

A Kantian Critique of Current Approaches to Self-Knowledge 2: Transparency and First Person/Third Person Asymmetry

Patricia Kitcher

Columbia University

1. Implications of Kantian Unity for Transparency

The implications of Kant's account of higher cognition are as direct for the 'transparency' approach to self-knowledge as they are for the view that self-ascription can be addressed without consideration of mental unity. In brief, the transparency approach assumes that knowledge of the 'inner' is grounded solely in knowledge of the 'outer.' We can see the implications of Kant's theory as soon as we try to use Evans's strategy for the self-ascription of belief. According to Evans, someone, Franz, for example, can self-ascribe a belief about the likelihood of a third world war by first putting into operation his procedure for answering the question: Will there be a third world war?

Since knowledge of such a proposition would be indirect, we will be looking at a case of inferring rather than judging, but similar points would apply in cases of judging. Suppose that Franz reasons as follows:

Wars are brought about by the preparation for war.

Major powers are now disarming.

Therefore, a third world war is unlikely.

Franz forms this belief by looking outward, by remembering the correlation between war and levels of armaments and by considering the current level of arms. He cannot, however, just be looking outward. The act of inferring through which he derives the truth of the conclusion from the truth of the premises informs him that the cognition that a third world war is unlikely depends on other cognitions. It is necessarily connected to them and would not exist without them.¹

Further, this is not an exercise in reflective or critical reasoning (Burge, 1996, 98-99). Like the Christmas shopper who learns the price through seeing a Euro symbol and numerals on a sign, Franz draws the conclusion that a third world war is unlikely and recognizes the cognition(s) through which he has that indirect cognition in a single

mental act. Thus, in inferring that a third world war is unlikely, he knows things about the mental as well as about the external world: The judgment that a third world war is unlikely did not arrive as an inspiration from the gods. It was produced by combining other cognitions.

In the third talk, I tried to reply to the objection that Kant confused a philosophical theory of cognition with an account of the psychological capacities required for cognition. It may be tempting to try to defend transparency by raising a special case of that objection: Although epistemologists distinguish between immediate and mediate (inferred) cognition and psychologists know that mental processing is required to make inferences, cognizers can make inferences without taking account of any mental conditions. All that Franz needs to come to the belief that a third world war is unlikely is knowledge of the correlation and of the facts and the ability to appreciate the logical relations between the premises and the conclusion. Since the correlation, the facts, and the logical relations are independent of Franz's mind, he can form the belief just by looking 'outward.'

To see why this cannot be right, let's examine Evans's case in more detail. What are the procedures that Franz puts into place to figure out whether there will be a third world war? Although Evans suggests deliberately setting up procedures, cognizers often acquire inferred knowledge without having tried to do so. They simply hear some claims and make the obvious inferences. But what conditions must these procedures meet to produce something that counts as an (inferred) belief, a belief that can then be self-ascribed? It would not be enough for Franz to have inputs containing the information in the premises and to output the information contained in the conclusion. That could happen unconsciously and would not be a case of (inferred) belief. Suppose that we add that Franz is conscious of the premises and conscious of the conclusion, and even that his mind *moves* from the premises to the conclusion—but that he is not conscious in inferring the conclusion from the premises.

In 'What the Tortoise Said to Achilles,' Lewis Carroll argued that reasoning requires more than premises and inference rules. It requires actual inferring, mental movement from premises to conclusion (1895). He presents the tortoise as accepting the premises

of an argument, 'If P, then Q' and 'P,' but then as denying the conclusion, 'Q'—because his mind did not move. On Kant's view, mental movement is still not sufficient for inferred belief. Taking up the role of the tortoise, Franz could say that he is conscious of the premises and that, when he is conscious of the premises, he quickly becomes conscious of the conclusion (because his mind does move), yet deny that he sees any relation between the conclusion and the premises. That denial seems as defensible as the original tortoise's denial. Someone who infers must not only have a capacity to recognize logical relations, he must exercise the capacity. Whatever psychological processes he goes through, he must be able to see the relation between the conclusion and the premises in the actual case. But, as we have just seen, that is not possible if the mental movement is unconscious. It is not possible if he blindly moves from some representations to another representation.

If Franz is conscious in drawing the conclusion from the premises, then he recognizes that his belief that C depends on his beliefs that P_1 and P_2 , and thus that his mental states are necessarily connected (though he wouldn't express the point in these philosophical terms). In this way, he also knows that C is true because P_1 and P_2 are true. There are other ways for Franz to discover the relation between the truth of P_1 and P_2 and the truth of C. He could set up a formal proof; he could construct truth tables. In the ordinary case that Evans seems to have in mind, however, he discovers the relation between these claims simply through inferring C from P_1 and P_2 . And in that case, his recognition of the relation between P_1 and P_2 and C depends on his recognition of a relation of necessary connection among his mental states. Since recognizing the relation between the truth of the conclusion and the truth of the premises is necessary for inferred cognition, his recognition of a relation of necessary connection among his mental states is also necessary for inferred cognition. Thus, Franz's belief that a third world war is unlikely involves mental as well as external factors.

A defender of transparency might reply that the Kantian position should be rejected, because it is circular or it leads to an infinite regress. If Franz believes that *his beliefs* in the correlation and in the facts produce his belief that it is unlikely that there will be a third world war, then how did he self-ascribe the beliefs through which he knows the

target belief? Evans tried to explain the self-ascription of beliefs; the Kantian critique of his approach seems to presuppose what Evans was trying to explain.

I think the Kantian position can be defended against the charge of circularity. Suppose that Franz remembers the correlation, learns about the level of arms, and concludes that a third world war is unlikely. He does not have to pre-identify the mental states that are the premises as 'his;' he does not have to pre-identify them as 'beliefs.' He simply has various mental states and consciously draws the conclusion from them, thus creating and recognizing the relation of necessary connection among them. On Kant's view, inferred cognition does not require the pre-identification of beliefs as 'beliefs.' It requires the ability to combine mental states and to recognize the necessary connection between the 'combination' and the mental states that are 'combined' in it as such. It requires the 'post-thinking' identification of beliefs as products and components of thinking, as thoughts that stand in a relation of necessary connection to other thoughts.

Kant gives the functionally characterized ability to combine mental states and to recognize them as combined the philosophically informative label, 'transcendental apperception.' But he offers no hypotheses about the actual psychological faculties that realize or underlie the functionally described capacity. He also maintains that cognizers have an *a priori* concept of an 'I-think,' the concept of a subject whose states stand in the relation of necessary connection. This is not a psychological hypothesis. It is a logical implication from the fact that the senses cannot provide evidence of necessary connections. If humans have the concept of a cognitive subject whose states are necessarily connected (and Kant thinks this can be shown) and they could not have acquired that concept from the senses—*a posteriori*—then it must be *a priori*.

Many capacities make Franz's inferred belief possible, most obviously, perhaps, the capacity for language. Kant's point is that more is needed for inferred cognition. Franz must not only have the ability to recognize logical relations, he must use that capacity. And in the case of ordinary inference, that means that he must also have the capacity to recognize a relation of necessary connection across his mental states as such—when that necessary connection has been made through acts of thinking. Given that capacity

and his *a priori* concept, he can recognize his mental states as the states of an 'I-think' and so recognize them as thoughts or beliefs.

Franz is able to recognize his beliefs as beliefs through thinking, through creating and recognizing relations of necessary connection among his mental states. Kant presents the concept of a thinker as the concept of an 'I-think.' That concept is (of course) general, not singular. As noted in the last talk, on a Kantian view (but not Kant's view), Franz could self-ascribe his belief about the third world war through an entitlement principle:

When a cognizer (consciously) performs an act of higher cognition A that produces a state B, then he is entitled to self-ascribe A, and so B.

At this point we can make the principle first more explicit and then more focused. The principle can be understood as having two parts:

1. When a cognizer (consciously) performs an act of higher cognition A that produces a state B, she recognizes the act as the act of thinking, and its products and components as thoughts or beliefs.
2. When the cognizer recognizes what is in fact her thinking as thinking, and its product and components as thoughts, then she can self-ascribe the thinking and the thoughts.

It follows from the argument just given that when a cognizer performs an act of higher cognition—when she thinks—she recognizes the act as thinking. Although step one is thus always correct, it is not directly related to self-ascription. So a more specific Kantian self-ascription principle would just be part 2. In order to avoid obvious circularity, the principle needs to be expressed carefully: the cognizer recognizes what are in fact, her acts, as acts of thinking; she does not recognize *her* acts of thinking. Even though a subject is not conscious that *she* is thinking, neither is he conscious that someone is thinking or that thinking is going on. One might object that it is impossible to recognize an activity without recognizing the agent. If that is true, then the principle would be useless. This is, however, just what is different about mental activity. Thinkers can recognize what are, in fact, their acts of thinking without first having identified an agent.

In some ways, circularity is avoided on Kant's position in the same way that it is avoided on Evans's. It is not the belief that you have a thought that produces other thoughts, but combining representations or thinking that produces the thought that is the conclusion. Further, self-ascription is not based on a special kind of perception. It operates through an entitlement principle that reflects the fact that self-identification does not rest on any evidence. Still Kant's view differs from Evans's, because he maintains that thinking involves recognizing that activity and its products and components as such, as acts of thinking and as thoughts or beliefs.

We can also see the plausibility of Kant's view that thinking involves recognizing thinking as such from a different perspective. I have noted many times that Kant takes higher cognition to involve acts, but I have not considered why he takes thinking to be acting. I should note that the argument I'm about to present will not be persuasive to those who deny that thinking as a kind of acting. But for those who do, it is important to consider why thinking is acting. Henry Allison has argued that, for Kant, thinking is acting, because it is free (1990, 36, 2006, 389). I don't think this is the way to understand Kant, and the examples already given indicate why it would be an implausible account of thinking. Humans are not 'free' to refrain from making simple inferences any more than they are free not to hear their own names when the name is spoken. It seems equally implausible to claim that thinking must be understood as acting because it is produced by intentions. Unhappy souls may intend to block thinking through drink or drugs, but their luckier fellow humans do not form intentions to think. So how can Kant or anyone be right that thinking is a kind of acting?

Kant originally presents thinking as acting by contrasting it with sensibility, which involves the passive receipt of information rather than its active combination (A51/B75, A68/B93, B129-30). Insofar as the distinction is just that between 'passive and active powers,' however, thinking would no more be a case of acting than digesting your food is. As we have seen, he also takes the synthesizing or combining that produces conceptual cognition to be conscious. When contrasting 'inner sense' and 'transcendental apperception,' he emphasizes that the power of apperception or understanding is conscious of its acts, even apart from sensibility (B153). He denies that consciousness is necessary for all mental action (B130), because he believes that

cognition requires a great deal of unconscious processing or combining. Still, the combinings at issue in higher cognition are conscious and that is probably the best way to justify his claim that they are acts.

Ironically, the classic argument connecting actions to consciousness was given by Anscombe in her important book, *Intention* (1957/1979). She argues at great length that action must be intentional, but not preceded by intentions. Rather, the crucial fact about action is its connection to a special sort of refusal. A person denies that she acted intentionally by saying 'I didn't know I was doing that.' On the other hand, if a person cannot deny that she knew she what she was doing, then she cannot deny that she was acting intentionally. Since higher cognition must be conscious for Kant, thinkers cannot deny knowing that they are thinking. Thus, higher cognition has the crucial feature of intentional action. Anscombe argues further that subjects have knowledge of their intentional actions without observation (e.g., 1957/1979, 14). As just noted, Kant maintains that the action awareness involved in thought is non-perceptual. Given Anscombe's analysis of what (intentional) action involves, Kant's claim that higher cognition must involve (intentional) action would be correct.

If thinking is (intentional) acting, because the subject (non-perceptually) knows what she is doing, then what does she think that she is doing? The answer must be that she thinks that she is changing her mind. If Franz knows what he is doing, then he knows that in judging that a third world war is unlikely he is not walking or singing—or changing the course of history. He is changing how he thinks about the likely course of history. Thus insofar as Kant is right that higher cognition is a kind of (intentional) action, then subjects who think must understand themselves to be thinking. And in that case, 'higher' cognizing must involve metacognition. The subject who comes to believe that a third world war is unlikely does not acquire infallible self-knowledge of the mental solely on the basis of his relations to the 'external' world, even his mental relations—his grasp of facts or of logical relations. Reasoning involves a mental appreciation of the mental—that the belief is a product of thinking and that it depends on other thoughts. Thus cognizers cannot acquire self-knowledge just by looking outward, because they cannot form inferred beliefs just by looking outward. As noted, this further argument will be persuasive only to those who take thinking to be a kind of acting and also accept

Ancombe's theory that the hallmark of (intentional) action is that it is characterized by its relation to a certain sort of refusal. I offer the argument because I think that many contemporary philosophers do consider thinking to be a form of acting and that many also find Anscombe's account of (intentional) action to be the best currently available.

The Kantian critique of transparency is similar to its critique of Anscombe's attempt to explain self-consciousness prior to considering the unity of the self. Evans and his many followers hope to explain self-knowledge, self-consciousness and self-reference without first addressing the question of are the necessary conditions for being a cognizing self. The lack of focus on this issue is not surprising, since a primary goal of Evans's work is to refute and replace the Cartesian conception of humans as first and foremost thinkers. He offers a rule that can get around the issue: Whenever you are in a position to assert that p (whatever that might involve), you are in a position to assert 'I believe that p '. As a result, he doesn't see that higher cognition already involves the mental, in particular, knowing of your states that they are necessarily connected mental states and knowing of your actions that they are thinking kinds of actions. Because inferred cognition of the external must include these mental elements, it cannot provide a reductive account of self-knowledge. In brief, self-ascription of belief cannot depend only on external factors because having beliefs is not just a matter of external factors.

Evans's defenders could throw the 'leaving out important factors' objection back at Kant. They could object that his focus on the sort of cognition distinctive of humans led him to the false position that humans are only or primarily cognitive subjects. This weakness in his account can be brought out by a skeptical challenge of the sort that Evans uses to motivate some of his views about 'I's. Suppose that one body counted from one to five, then passed its thoughts along to another that completed the count to nine? It seems reasonable to believe that the judgment 'I think that the number is nine' would be false. To avoid such cases, the Kantian account would have to be filled out by something like Evans's condition that the subject acquires information only in relevant ways. And that additional condition may require reference to particular bodies (1982, 216). This criticism seems fair and it turns on Evans's point that Early Modern accounts of subjects are too thin. Even if what is most distinctive about humans is their ability to think that does not mean that many other factors aren't involved in self-knowledge and

self-identity. Still, the skeptical challenge to Kant does not weaken the Kantian critique that Evans's account of self-knowledge of belief is itself too thin.

2. *Kant's Argument for First Person-Third Person Symmetry*

Although the expression 'standard view' is used too much, the Strawsonian thesis that it is a fundamental fact about psychological predicates that they have different application conditions for the first-person and second/third person cases should be characterized in these terms. As I noted in the third talk, that doctrine was the key premise in the argument of 'Persons.' The claim is repeated in countless subsequent discussions, including Evans's (1982, 226).

Kant's view that thinking is acting may seem, at first, to offer strong support for the asymmetry position. On the Kantian view that I have developed subjects are entitled to self-ascribe thinkings and thoughts just by thinking. To ascribe thoughts to others, they must do more. In the case of knowledge of 'other minds,' we do not have to work out a Kantian position. He drew out the implications of his theory of the 'I-think' for the issue as he began the Paralogisms chapter. As I argued earlier, he intends to use his positive discussion of the 'I-think' as a basis for criticizing the claims of Rational Psychology. For this reason, he prefaces his discussion with an explanation of how he is able to do so:

Es muß aber gleich anfangs befremdlich scheinen, daß die Bedingung, unter der ich überhaupt denke, und die mithin bloß eine Beschaffenheit meines Subjekts ist, zugleich für alles, was denkt, gültig sein solle, and daß wir auf einen empirisch scheinenden Satz ein apokirtisches und allgemeines Urteil zu gründen uns anmaßen können, nämlich: daß alles, was denkt, so beschaffen sei als der Ausspruch des Selbstbewußtseins es an mir aussagt. Die Ursache aber hiervon liegt darin: daß wir den Dingen a priori alle Eigenschaften notwendig beilegen müssen, die die Bedingungen ausmachen, unter welchen wir sie allein denken. (A346/B404-405, my underscoring)

Someone who is convinced by Lockean Empiricism will believe that I know the conditions under which I think empirically, namely, by observing my thinking through inner sense. Then he will question how I can generalize that others must think as I do.

Kant writes that this proposition merely *seems* empirical. When someone counts, for example, she does not observe herself think through inner sense. She consciously

combines. Because she has the ability to consciously combine some representations in others and to recognize the resulting judgments and the former representations as necessarily connected, she can both perform acts of higher cognition and can understand what higher cognition involves: In the case of judging, it involves applying representations on the basis of partial representations; it involves applying representations through having sensory indications of similarity to past cases where the term was used; more generally, it involves combining representations that are thereby understood as necessarily connected. In the case of inferring, it is a matter of knowing some things through knowing other things, and so of combining representations that are thereby recognized as necessarily connected. Thus what subjects come to understand through 'higher' cognizing is not just how they apply concepts or make inferences, but what higher cognition involves, and hence how any cognizer *must* think.

Although they have some metacognition, this does not mean that ordinary cognizers have theories of cognition. Rather, they have a practical understanding of thinking as creating a dependence of some thoughts on other thoughts. They apply that understanding to other minds and thus take everything that thinks to have the same character as their thinking. This practical understanding may be most evident in cases where they deal with humans who fail to meet the standard: they can see that a small child who calls a Newfoundland dog a 'horse' is thinking, but hasn't got the right similarity relation; they can understand that a scientist who didn't do any experiments must have made up his conclusion.

The procedure of using yourself as a model is justified, because the only access that humans have to the nature of higher cognition is through judging or inferring—through making and recognizing relations of dependence and so necessary connection across mental states. This unusual feature of higher cognition—that it can be understood only by those who perform acts of higher cognition—does not apply to most concepts. A subject could, for example, understand that the states of a substance can change merely by receiving sensory data and by possessing the category 'substance.' He doesn't have to be a substance. But cognizers can apply the predicates 'believe,' 'judge,' 'think,' *etc.*, to others only by using themselves as the model. Far from having divergent

application conditions in the first and third person cases, the application conditions are and must be exactly the same.

Strawsonians can reply to this argument with the obvious point that there is a still large difference between the first and third person cases: I know my thoughts just by thinking them, whereas I have to figure out someone else's thoughts from what she says and does. At this point, the disagreement may seem to be over which considerations are more important for knowledge of psychological states, the asymmetry between self-ascription and ascription of states to other minds or the necessary symmetry in understanding thinking and thoughts across all cases, first, second and third person. Part of my point is historical. Insofar as the asymmetry thesis is supposed to be traced back to Kant's theory of the 'I-think', that cannot be right. His view was that people self-ascribed beliefs through inner sense. More importantly, he developed his theory of the 'I-think' in the Transcendental Deduction and in the Paralogisms chapter. What I have just argued is that those texts provide a compelling argument for symmetry. In claiming Kant as a historical source of the asymmetry view, I think Strawson and his followers have missed his actual and very interesting argument for symmetry.

Strawson's defenders could argue that even if Kant had an argument for symmetry, it rests on questionable assumptions. In particular, they might argue that in maintaining that thinking cannot be represented by predicates that are applied through 'outer sense' (A22/B37, A347/B405, B419-20), he simply begs the question against the Wittgensteinian thesis that the human body is the best picture of the human soul (1953/1997, 178). They could object further that his solution to the alleged inability of outer or inner sense to provide a representation of a thinker/thinking—*viz.* the hypothesis of an *a priori* representation of a cognitive subject as a subject whose states are necessarily connected—is lazy, extravagant, and has its roots in the metaphysics he claims to leave behind.

These likely objections can, however, be met. At one level, Kant agrees with Wittgenstein's aphorism. He thinks that human form is, in fact, the best evidence for the presence of a mind. In *Anthropology from a Pragmatic Point of View*, he notes that we

know rational nature only through the human species (7.321). His equation of rational creatures and humans is also evident in the ethics where he refers almost indifferently to humans and to rational beings. This is especially obvious when he presents the second formulation of the categorical imperative. He switches from talking about how rational beings must be treated to stating the principle in terms of how 'humanity' should be treated (4.429). Further, as Strawson pointed out long ago, he also recognized that bodily continuity was the invariable evidence for mental continuity (B415). Kant also notes in the *Anthropology* that humans constantly observe the behavior of other humans and that they should study human behavior to become more astute moral agents (4.119ff.).

The argument for symmetry rests on a narrower claim: In the case of *higher cognition*, the predicates applicable through outer sense are irrelevant (A22/B37, A346/B405, B419-20). They bear no relation to the key idea of the dependence of one mental state on another, the relation of 'necessary connection' between the states. The notion of such dependence or 'necessary connection' can be understood only by making such connections, by engaging in mental acts in the course of cognizing.ⁱⁱ For that reason humans can understand other humans to be thinkers only by using their own activities as models—so the application conditions must be symmetric in all cases. This argument is compatible with holding that humans can learn a great deal about what and, indeed, whether others are thinking through observation. What 'outside' observation cannot supply is an understanding of what thinking is.

Kant described nativism as a 'lazy hypothesis' and tried to differentiate his claims about *a priori* concepts from it. Humans are not born with '*a priori*' representations, but acquire them through mental activities that are brought into play with the receipt of sensory data (B1-2). Showing that a representation is '*a priori*' is not suitable work for the lazy, because it requires demonstrating that certain mental activities are required for the production of cognition out of the materials of sense. In particular, Kant's claim that 'I-think' is *a priori* does not hypothesize an innate representation. What he takes to be innate in humans is not the representation, 'I-think,' but the capacity to combine representations and to grasp them as standing in relations of necessary connection when the occasion arises (cf. 8.221-23). That capacity is required for conceptual and

inferred cognition. Kant's claim that humans must be able to represent mental states as necessarily connected is not an additional hypothesis, but a direct implication of his analysis of higher cognition. If, as I have argued, that analysis is correct, then it follows that humans must have the capacity to recognize their mental states as necessarily connected. The only 'extra hypothesis' is Kant's explanation of the relation between this capacity and the representation 'I.' The 'I-think' collects the mental states that stand in the relations of necessary connection that are necessary for higher cognition under a single expression. To say that the representation 'I-think' is *a priori* is to claim nothing more than that humans must have the capacity to make and recognize relations of necessary connection dependence across states and to collect the states so related under a single expression, 'I-think.'

As we have just seen, Kant's argument for symmetry is narrowly focused. It does not cover the full range of mental predicates, but only those involving higher cognition. Cognition has a number of special features that may or may not be generalizable to cases of sensation and desire. Still, it is a sufficiently important class of cases that it demonstrates that first person third person symmetry may be at least as important as any asymmetry when trying to explain self-knowledge and knowledge of other minds.

Despite his intense study of the Deduction and the Paralogisms, Strawson took away the wrong lesson about the 'I-think.' He thought that Kant's claims about mental activity were supposed to provide an explanation of the construction of a known phenomenal world out of an unknowable noumenal one. Correctly seeing such a project as incoherent, he dismissed the theory of mental activity as an aspect of the *Critique* that had no relevance for current research. By contrast, I've argued that the painstaking study of the mental activity involved in cognition has many interesting implications about how we should understand mental unity, self-knowledge of belief and how we represent each other.

ⁱ Of course, Franz could come to the belief that a third world war is unlikely by appealing to other considerations. The assumption is that he didn't. Hence in these circumstances, his belief is dependent on his beliefs that wars are correlated with the level of arms and that the current level of arms is low.

ⁱⁱ This point should be clarified. It is certainly possible to have an abstract understanding of necessary connection. A is necessarily connected to B just in case A could not exist if B did not exist. But this merely nominal understanding of necessary connection does not suffice for the subject to understand how any items could actually stand in this relation. See Kant's discussion of understanding freedom 'in name only' as opposed to understanding how it is really possible in the preface to the *Critique of Practical Reason*. (5.5)