

Replies

PATRICIA KITCHER

Columbia University

Email: pk206@columbia.edu

A central claim of *Kant's Thinker* is that the Transcendental Deduction succeeds in arguing that conceptual cognition and the unity of self-consciousness are necessary and sufficient conditions for each other. Ralf Busse, Tobias Rosefeldt and Thomas Sturm question whether Kant establishes that cognition requires apperception (CRA); Falk Wunderlich objects that apperception could not imply conceptual (or categorial) cognition.

Kant argues for CRA while maintaining that there is no consciousness of a self. Paradox is avoided, because he takes self-consciousness to be a matter of recognizing relations of rational and so necessary connection across mental states. Hume had rejected the notion of a 'self', claiming that there is no constant impression of a self and no 'real bond' or necessary connection across mental states (Hume 1739/: 250, 252, 259). The argument for CRA is that rational cognition requires such a bond and a bond that is recognized.

Kant does not believe that all cognition is rational (based on reasons) (*L*, 9: 64–5). Only conceptual cognition is and so requires the unity of self-consciousness. Conceptual cognition makes representations of marks that are common to many things the ground of cognition (*L*, 9: 58, *R*, 16: 300, R2288). Cognizers regard concepts in this way (*R*, 16: 299, R2285). His claim is not that ordinary people use terms such as 'concept' or 'representation', but that they understand that in calling something 'orange', for example, they are saying that it shares something with other things they call by that name. Concept users can also use concepts that are contained in others ('partial' concepts) as the basis or ground for applying other (whole) concepts (*L*, 9: 58). So a concept user could call something an 'orange', on the grounds that it is orange, round, edible, and so forth. Again, he does not claim that ordinary cognizers speak of grounds of cognition, but only that they understand that anything they call an 'orange' must be e.g. more or less round

and that roundness, edibility, etc. are reasons for calling something an ‘orange’.

Kant uses the example of counting to argue that concept application requires the unity of apperception:

If, in counting, I forget that the units that now float before my mind or senses were added together by me one after another, I would not ... cognize the number.

The very word concept could on its own lead us to this observation. For this one consciousness is what unites in one representation what is manifold, intuited little by little, and then also reproduced. Often this consciousness may be only faint, so that we do not [notice it] in the act itself, i.e. do not connect it directly with the representation’s production, but [notice it] only in the act’s effect. Yet, despite these differences, a consciousness must always be encountered, even if it lacks striking clarity; without this consciousness, concepts, and along with them cognition of objects, are quite impossible. (Kant 1996: 156 amended translation; *CPR*, A103–4)

Representations of the partial concepts, ‘one’, ‘two’, etc., must not merely be retained, but consciously synthesized in making the judgement, say ‘four’. Thinkers need not pay attention to each step, but they must be conscious in applying the concept ‘four’ of doing so on the basis of the partial concepts. Otherwise, they would be unable to use concepts. For, to cognize through concepts humans must not merely have representations that lead to other representations. In the case of complex concepts, they must recognize the partial representations as the grounds of cognition of the whole representation (judgement) and so the existence and rationality of the judgement as dependent upon them.¹ A related point applies to concepts that have no parts or are used like simple concepts. The cognizer does not merely have an intuitive representation of e.g. a unit. She must be conscious of applying the partial concept ‘one’ on the basis of that intuitive representation (that she takes to present the same property to which she has previously applied the concept ‘one’) and so of the dependence of her use of the partial concept on that intuition (and on previous uses and their intuitions).²

Kant moves directly from his analysis of conceptual cognition to his claim about the unity of apperception, because he takes the *a priori*

representation ‘I-think’ to be governed by what I label the ‘I-rule’, the principle that representations must belong to some single ‘I-think’ (Kitcher 2011: 123–6). Once he has shown that conceptual cognition requires that some faculty (or faculties) consciously combine some mental states in others and recognize the relation of dependence thus produced, he has also shown that it requires that different representations must belong to a common ‘I-think’, i.e. the unity of apperception.

Busse objects to this interpretation by appealing to Kant’s example of drawing a line. He reads that discussion as showing that a subject can represent different segments (a, b, c) as integrated into one line only by actively integrating her intuitions of them into one intuition. This requires a synthetic unity of consciousness only in the sense that the synthesizing act must be ‘a unified conscious ... performance’ (III). There need be no reflexive awareness of the act and so no *self*-consciousness. In this passage, however, Kant is not exploring the requirements of drawing a line, but of ‘cognizing something’ under a concept (B138). This requires intuitions that follow the rule for lines, so the subject mentally draws a straight non-gappy figure and judges ‘Line’. In making that judgement, she simultaneously cognizes the line and understands the judgement as rationally dependent on and so as necessarily connected to the intuitions of the parts a, b and c. Hence Kant concludes the discussion by asserting that the synthetic unity of consciousness is a necessary condition for all object cognition (B138).

Against this background, Rosefeldt’s counter-examples seem off the mark. Is the 3-year-old who hears mommy arrive a concept user? The dog might hear the same sound and start wagging its tail. Some philosophers would describe both dog and toddler as believing that mommy is home. It seems a stretch, though, to characterize the dog’s belief as based on a reason. Not wanting the debate to turn on who owns the terms ‘belief’ or ‘rational belief’, my strategy was to defend Kant by arguing that only his account of conceptual cognition, which implies CRA, could explain the fact that (normal) mature humans have self-knowledge of their reasons for beliefs. In that case, although there may be good reasons to use ‘belief’ more widely than he would use ‘judgement’, he would still have shown that an interesting, distinctly human level of cognition requires apperception.

I tried to establish this thesis by arguing that the best alternative account of self-knowledge of reasons for belief, that of Kieran Setiya, was vulnerable to systematic objections. Setiya tries to explain how

subjects can move from having beliefs based on reasons to having self-knowledge through an epistemic principle:

If you believe that the fact that q shows that p , form the belief that you believe that p , because you believe that q . (n.d. 20, n. 41)

Setiya uses the ‘fact’ locution to make it clear that we are not already dealing with a belief about your belief. You believe that the fact that q shows that p and not that your belief that q shows (you) that p . My objection was that this rule could be learned only by subjects who understand the ‘showing’ relation and that that involves knowing the appropriate *relata*, namely, facts and a fact appreciator and so reference to a cognitive subject. The problem seems systematic, because ‘shows’ or some similar locution must be used to express the content of a belief when a person believes one thing on the basis of something else.

Rosefeldt counters by offering two principles that would capture evidential relations without their antecedents referring to evidence appreciators. He removes the reference to ‘showing’ in Setiya’s principle in formulating Rule 2*:

Whenever a rational subject believes that the fact that q makes it sufficiently probable that p she is entitled to believe that she believes that p because she believes that q .

Rule 2* is, however, a bad epistemic rule. There are many reasons for believing propositions and although a subject may believe that the fact that q makes it sufficiently probable that p , she may believe that p for a different reason.

Rosefeldt’s second principle, Rule 3, offers an epistemic standard:

Whenever a rational subject believes that p because she believes that q , she ought to believe that the fact that q makes it sufficiently likely that p .

He suggests that cognizers could move from having reasons for a belief to having reflected reasons via Rule 3, and then from that step to self-knowledge of beliefs based on reason via Rule 2*. The passage from having reasons for beliefs to self-knowledge of those beliefs and reasons via these two rules is, however, too leaky to explain the infallible character of self-knowledge of belief.

There is a further problem with Setiya's rule and Rosefeldt's pair of rules, because they are supposed to address two issues. One is justification. At least in Setiya's case, that is handled by the form of the rule. If a subject uses this rule, then her self-knowledge claim must be correct. As Rosefeldt's language makes clear, the rules are also intended to explain how self-knowledge is possible without appealing to inner sense. It is by learning these rules along with acquiring the concepts of 'belief' and 'reason'. These acquisitions will not, however, enable a subject to apply the rules. She must not only believe that p because q . She must have access to this information or self-knowledge of belief would be miraculous. Given the serious failures of alternative accounts of self-knowledge of reason-based beliefs, Kant's theory of conceptual cognition seems more attractive, because it implies that cognizers must always be implicitly conscious of the beliefs they have formed through judging and inferring.

Busse takes me to understand Kant's theory of cognition in terms of a second-order consciousness of mental states, acts and subjects. I tried to avoid this reading by describing cognizers as aware *in* rather than *of* the act of cognizing. As I understand Kant, his view is that the conscious synthesizing involved in 'higher cognition' enables subjects to know the rational and necessary connections across their states, not because they observe the synthesizing, but because they have carried it out. Kant is clear that the consciousness required for rational cognition is non-sensory, but he has no positive model to offer (B1 53). Still, the inability to model the consciousness by analogy with something else does not weaken his argument that cognizers who judge or reason must do so consciously. It is this practical consciousness in reasoning from q to p or in judging 'four' through counting that provides the access required for particular claims of self-knowledge of beliefs based on reasons.

Given these considerations, Kant's theory should not be characterized as requiring second-order consciousness *of* mental acts. As noted, what he means by 'apperception' is not being conscious of a self, but recognizing (by having produced) necessary connections across mental states. The case of representations may seem different. If a cognizer recognizes the dependency of some representations on others, then it may seem that he is reflecting on his thought, as Busse reflects 'that thought a second ago was rubbish'.

The similarity is, however, illusory. Busse's earlier thought could exist without any subsequent reflection; on Kant's theory, consciousness of

the relations of dependency and necessary connection across representations is a necessary condition for a possible judgement to be an episode of conceptual cognition. We can also see the difference between Busse's second-order reflection on a thought and Kant's analysis of higher cognition by considering Busse's case. He could not have the thought that his earlier thought was rubbish unless he saw that it was inconsistent with itself or with obvious truths or objectionable in some other way. In those circumstances, however, Kantian consciousness of the dependence of some thoughts on others, i.e. the dependence of the thought that the first thought was rubbish on both the first thought and the thoughts that give it the lie, is a necessary condition for the sort of second-order reflection that Busse presents and so cannot be assimilated to it.

Although Busse's distinction between thinking in representations and thinking about them seems exhaustive, it does not capture Kant's view. Animals recognize objects in or through representations, with no understanding of what they are doing; theorists and people in critical or reflective moments think about thoughts and so representations. On Kant's theory, when humans engage in rational cognition, they consciously combine representations and so know what they are doing. Reflecting on the representations (or the mental actions) would interfere with what they are doing, which is using the representations to produce further judgements.

Busse notes that Kant makes the activity of thinking or integrating different representations in consciousness the basis of I-consciousness. He objects that Kant did not appreciate the gap between a requirement that a single thinker integrate a multitude of representations and the requirement that she be aware of herself as an 'I-think'. This criticism is only half right. Kant appreciates (as Busse does not) that the capacity for rational cognition requires not just the ability to recognize e.g. that 'Socrates is mortal' follows from the usual premises, but also the ability to recognize a thought as rationally dependent on other mental states.

Still Busse is right that Kant has no satisfactory account of how cognizers can use 'I-think' of themselves. Sturm objects that my Kantian solution to the problem of self-knowledge of belief fails, because the theory does not address the 'mineness' of mental states, but only their 'togetherness', their unity. On Kant's theory we understand that and why humans must be conscious of a unified self that is theirs. But how are they conscious of it *as theirs*? He did not address the issue of the use

of 'I', because he believed that Locke's theory of inner sense provided the answer.

Still, there seems to be a fairly straightforward way to supplement Kant's theory to handle 'mineness'. Consider the self-reference rule for 'I': "I" refers to the thinker of this thought' (e.g. O'Brien 2007: 7). Since self-reference is infallible, cognizers must have some way of referring to themselves as subjects of thought that necessarily accords with the self-reference rule. On Kant's theory, cognizers who think are conscious in combining some representations in others. Hence they could use the following rule:

When you (consciously) perform an act of higher cognition producing a judgement *p*, self-ascribe the thinking and the thought, 'I think' and 'I think that *p*'.

I put 'consciously' in parentheses, because on his theory, it is pleonastic. The consciousness is not of one's self performing the act (or the account would be circular), but of performing the act. It does not rest on observation of, or reflection on, acts. It is, rather, the conscious combining of some representations in others. Since this consciousness must be present for the act to be one of rational cognition, cognizers who learn some such rule for using 'I-think' will always be able to ascribe their judgements and inferences and the self-ascriptions will always be correct.

Whether or not possessing a rule for using *I-think* is a necessary condition for rational cognition, the act-consciousness and the consciousness of relations of rational and so necessary connection across representations (which make the states and activities those of a unified thinker, an 'I-think') that would make such a rule usable and infallible are necessary conditions.³ That seems to me sufficient to defend Kant's view that conceptual cognition requires the unity of self-consciousness.

Following Henrich, Wunderlich wonders how Kant could establish that self-consciousness is sufficient for object cognition according to categorical rules. Henrich maintains that cognizers know with Cartesian certainty that different thoughts are united in the same thinker, but he believes that it is impossible to move from that unity to claims about categorical principles (1989: 256). In Wunderlich's formulation (144), it seems impossible to argue for the categories from the tautology that 'all my representations are mine'. *Kant's Thinker* does not present the argument from apperception as starting from a tautology, but from the

assumption that (*pace* Hume) cognizers know themselves to be continuing selves. It presents Kant as making a twentieth-century-type transcendental argument (p. 133): Since cognizers are not aware of a constant self (A107) and the consciousness that must accompany all mental states does not establish sameness of consciousness across different states (B133), the only explanation left standing is that they engage in rule-governed conceptual cognition (B420). I now think that Kant's argument is stronger. The kind of unity that is constitutive of being an 'I-think' is not merely necessary connection, but rational (necessary) connection across different mental states and that unity can only be produced by thinking about objects. Neither this argument nor the one I present in the book completes the arguments for individual categories, but they do yield a central result of the Deduction chapter: The unity of self-consciousness requires object cognition.

Wunderlich and Sturm raise a crucial question about the nature of Kant's thinker. Kant claims that the act of thinking 'determines his existence' (B157n). What sort of existence is determined by thinking? Is Kant's characterization of the thinker purely functional or does he imply that there is some minimal underlying substratum?

The note at B157–8 asserting that thinking determines the existence of a thinker is confusing, because it also claims that cognitive subjects cannot determine themselves as self-active beings, because they have no self-intuition (except inner sense) to give the determinable (the matter to be determined by concepts, see A261/B317). Although cognizers cannot attribute anything determinate to themselves as self-creators, they can attribute determinate states and conditions to themselves as self-created beings. When cognitive faculties synthesize the materials of intuition, they create representations of objects and so new mental states, and rational (causal) connections across mental states. Through their conscious synthesizing in rational thought, cognizers have access to determinate states and conditions of themselves as self-created cognitive subjects, as e.g. now thinking *p*, because *q*, or now judging 'four' through counting.

Since synthesizing brings about the real connections across mental states, it seems apt to characterize Kant as offering a functionalist characterization of the cognitive subject. Appreciating the centrality of conscious synthesis to his theory of cognition, however, now (as opposed to Kitcher 1990) leads me to qualify that position. Kant understood that reasons were a special kind of cause. When that crucial

point is forgotten, characterizing the mind functionally makes it seem to be a realm of ‘blind’ causal relations – which is just what he denied (although he agreed that there were many unconscious activities underlying rational thought, *Anthropologie* 7.135). Hence, we need a more precise characterization: Kant characterizes the cognitive subject as a set of states that have been or can be connected through rational thought. This characterization is functionalist and so neutral about any underlying substratum, but it differs from many familiar kinds of functionalism because of the nature of causation involved.

Since Kant’s thinker requires active and receptive powers as well as real connections across mental states, one might argue that his subject is a substance, because active powers must reside in substances. He has some confusing things to say about this issue at A205/B250, but he could not have accepted the power or activity to substance inference without contradicting the discussion of the First Paralogism (A349) and rendering the reasoning of the First and Second Analogies otiose.

Wunderlich objects that my reading of Kant’s discussion of whether the existence of the thinker can be established by thought alone mixes *a priori* and empirical matters. The empirical proposition ‘I think’ is the topic of the long note in the B Paralogisms where Kant clarifies the relations among that proposition, the representation ‘I-think’ and intuition. As do all others for Kant, this empirical proposition has an *a priori* component, namely, the representation ‘I-think’ (see B423, n.). *Kant’s Thinker* does not claim that Kant comes to realize that thought bereft of intuition can establish existence, but that he may have come to see why Cartesians think it can (2011: 194–7). Since ‘I-think’ is an *a priori* representation, it is neither acquired through experience nor paired with a particular kind of intuition. In that case, humans can imagine with Descartes that all their past experience is erroneous and still consider themselves to be conscious of their unity in and through apperceptive act-consciousness. Some (given) perceptual materials are required, but these can be completely indeterminate. They would imagine combining indeterminate perceptions according to some rule and of being conscious in that combining. In this way, they can imagine themselves continuing to exist as thinkers, e.g. as believing *p* because *q*, without having spatiotemporal or any particular intuitions at all.

Heiner Klemme rightly wonders why I present myself as supporting Onora O’Neill’s claim that the categorical imperative is the highest principle of practical and speculative reason, when my central claim is

that the spontaneity of Kant's thinker differs in important ways from that of his moral agent. Akeel Bilgrami (2006: 170) suggested that Kant pointed the way for understanding cognition as responsible, because spontaneous and agential, by assimilating it to morality (2006: 170). Henry Allison has long maintained that philosophers should not resist the absolute spontaneity that Kant requires for moral action, because he requires the same spontaneity for higher cognition (e.g. 1990: 36). Against both I argue that Kant presents theoretical and practical reason as involving significantly different kinds of spontaneity. In his view, theoretical reason is autonomous at four levels: reason thinks in accord with its principles (as opposed to principles borrowed from experience), it makes those principles explicit, it evaluates some of its principles by appeal to others, and philosophers can evaluate its most basic principles and establish their legitimacy through a transcendental deduction. Since there can be no deduction of the moral law, practical reason is autonomous on only the first three levels. On the other hand, the consciousness of the moral law and its 'ought' in practical reasoning provide moral agents with practical cognition of an entirely different sort of freedom, the freedom to follow the moral law regardless of the state of the world (including their prior actions) up to that point (2011: 247).

Passages in the *Logic* claim that intellectual mistakes are imputable (9: 54, 74), and so presumably free. As I read those discussions, however, it is not errors of theoretical reasoning *per se* that are imputable, but failures to develop reasoning skills and to take up controversial topics that can be faulted (2011: 247–8). Hence there is a sense in which thinking is subject to the demands of the categorical imperative, but that is only because thinkers can choose to develop their talents and can decide which issues are worth studying. Klemme is right that this is too weak a sense of being 'governed by the categorical imperative' to provide support for O'Neill's bold claim.

Klemme believes that Kant takes the *Critique of the Power of Judgement* to offer a solution to the unity of reason question. I do not see how Kant could, since he forecloses the possibility of a common principle of practical and theoretical reason with his claim in *Religion* (6: 26, n.) that rationality is possible in the absence of moral personality. Insofar as his views about the sources of moral obligation are plausible, Kant's contribution to contemporary debates about assimilating moral and cognitive agents would fall on the negative side. *Kant's Thinker* offers only that hypothetical conclusion, because it leaves the assessment of Kant's moral theory to another occasion.

Notes

- 1 A judgement 'four' could have many grounds. In Kant's exemplary case, the judgement is the result of counting and so understood as necessarily connected to the partial representations, 'one', etc.
- 2 Different intuitions could serve as the basis for applying the partial concept. In any particular application, however, the cognizer understands that her application would be impossible without the intuition on which it is based.
- 3 The rule for using 'I-think' presupposes act-consciousness and Kant's principle that different representations must belong to a single thinker (which I call the 'I-rule', 2011: 123–6). This is clearer if the self-ascription rule is fully spelled out:

When you (consciously) perform an act of higher cognition that produces the judgement *p* (and so recognize the necessary connection of that mental state to others), self-ascribe the thinking and the thought, 'I think' and 'I think that *p*'.

References

- Allison, Henry (1990) *Kant's Theory of Freedom*. New York: Cambridge University Press.
- Bilgrami, Akeel (2006) *Self-Knowledge and Resentment*. Cambridge, MA: Harvard University Press.
- Henrich, Dieter (1989) 'The Identity of the Subject in the Transcendental Deduction'. In E. Schaper and W. Vossenkuhl (eds), *Reading Kant: New Perspectives on Transcendental Arguments and Critical Philosophy* (New York: Basil Blackwell), 250–80.
- Hume, David (1739/1978) *A Treatise of Human Nature*, ed. P. H. Nidditch (revision of L. A. Selby-Bigge), 2nd edn. Oxford: Oxford University Press.
- Kant, Immanuel (1902–) *Kants gesammelte Schriften*. the Koeniglichen Preussischen Akademie der Wissenschaften, 29 vols. Berlin: Walter de Gruyter and predecessors.
- (1996) *Critique of Pure Reason: Unified Edition*, trans. Werner Pluhar. Indianapolis: Hackett Publishing Company.
- Kitcher, Patricia (1990) *Kant's Transcendental Psychology*. New York: Oxford University Press.
- (2011) *Kant's Thinker*. New York: Oxford University Press.
- O'Brien, Lucy (2007) *Self-Knowing Agents*. Oxford: Oxford University Press.
- Setiya, Keiran (n.d.) 'Knowledge of Intention' (draft, 30 April 2009). The published version appeared in Anton Ford, Jennifer Hornsby, and Frederick Stoutland, eds, *Essays on Anscombe's Intention*, Harvard, 2011, 170–197.