sup> A brute error is one that occurs when the subject's cognitive systems are functioning well and the subject is performing in a fully warranted way. Brute errors occur in perception, but not, I claimed, in authoritative self-attributions. I think that understanding this feature of self-knowledge helps elicit the special authoritative, nonperceptual character of some of our self-knowledge.

Where the content of first-level attitudes is simply reused in second-level attributions, there is no possibility of error regarding content.<sup>12</sup> But in *noncogito* cases (and indeed, in a sense, in performative-type-*cogito* cases), error is possible about whether one holds the attributed attitude toward the specified content.

Peacocke offers an explanation of the impossibility of brute error in the case of the conscious memory that Napoleon lost at Waterloo. He points out that the memory and its endorsement are sufficient for, indeed constitutive of, one's believing that Napoleon lost at Waterloo. Thus they are sufficient for the truth of the self-attribution they cause. He holds that this constitutes a disanalogy with perceptual knowledge. A perceptual experience is never sufficient for the correctness of a perceptual belief, even if the perceiver is operating well and with optimal warrant. He holds that the lack of a layer of experience between the memory representation and the self-attribution is the key to explaining the impossibility of brute error of self-attributions "rationally caused" by endorsements of the memories. He writes, "For self-ascriptions made in this way, brute error is impossible. It is impossible precisely because, in these psychological self-ascriptions, there is nothing that plays the role that experience plays in genuine observational knowledge of physical objects."<sup>13</sup>

Now I think that all of this is close to being right, as far as it goes. But I do not find the explanation of our immunity to brute error, much less the explanation of the authority of our warrant for self-knowledge, satisfying. The intervening conscious memory case is fixed as one in which the self-attribution is veridical: It is appropriately caused by the state that it attributes. Perceptions that are appropriately caused by the objects or states that they represent are guaranteed (trivially) to be veridical also. Peacocke and I agree that there remains a disanalogy between the self-attribution case and the perceptual case.

What distinguishes this case of self-attribution from cases in which one makes a self-attribution that does not properly connect to the memory trace? We cannot usefully say that in that case there is necessarily malfunction or irrationality. For that is what needs to be explained.

---

It is not specific enough to say, as Peacocke does, that “nothing plays the role that experience plays in genuine observational knowledge of physical objects.” That role is perceptual and is subject to brute error. Granted, but the difference is what needs to be explained. The memory and endorsement are, unlike the objects of perception, causes that provide warrant. We may call them “rational causes.” But again the nature of the rationality needs explanation.

The core of Peacocke’s explanation consists of two points: There is a reliable mechanism that is associated with an attributed conscious state. And there is no intervening layer of representation between the conscious state and the self-attributive judgment it causes.

I think that these points do not suffice to explain what is normatively special or authoritative about the warrant. The role of consciousness of the attributed state in the explanation is perhaps significant in the case of the conscious intervening memory state.<sup>14</sup> I am sceptical, however, about whether appeal to consciousness is *sufficient* to explain what is normatively special in the case.

In the first place, I doubt that it is a sufficiently normative notion to carry the burden of explaining the special normative character of the relevant authoritative warrant. To be explanatory of authoritative access to one’s mental states, the relevant notion of consciousness cannot be explained in terms of access, on pain of circularity. So it must be something like phenomenal consciousness. But phenomenal consciousness together with reliability—apart from some further explanatory supplement—do not seem to me to carry the normative elements needed for explaining the relevant authority.

In the second place, there are conscious states that we self-attribute with no intervening representational state but with no authority. We are conscious of certain of our moods, but not authoritative about their nature. We may even be reliable at getting them right, but not authoritative in our judgments. For example, I may be rather reliable at telling whether my unease constitutes anxiety rather than anger or depression, and I may be conscious of the relevant self-attributed state. It does not follow that my self-attributions of my mood are authoritative. I believe that whether they are authoritative depends on further issues regarding the nature of the relation between the mood state and perspective of the self-attribution. This same sort of point applies to construals of the nature of many conscious propositional attitudes. In many ordinary senses, we can be conscious of a state and reliable at characterizing its real nature, but not authoritative about our characterizations.

Moreover, consciousness does not help with most cases of authoritative self-attribution. It does not seem a *necessary* feature of explanations

of the special authoritative warrant. Normally when one authoritatively self-attributes a standing belief, as when I judge that I believe that California is south of Alaska, the attributed belief is not conscious in any other sense than that one brings it to consciousness in self-attributing it. It is the special rational status of this self-attribution that needs explaining.

Many such beliefs, like many perceptual beliefs that can be authoritatively self-attributed, are not acquired through storing any previous conscious act of judgment. Many of the propositional attitudes (especially many perceptual beliefs and beliefs unconsciously derived from perceptual beliefs) that we acquire and are authoritative about enter our system through no conscious endorsement.<sup>15</sup>

Consciousness plays a central role in understanding mind and even rational deliberation. But I doubt that it is ubiquitous or fine-grained enough to play a central role in explaining all instances of authoritative self-attribution. It does not seem to be sufficient to explain the authority of even those cases in which the self-attributing individual is conscious of the attributed state.

Recall that in the case in which a standing state is directly self-attributed without intervening conscious state, Peacocke adds the condition that the reliable mechanism is such that it would be given up if it came in conflict with other conscious bases for attributing the underlying standing beliefs. This suggests that the judgments about unconscious cases are warranted only because the individual is disposed to make them cohere with cases in which the self-attributed state is already conscious. I think that such a suggestion would be mistaken. The condition that Peacocke adds is not essential to the explanation of the entitlement: The self-attribution of standing beliefs, like my belief that California is south of Alaska, would be just as authoritative if I had *no* conscious, separated memories, or other conscious bases for attributing such beliefs, to check the attributions against, and *no* inclination to let the attributions of unconscious states be affected by attributions based on conscious ones.

Perhaps the *self-attribution* must be conscious to be authoritative. Perhaps for a self-attribution to be authoritative, it must be possible to make the attributed state conscious, in the sense that it must be susceptible to being consciously self-attributed.<sup>16</sup> But I do not see that the self-attribution need rely on any further conscious states, actual or potential, to make it warranted or authoritative. A disposition by the individual to require coherence with such conscious bases is not an essential factor in the authoritative warrant (though absence of actual incoherence with such conscious bases, may be essential). The case of intervening conscious memory traces is, I think, relatively unusual, not the central type.

---

What I think is important to the warrant is not intervening conscious memory traces—actual or potential. It is the role that the attributed states play in the rational system in which the attribution is carried out. Also important is the individual's ability to withdraw commitment to memories on the basis of understood rational norms. I shall elaborate these points shortly.

There remains Peacocke's point that the mechanism is reliable, and that there is no intervening layer of representation between the first-level state and the self-attribution it causes. Normal reliability of a causal route to self-attribution is surely necessary but not sufficient to the authoritative warrant.<sup>17</sup>

So what of the lack of an intervening layer of representation? This lack does not, I think, suffice to explain the authoritativeness of the warrant or the immunity to brute error either.

It must be remembered that there *is* a layer of representation, present in the attribution but not present in the attributed state. The conceptual analog of the that-clause in authoritative self-attributions redeploys a content already present in the attributed state. But the representation of the type of propositional attitude that is attributed (e.g., that it is a belief) and the attribution of the attitude *in the first-person way* are additional layers of representation. The former, the representation of the type of propositional attitude, can go awry. Moreover, we authoritatively redeploy the content of the first-order attributed state only when we are correctly and authoritatively related to the state. It is certainly possible to get the contents of some of our attitudes wrong—to misrepresent them, in *noncogito*-like self-attributions. So except in pure *cogito* cases, redeployment is not necessitated by the form of the attributions, but is an aspect of the authority. So it too needs to be explained.

I believe that brute error, hence lack of authority, is compatible with high reliability and lack of intervening layer of representation. Suppose that we were wired through evolution so that in attributing standing states, we were able about 98 percent of the time—entirely unreflectively and without going through any procedure—to correctly self-attribute otherwise unconscious states. Suppose that these attributed states are counterparts to states that in our actual psychologies are modular.<sup>18</sup> Suppose that we were unable to control or integrate these early-perceptual states any better than we can in the actual circumstances. Suppose that we were also wired to get our mothers' propositional attitude states right about 96 percent of the time. Suppose that these are natural, evolved cognitive capacities. I stipulate that in neither case is there an intervening layer of representation. (Some extra-sensory causal vibrations put us on to our mother's states.) Suppose

that mistakes occur because a “wrong” standing state under certain conditions triggers the causal mechanism in a way that is indiscernible from the inside. Such errors do not necessitate any malfunction of our capacity, certainly not in discerning one’s mother’s states. The reliable mechanism is only *so* reliable because of interfering conditions external to its operation. No failure of warrant need occur when we make mistakes. I believe that brute error would clearly occur in attributions of one’s mother’s states, and might well occur in attributions of one’s own.<sup>19</sup>

The source of our authoritative warrant in actual self-attributions cannot be fully explained in terms of reliability together with an absence of representational layer. These mechanistic and reliabilist resources are *part* of the account of our warrant for *some* authoritative attributions. But they do not come to grips with the essential matter. One must say more about the epistemic norms governing the transitions between the relevant intentional states.<sup>20</sup> To prepare development of this point, I turn to Peacocke’s primary criticism of my position. He agrees that anyone who has critical reason must have an authoritative warrant for self-attributions. But he holds that the role in a system of critical reason is not a source of such a warrant.

Peacocke’s own view is that critical reason is the result of combining two more primitive capacities, which are warranted independently of their role in critical reason. One capacity is thought which is guided by evidence, but which involves no assessment of beliefs or reasons for belief as such.<sup>21</sup> The other capacity is the ability to self-attribute beliefs. The warrant for the latter is supposed to be independently explained by the mechanisms that we have already discussed. I have maintained that these mechanisms are insufficient to explain the authoritative nature of the relevant warrants. They are insufficient because they do not touch the nature of the rational norms governing mature, authoritative self-attribution.

I believe that Peacocke is pressing an important question about the source of our warrant for authoritative self-knowledge. There are levels of self-attribution, associated with different levels of cognitive ability. Underlying all levels is the reflexive ego-centric sensitivity present in the senses of bodily awareness, touch, and spatially oriented vision, that is the primordial base of self-consciousness. This sensitivity is present even in lower animals. It is developmentally prior to propositional thought. It is certainly prior to the conceptual abilities appropriately associated with the self-consciousness of persons. As I indicated earlier, higher animals have beliefs about their pains. This is a propositional self-attribution, involving some sort of egocentric self-representation by the animal. I am sure that the epistemology of this sort of self-attribution is special. Perhaps it is free from brute error.<sup>22</sup> Presumably the causal

---

mechanisms involved are analogous to mechanisms involved in us. Such mechanisms may provide the primitive basis for the reliability of our own self-attributions. But the animal's self-attribution of pain lacks the authoritative warrant that I am pursuing.

Perhaps at a yet higher stage—but before understanding the concept of warrant for attitudes—higher animals and young children have rudimentary propositional attitude concepts that play a role in the anticipation of others' behavior. Perhaps they even have an analog of a representation of false belief. But I believe that they lack an ability that is central to a mature mastery of our concept of belief. Such mastery involves a rudimentary conception of good and bad warrant for belief—some ability to think *I believe it because*, where there follows a specification of not *merely* a cause, but a specification of a reason backed by some conception of norms of reason.

The chimp cannot come to doubt its belief that it is in pain by thinking that a reason to doubt is that the sensation was caused by the sudden surprising application of ice, rather than by a cut. The chimp might perhaps retrospectively change its former belief (so that it does not affect memory). Such a change would be rational, but it would not be on the basis of an acknowledgement of reasons as such.

An understanding that extended only as far as the functional role of beliefs in being caused by other matters and in causing behavior, even in oneself, would not, I think, be a mature mastery of our concept of belief. One needs a rudimentary understanding of norms governing beliefs and inferences, a conceptualization of warrant—whether the warrant is entitlement or reason.<sup>23</sup>

In my view, the authority of the warrant for mature self-attributions of belief by persons lies partly in the role such attributions play in a system of conceptualized reasoned revision of propositional attitudes. It lies in the adjudication of error according to *understood* norms for confronting error. I believe that knowing what a reason is, and understanding normative grounds for belief revision (not merely thinking and changing attitudes in accord with rational evidential procedures) are so basic to understanding belief that these abilities are criterial for having a full or mature concept of belief (or other propositional attitude). So mature self-attribution of belief requires understanding the concept of belief. Such understanding requires not only reliable unreasoned self-attribution but having the capabilities integral to critical reason. The authority of our warrant for mature self-attributions of belief is associated with this conceptual mastery, and more basically with the ability to utilize self-attributions as baselines for the application of critical reason. Authoritative mature self-attributions cannot precede critical reason.



This point about relations between the concept of belief, critical reason, and the authority of our warrant in self-attribution is associated with a further point about abilities associated with the first-person concept. The further point seems to me more strongly evident. Having the first-person concept that expresses the self-consciousness of persons and is associated with having the relevant authoritative warrant itself requires having a concept of a warrant or reason for propositional attitudes.

Young children, and old people in certain stages of dementia, are persons even though they lack critical reason. But they are persons partly by virtue of having a history of or potential for critical rational abilities. The authoritative self-attributions whose warrant we are discussing are the self-attributions of mature persons. I assume that at some stage beyond what higher animals can do, but prior to having the concept of warrant for a belief, a child can, by understanding language, understand a prototype of the full first-person concept. But the self-consciousness of persons goes beyond marking an egocentric file associated with bodily sensitivity, or even mental agency. The self-attribution of mature persons requires that one have the intellectual abilities fundamental to being a person.

These abilities are so basic to marking an important kind that it seems to me that they should be taken as criterial for having the first-person concept that is fully expressive of personhood. Having such a concept and having the capacities that are central to being a person go together. Both require having a concept of a system of evaluable beliefs—of a point of view—held together with reasons and subject to revision under the force of reasons. They also require having some conception of rational plans that span nontrivial time periods.

I have argued elsewhere that what is central and unique to our first-person concept is that it marks an ability and disposition to acknowledge the immediate relevance of reasons to implementation of the reasons in thought or action—and a disposition to actually implement reasons where appropriate.<sup>24</sup> I believe further that having the full first-person concept involves having a rudimentary understanding of its function. It requires not merely understanding the rule that the concept applies to the author of the thought containing it. It requires a rudimentary ability to reason critically—applying reasons and recognizing through first-person acknowledgement of beliefs and plans that reasons governing one's beliefs and plans are immediately relevant to changing or maintaining them. This ability requires having the concepts of propositional attitude and reason, and an ability to acknowledge and implement reasons when they are relevant to guiding thought. Such abilities carry warrants associated with critical reason. So utilizing the first-person concept that marks the abilities and self-consciousness

---

of mature persons cannot precede the ability to reason critically. So self-attributions using this concept cannot precede the ability to reason critically.

I think that this fact is connected with our recognition that certain first-person self-attributions are authoritative—not merely immediate and highly reliable. More basically, the fact that the intellectual abilities definitive of critical reason are necessarily associated with mature self-attributions by persons underlies the authoritative nature of the self-attributions. In my view, attributions involving subpersonal ego-centric sensitivities or concepts, though warranted—perhaps even epistemically special, perhaps even free of brute error<sup>25</sup>—lack the authoritative warrant that marks mature persons' self-attributions.

Self-attributions by mature persons have a normative authority that is associated with both understanding rational procedures for dealing with error and with understanding one's own responsibility to rational norms in the implementation of reasons. The authority is in large measure epistemic, but it is associated with the inviolacy and respect owed to persons. Acknowledging in the first-person way that attitudes are one's own is fitting them into a structure of reflective, critical reason that helps constitute personhood.

An account of epistemic warrant should explain why a warranted attitude is in normal circumstances a reliable guide to truth. This is why at certain levels of explanation and in some cases, it is correct, as Peacocke emphasizes, to mention causal mechanisms that mediate first-level mental states and some authoritative self-attributions. But an account of the authority of self-attributions must explain what is common to both causally-based and non-causal, *cogito*-like cases.

Transitions between first-level propositional attitudes and self-attributions provide basic links in a critically rational structure that help constitute our self-concepts, and indeed our personhood. The ability to move freely, rationally, and immediately, from first-level attitudes to second-level self-attributions of them, and back, is an element in self-conscious critical reason.

The authoritative status of warrants in self-attribution presupposes, for their being warrants at all, a capacity to make reliable self-attributions. But the warrants' authority depends on the role of self-attribution in critical reason. This dependence is explained by the fact that our mature first-person concept—and, I think, our mature propositional attitude concepts—are constitutively associated with critical reason.

More basically, the warrant's authority depends on critical reason because that status derives from two aspects of critical reason.<sup>26</sup> It derives, first, from our using self-attributions as a baseline for applying

understood rational procedures to deal with error. One's status as a judge of one's mental states is authoritative partly because in addition to being reliable, *one knows the law*. One understands procedures for adjudicating difficult cases, even including self-attributions.

The authority derives, secondly, from the role of our acknowledging responsibility to rational norms in constituting our special status as persons. The employment of the mature first-person concept in its paradigmatic function of acknowledging intellectual and practical responsibility for thoughts and reasons generates the authoritative status of the warrants that we have been discussing. The self-attributions are authoritative partly because they help constitute our concept of ourselves as *persons*, as rational, deliberative beings. This concept in turn is fixed by its function in acknowledging intellectual and practical responsibility for thoughts and reasons.

A reliable mechanism of self-attribution in a being that lacked any conception of reasons for beliefs could be warranted. The warrant might be interestingly different from the warrant for perceptual beliefs. But it would lack the authority of mature self-attributions by persons because it would not be an epistemic constituent in personhood and critical reason. To play that role, the attributions must be associated with the full first-person concept, or with the abilities I associated with such a concept. Being part of a system that underwrites the transfer of reasons immediately between first- and second-levels within a critically rational point of view is part of what is epistemically distinctive about authoritative self-attributions. Acknowledgement of one's attitudes is authoritative because the acknowledgement is essential to a system that constitutes one's personhood.

This account of the authority of mature self-attributions (which I do not claim to be exhaustive) helps explain why self-knowledge has never been a serious target for scepticism about either the truth of or the warrant for self-attributions of mental states and events. The explanation does not derive purely from the presence of a reliable mechanism without an intermediate perceptual layer of representation. For the reliability of the fallible *noncogito* self-attributions would itself be a potential target for scepticism. I believe that scepticism has tended to avoid targeting self-knowledge for three reasons: One is the sheer obviousness of the relevant judgments. A second is the fact that warrant for them is implicit in mastery of the first-person concept and propositional-attitude concepts, used in the relevant self-attributions. But I think the central reason is our implicit ability to connect the warrant for the self-attributions to critical reason. Sceptics must acknowledge critical reason to urge scepticism as a reasonable position. A theory that

---

appeals only to reliability and consciousness, and to the lack of intervening perceptual representations, does not answer sceptical worries about whether the causes of our self-attributions are truly rational, or warrant-giving, causes. It does not call on all the resources that we have in explaining our confidence that scepticism about self-knowledge is rationally self-defeating.

What I have said here is too cursory to do justice to these complex issues. I wanted just to illustrate the kind of problem about the specialness of human beings that I think philosophy can confront and make progress on. You will be glad to know that I shall postpone further efforts in this direction, and that you will get to dinner before the century expires.

## NOTES

1. Given as Presidential Address to the Pacific Division of the American Philosophical Association in Berkeley, April 1999. I would like to thank Ned Block, Dorli Burge, Paul Coppock, Deborah Brown, Chris Peacocke, and Seana Shiffrin for discussions that improved various aspects of this paper.
2. Tyler Burge, "Individualism and the Mental," *Midwest Studies in Philosophy* 4 (1979), pp. 73–121; "Other Bodies," in *Thought and Object*, ed. Woodfield (New York, Oxford University Press, 1982); "Individualism and Psychology," *The Philosophical Review* 95 (1986), pp. 3–45; "Cartesian Error and the Objectivity of Perception," in *Subject, Thought, and Context*, eds. Pettit and McDowell (New York, Oxford University Press, 1986); "Intellectual Norms and Foundations of Mind," *Journal of Philosophy* 83 (1986), pp. 697–720; "Perceptual Individualism and Authoritative Self-Knowledge" in *Contents of Thought*, eds. Grimm and Merrill (Tucson, Arizona University Press, 1988); "Individualism and Self-Knowledge," *Journal of Philosophy* 85 (1988), pp. 649–663; "Our Entitlement to Self-Knowledge," *Proceedings of the Aristotelian Society* 96 (1996), pp. 91–116.
3. "Individualism and Self-Knowledge," op. cit.
4. "Our Entitlement to Self-Knowledge," op. cit.
5. *Ibid.*
6. This is something that I have noted from my first writings on the subject. Cf. "Individualism and Self-Knowledge," op. cit., pp. 649, 658, 663 n11. There is another range of cases involving knowledge of one's nonpropositional states—sensations, or perceptual representations, for example—that does not fit directly into the picture I have just outlined. I believe that this sort of knowledge also gets its authority from the role of such states in a point of view. But the relation to critical reason is more complex. I hope to develop these aspects of self-knowledge on another occasion.
7. Peacocke argues strongly and plausibly for the role of causation in some of the entitlements. Cf. Christopher Peacocke, "Our Entitlement to Self-Knowledge: Entitlement, Self-Knowledge and Conceptual Redeployment," *Proceedings of the Aristotelian Society* 96 (1996), pp. 117–158, esp. sections I and II. His arguments may give the impression that I deny such a role. But I only deny the role of causation in *cogito*-like cases. I maintained

that in none of the authoritative cases (including nonbasic, *noncogito* cases) does the entitlement depend on a causal-*perceptual* mechanism. The presence of “perceptual” is crucial. I cited the case of *noncogito* attributions of standing states and indicated that they were authoritative as well. I grant that there is a causal element in some of these self attributions, and in correct formulations of the entitlement, at least at some levels of abstraction. Perception, I argued, is a poor paradigm for *any* authoritative self-knowledge because it is subject to brute error and is in a certain sense impersonal, whereas authoritative self-knowledge lacks both of these features. Cf. “Individualism and Self-Knowledge,” *op. cit.*, pp. 657–659; “Our Entitlement to Self-Knowledge,” *op. cit.* pp. 91–93, 106 n12.

There are yet other cases which are authoritative but not *cogito*-like, in the sense that they are not performatively or logically self-verifying. For example, in *I am wondering whether I should lengthen this footnote*, I may be tracking an occurrent wondering that is conceived as independent of the second-level judgment. Then the judgment is not self-verifying, and is not *cogito*-like. The second-level judgment need not constitute the wondering; it does not make itself true. One could certainly mistake the modality—thinking one is wondering, when one has already decided. But judgments of this sort can still be authoritative, in something like the way judgments about standing states are. (Whether they are authoritative depends on more than their form.) I am indebted to Ned Block for remarks that led to this point.

8. This type of case was originally emphasized by Gareth Evans in *Varieties of Reference*, ed. McDowell (Oxford, Oxford University Press, 1982), chapter 7. Such cases have been usefully discussed by Richard Moran, “Making Up Your Mind: Self-Interpretation and Self-Constitution,” *Ratio* (1988), pp. 135–151.

9. This type of case may or may not differ from the *cogito*-like cases. In many instances, one will reason about the first-level question and conclude with a performative-like self-attribution: *I do (in this very act) form the intention to stay up or I (now) think that I can afford it*. In fact, as so far described, there is not here a single type of case. The making-up-one’s-mind example is commonly represented as following a conscious procedure. Peacocke describes it this way: “When you are asked ‘Do you intend to go to next year’s Joint Session?’ you may be considering that question for the first time. You can answer the question by putting into operation whatever procedure you have for deciding whether to go to next year’s Joint Session, and answering ‘I do intend to go’ if and only if you do then decide to go.” Cf. Peacocke, *op. cit.*, pp. 121–122. But to answer the explicit question whether one intends to stay up late by following the procedure “answer that one does so intend if and only if one decides to stay up late,” one must know whether one decides to stay up—which is exactly the matter that is supposed to be illuminated. To insure that making-up-one’s-mind examples constitute a new type of case, it must be stipulated that one follows the procedure unconsciously and one is not making the self-attribution in a *cogito*-like or performative way. That is, the second-level judgment is not in any way self-verifying, or constitutive of the first-level attitude, but is fallible and conceived as independent; yet it does not regard the first-level judgment from an outsider’s point of view.

10. I think that this ability to check against other methods of self-attribution is quite inessential to the authority of our primary warrant for self-knowledge. As I shall note below, conscious representations of the self-attributed attitudes are usually not available.

11. “Individualism and Self-Knowledge” *op. cit.*, pp. 657ff.

12. Cf. "Individualism and Self-Knowledge" op. cit. and Peacocke, "Our Entitlement to Self-Knowledge: Entitlement, Self-Knowledge, and Conceptual Redeployment" op. cit.

13. Here is a fuller rendering of his account: "If one's conscious memory representation were somehow made available to one only through a further layer of experience of the memory, for which the question of the veridicality of the experience in respect of the memory arises, then there would apparently be a possibility of brute error. But . . . the conscious memory representation is already, as a conscious state, something capable of giving reasons for forming beliefs. No further layer of experience of the memory exists, and none is necessary. In the presence of a willingness to take its deliverances at face value, a conscious memory representation can make reasonable self-ascriptions of attitudes. . . . For self-ascriptions made in this way, brute error is impossible. It is impossible precisely because, in these psychological self-ascriptions, there is nothing that plays the role that experience plays in genuine observational knowledge of physical objects." Peacocke, op. cit., pp. 125–126.

14. In conversation Peacocke has informed me that the feature of consciousness is fundamental to his view. I am quite sure that the issues involved are more complex than the remarks that follow may suggest.

15. There is a psychological literature that supports this common sense point: Daniel T. Gilbert, "How Mental Systems Believe," *American Psychologist* 46 (1991), pp. 107–119; "The Assent of Man: Mental Representation and the Control of Belief" in *Handbook of Mental Control*, eds. Wegner and Pennebaker (Englewood Cliffs, NJ, Prentice Hall, 1993); Gilbert, Krull, Malone, "Unbelieving the Unbelievable: Some Problems in the Rejection of False Information," *Journal of Personality and Social Psychology* 59 (1990), pp. 601–613.

16. Cf. Immanuel Kant, *Critique of Pure Reason*, B131.

17. One might allow that a person remains warranted and authoritative in self-attributions, at least temporarily, if some diseased memory mechanism that normally funds the self-attributions begins to slip into unreliability. In normal functioning individuals, however, the causal route must be reliable in normal circumstances if it is to be warranted—if it is to carry the relevant entitlement.

18. I do not assume that the reliability has to be this high, or that it has to have any particular number attached to it. As I have noted, I am open to allowing special circumstances when individuals are unreliable but still warranted in their attributions.

19. It does not matter whether one is warranted in relying on the mechanisms, although I am inclined to think that one could be. Peacocke agrees about the cases just described and would take his stand on the consciousness provision that I discussed previously.

20. I think that the presence of additional representative layers also does not entail vulnerability to brute error. To requote a passage from Peacocke: "If one's conscious memory representation were somehow made available to one only through a further layer of experience of the memory, for which the question of the veridicality of the experience in respect of the memory arises, then there would apparently be a possibility of brute error." Cf. Peacocke, op. cit., pp. 125–126. The qualification "for which the question of the veridicality of the experience in respect of the memory arises" is bearing all the weight. For intervening memory representations do play an actual psychological role in our reasoning in long proofs. Such supplemental representational layers do not entail vulnerability to brute error in preservative memory. Everything depends on their normative role in reasoning. If

they play the role of perceptual experience of the memory—taking the memory as a perceptual object—then of course brute error is possible. Perhaps Peacocke is alluding to this point in the phrase “for which the question of the veridicality of the *experience* in respect of memory arises.” The reference to experience and the point that it is an open question whether the experience is veridical allude to a difference in normative or rational role. But the difference with experience is the starting point of an explanation, not the end point. And what is meant by a question of veridicality arising needs to be developed.

Peacocke comes closer to what I think is needed in an almost parenthetical paragraph in which he notes the possibility of holding that necessarily if one has the concept of belief and a proper mastery of the first person, then if the question arises, one will judge that one has the relevant first-order belief when one has the standing belief or when one endorses a conscious memory. *Ibid.*, pp. 126–127. The issue here concerns “the relevant beliefs” in the standing cases. In some cases we are authoritative; in others we are not. I think that the differences cannot be explained purely in terms of relative reliability of correct self-attributions of conscious states. They need to be explained in terms of the special authority that exempts us from brute error. But I think that Peacocke is on to something important in connecting the attributions to conceptual understanding of belief and the first person.

21. *Ibid.*, pp. 127–131. Peacocke calls this “second-tier thought” because it involves thought about relations of support, evidence, or consequence between contents. I think that this sort of thought is an abstraction from thought about norms governing propositional attitudes. Children do not have the concept of intentional content (or of truth, evidence, and consequence) independently of the concept of propositional attitudes. Peacocke’s cases of someone engaging in belief revision at a stage prior to the acquisition of propositional attitude concepts are, in my view, improperly described, and should be redescribed without assuming that the person conceives of evidence as evidence. Such persons are sensitive to evidence in belief revision without conceiving it as evidence for propositions or for beliefs.

22. Immunity to brute error is, in any case, necessary but not sufficient for the relevant authoritative warrant for self-attribution.

23. One cannot understand warrants (entitlements or reasons) without understanding reasons. My term “justification,” which I often contrast with “entitlement,” is to be understood in terms of reasons.

24. Tyler Burge, “Reason and the First-Person” in *Knowing Our Own Minds*, eds. Wright, Smith, Macdonald (Oxford, Clarendon Press, 1998).

25. As noted earlier, I believe that freedom from brute error, though a necessary feature of the authoritative warrant of mature self attributions, is not sufficient. Of course, certain types of reasoning that are not pieces of self-knowledge at all are immune to brute error—deductive reasoning for example.

26. I say “more basically” because I think that it is less important to the argument exactly what the boundaries of the mature first-person concept and mature propositional attitude concepts are. What is important is that these concepts are related in their mature use to the relevant aspects of critical reason. These aspects are what fund the authoritative character of the relevant warrants.

---