Kantian Review (2024), **29**, 289–298
doi:10.1017/S1369415424000128

KANTIAN REVIEW
CAMBRIDGE UNIVERSITY PRESS

AUTHOR MEETS CRITICS

Understanding the First Paralogism: A Friendly Disagreement

Patricia Kitcher

Columbia University, New York, NY, USA
Email: pk206@columbia.edu

Abstract

My comments focus on Proops's treatment of the Paralogisms. I agree with many aspects of his discussion, including his views about the project of Rational Psychology and his analyses of how, exactly, the arguments of the Paralogisms are defective in form, but I disagree with his interpretation of the First Paralogism. I argue that the source of confusion that Kant diagnoses is not the grammatical distribution of 'I' as singular, but the fact that the I-representation is both empty and necessary for cognition.

Keywords: I-think; Rational Psychology; transcendental deduction

1. Approaching the Paralogisms

Ian Proops has written a wonderful study of Kant's Dialectic,[1] to borrow a phrase from Jonathan Bennett (1974), one of his few predecessors in this field. Although the Dialectic has received less systematic attention than the Analytic,[2] the part I consider, the Paralogisms, has been widely discussed. P. F. Strawson (1966) devoted a section of *The Bounds of Sense* to the topic, and the focus of Wilfrid Sellars's Presidential Address to the American Philosophical Association (1970) was '. . . this "I or he or it (the thing) which thinks . . . "'. Both of these seminal treatments spawned large literatures on the relevance of Kant's insights into the self and consciousness to contemporary work. Karl Ameriks's groundbreaking, *Kant's Theory of Mind* (1982/2000) began a new line of inquiry, where the focus was less on mining Kant's text for ideas useful to current research and more on the historical context of the Rational Psychologists who were Kant's targets.

Others may divide the interpretative territory differently, but I think about approaches to the Paralogisms along two dimensions. One concerns how the Paralogisms chapter fits into the *Critique*. I follow Sellars and Strawson in seeing the Paralogisms as an extension of the discussion of transcendental unity of consciousness in the Transcendental Deduction.[3] By contrast, Ameriks emphasizes the context of the Dialectic's critique of pure metaphysics, in the case of the Paralogisms, the alleged discipline of Rational Psychology.

On the other hand, I agree with Ameriks that what is primarily at stake in the Paralogisms are metaphysical issues and only secondarily confusions about the use of

© The Author(s), 2024. Published by Cambridge University Press on behalf of Kantian Review. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/S1369415424000128 Published online by Cambridge University Press

language, confusions that Sellars and Strawson place at the centre of Kant's criticisms. Sellars presents Kant's diagnosis of the First Paralogism as involving an illicit move from features of the I-representation to features of the I that is represented:

> The representation of the I is not the representation of ... an aspect of something more basic ...
>
> Therefore, The I is not ... [a]n aspect of something more basic. (Sellars 1970: 13)

Sellars does not consider an aspect of Kant's discussion that Proops rightly takes to be crucial (p. 104): Why would such an inference seem plausible?

By contrast, Strawson thinks that the brilliance of the Paralogisms chapter rests on Kant's diagnosis of a confusion that is almost inevitable, given our normal ways of thinking about reference. The fact that

> lies at the root of the Cartesian illusion ... [is that w]hen a ... subject of experience ascribes a current or directly remembered state of consciousness to himself, no use whatever of any criteria of personal identity is required to justify his use of the pronoun 'I' to refer to the subject of that experience. (Strawson 1966: 165)

Strawson's reconstruction has been the definitive reading of Kant's critical insight in the Paralogisms.

On these two dimensions, Proops's approach is the opposite of what I would recommend. He sees the Paralogisms as exposing confusions about the use of the word 'I' and as mainly concerned with the arguments of Rational Psychology and not with Kant's doctrine of the 'I think' from the Deduction. Still, I view the disagreement I present as a friendly one, because as I argue below, Rational Psychology and the Deduction are complementary contexts for understanding the Paralogisms. Further, I agree with many elements of Proops's treatment of the Paralogisms, including his views about the project of Rational Psychology and with his analysis of how, exactly, the arguments of the Paralogisms are defective in form. My disagreement concerns only whether the source of confusion about the 'I' is linguistic or epistemological/metaphysical, and even here our views may not be that far apart.

2. The project of Rational Psychology

Drawing on Corey Dyck's *Kant and Rational Psychology*, Proops begins with the aims and methods of Rational Psychology. One of Dyck's central claims is that 'Rational Psychology' is a misnomer, because, as Proops explains, Rational Psychology's founder, Christian Wolff, uses two primary methods, derivation and observation. Rational Psychology starts from an empirical study of operations of the soul/self. This leads to a nominal definition capturing the observed features. From the nominal definition, Rational Psychology tries to derive and so identify the self's real essence. The derivation proceeds by reducing the observed operations of the soul to ever more basic faculties. As a cross-check on the derivation's adequacy, the procedure is

https://doi.org/10.1017/S1369415424000128 Published online by Cambridge University Press

reversed. Given the real essence, is it possible to derive all the observed operations of the soul? Proops cites Wolff's account of the task of Rational Psychology:

> we derive a priori from a unique concept of the human soul all the things which are observed a posteriori to pertain to the human soul. (p. 65)

As Proops notes, however, for all the talk of deriving empirically observed features of human psychology from their grounds in the real essence of the self, precious little deriving gets done in the *German Metaphysics* (p. 64). In a note to the B Paralogisms, Kant appears to mock the Rational Psychologists' claims to derive and so 'explain through grounds':

> The rationalist makes, out of our mere power of thought ... a self-subsistent being; and he does so merely because the unity of apperception in thought permits him no explanation of the soul from what is composite – whereas he would do better to admit that he does not know how to explain the possibility of a thinking nature ... (B415a).[4]

In whatever way the Rational Psychologist arrives at the real essence of a thinking self as a simple substance – perhaps by reasoning that thinking must be grounded either in something complex or in something simple, and it cannot be explained by a composite – he cannot verify his claim by returning to the empirical evidence. He cannot derive the operation of thinking from this substance, because he cannot explain thinking at all.

This weakness in Rational Psychology is not something Kant discovered in 1787. In his 1778–79 or 1779–80 (Carl 1989: 119) 'Metaphysics L_1' lectures, he observed that the project of deriving all the operations of the self from a real essence faltered at the earlier stage of deriving different relatively basic faculties from a single basic faculty:

> **[It] is a wholly other question: whether we are capable of deriving all actions of the soul, and its various powers and faculties, from one basic power.** This we are in no way in the position [to do], for we certainly cannot derive effects which are actually different from one another from one basic power. (28: 261–2, original emphasis).[5]

Kant repeats this point in the First Introduction to the *Critique of the Power of Judgement*:

> It **can** easily be demonstrated, and has already been understood for some time, that this attempt to bring unity into the multiplicity of faculties ... is futile. (20: 206)[6]

Given this background, the task of the Paralogisms chapter is, as Proops argues, less to refute the claims of Rational Psychology than to explain why they seem so seductive.

https://doi.org/10.1017/S1369415424000128 Published online by Cambridge University Press

3. Proops's interpretation of the First Paralogism in B

Proops prefaces his account of Kant's diagnoses of the mistakes of Rational Psychology with an analysis of paralogisms. The issue seems simpler than it is, because Kant says what he means by a 'paralogism': It is an argument whose form is defective, by virtue of *sophisma figurae dictionis* (A402), which is usually understood as 'ambiguous middle'. As Proops argues, however, it is difficult to fit Kant's criticisms of the Paralogisms into that one mould. He reads Kant as deploying at least three different 'styles of criticism' against the Rational Psychologists' arguments. All can be loosely characterized as involving 'problems of form', and all have problems with their middle terms. The simplest type is ambiguous middle. In a second type, the middle term of the minor premise matches the meaning of the middle term of the major, but with that common meaning, the minor premise is not known to be true. In the third, the middle terms match and the premises are known to be true, but the conclusion is so weak that it is irrelevant to the point at issue. In the third type, the formal fallacy is not ambiguous middle, but *ignoratio elenchi*. I discuss only the First Paralogism in the B edition, but Proops's taxonomy of formal errors is helpful in following Kant's objections to all the Paralogisms in both editions.

Proops takes the first B Paralogism to involve ambiguous middle. This is the reasoning to be criticized:

> **What cannot be thought otherwise than as subject does not exist otherwise than as subject, and is therefore substance.**
>
> **Now a thinking being, considered merely as such, cannot be thought otherwise than as subject.**
>
> **Therefore it also exists only as such a thing, i.e. as a substance**. (B410-11, Proops, p. 104; original emphasis)

Kant is explicit about where the ambiguity lies:

> In the first premise, one talks about things that cannot be thought of other than as subjects, in the second, however, one talks not about **things**, but about **thought** (in that one abstracts from every object), in which the I always serves as the subject of consciousness; hence in the conclusion it cannot follow that I cannot exist otherwise than as subject, but rather only that, in thinking my existence, I can use myself only as the subject of the judgement, which is an identical proposition that discloses absolutely nothing about the manner of my existence. (B411n, cited on p. 107)

I agree with Proops's analysis of the major premise. If there is an entity that cannot be thought as a property in the sense of 'conceived' as a property, then that entity cannot exist as a property (and so is a substance0. As Proops explains, the key to understanding the Paralogism is to determine the meaning of 'thought' in the minor premise, such that the premise is understood as known, and as necessarily true.

https://doi.org/10.1017/S1369415424000128 Published online by Cambridge University Press

Proops works backwards to the interpretation of the minor, by focusing on the conclusion that Kant thinks follows, namely:

> 1. In thinking my existence, I can use myself only as the subject of the judgement. (p. 107)

He takes Kant to be committing something like a use/mention confusion, so he offers a charitable interpretation of 1:

> 2. In thinking my existence I can use the representation of my self only as the subject of the judgement. (p. 108)

Since I could think 'something instantiates the property of being me', Proops amends 2 further to 3:

> 3. In thinking my existence I cannot use the representation of my self as the predicate of the judgement. (p. 108)

Working backwards, Proops suggests that the minor premise should be understood as:

> Minor (IP): The representation of a thinking being, considered merely as such, cannot be used as the predicate of a judgement. (p. 109)

On Proops's reading, 'something that cannot be thought otherwise than as subject' in the minor means 'something whose representation cannot play the grammatical role of a predicate' (p. 110). 'I' cannot occur as a predicate, because it is singular and only terms for universals can appear in the predicate position. Thus, Proops claims that

> Kant's key observation about the B-edition first paralogism is that this fact about the grammatical distribution of the first-person singular pronoun entails nothing about how the self can be conceived. *A fortiori*, it entails nothing about whether or not the self is a substance. (p. 109)

It is at this point in the interpretation that I want to object both on textual and plausibility grounds. The clarificatory note cited above says nothing about singular terms, but about what use the 'I' serves. It serves to indicate the subject of consciousness; it is used for the subject of judgement. Further, it seems to me that Proops's interpretation fails his criterion that a successful interpretation of the First Paralogism should present an argument that Rational Psychologists and other sophisticated thinkers would find plausible (p. 104). Why would anyone move from a claim about the grammatical distribution of the 'I' to a claim about the self being a substance?

As Proops observes, Kant's critique of the Paralogisms' reasoning comes after he offers a defence of how it is possible to make claims about thinking beings in general. I take Kant to be defending not the possibility of Rational Psychology, as Proops does (pp. 80–1), but his right to make the arguments in the chapter. This is Kant's defence:

https://doi.org/10.1017/S1369415424000128 Published online by Cambridge University Press

> It must, however, seem strange at the very outset that the condition under which I think at all, and which is therefore merely a characteristic of myself as subject, is to be valid also for everything that thinks; and that upon a proposition that seems empirical we can presume to base an apodeictic and universal judgement, *viz*: that everything that thinks is of such a character as the pronouncement of self-consciousness asserts of me. The cause of this, however, lies in the fact that we must necessarily ascribe to things a priori all of the properties that make up the conditions under which alone we think them [and] … solely through self-consciousness, can I have the least representation of a thinking being. Hence objects of that sort are nothing more than the transfer of this consciousness of mine to other things, which thereby alone are represented as thinking beings. (A346-7/B404-5, Proops cites parts of the passage on pp. 80–1)

This passage provides an invaluable clue about how to understand the arguments of the chapter. Proops observes, with surprise, that Kant thinks that his transference 'story' (p. 81) is obvious. In the Second A Paralogism, Kant writes:

> It is obvious that, if one wishes to represent a thinking being, one must put oneself in its place, and then substitute one's own subject for the object one wants to consider [which is not the case in any other kind of investigation]. (A353-4)

I do not think of Kant's transference account as a story, but as a plausible theory of how cognizers are able to represent other creatures as minded. This passage points back to the Transcendental Deduction, since that is where the 'pronouncements' of self-consciousness occur. Further, it is in the Deduction that Kant argues that the ability of a cognizer to use the representation 'I' or 'I think' is a necessary condition for the possibility of cognition, so that is where we need to look to understand why Kant thinks it is both necessarily true and known to be true that the 'I' serves as the subject of conscious judgements.

4. Self-consciousness and judgement

In the B edition, Kant explains how subjects can and must use the representation 'I' or 'I think' to refer to a subject of judgements:

> [T]he empirical consciousness that accompanies different representations is intrinsically sporadic and without any reference to the subject's identity. … [T]his reference comes about … through my **adding** one representation to the other and being conscious of their synthesis. Hence only because I can combine a manifold of given representations **in one consciousness** is it possible for me to represent the **identity itself of the consciousness in these representations.** (B133)

In thinking, the subject is conscious of synthesizing representations in a thought that she grasps as a whole comprising those representations (see also A77/B103). What subjects cognize through their self-consciousness is that thinking is a matter of

https://doi.org/10.1017/S1369415424000128 Published online by Cambridge University Press

combining representations and grasping them in a single thought and that they are intelligences who are:

> conscious solely of [their] power of combination. (B158-9)

They also understand, albeit inchoately, what thinking requires. Using the first person in which the Deduction is expressed, the representations combined in a single thought must all belong to me in the sense that they are all combinable by me. Given the Deduction, it is clear why Kant would believe that anyone who thinks carefully about the requirements of cognition would know the claim 'I am the [common] subject of thought' is necessarily true. What about the other part of the premise, the claim that I, as subject of thought, cannot be thought as a determination of something else?

Taking the obvious case, what would rule out thinking 'Humans are thinking selves', where humanity would be the determining ground of the property of being a thinking self? What the Deduction highlights and the preamble to the Paralogisms reiterates is that not only is there no intuition of a permanent subject of thought but also no intuition of anything that would distinguish a thinking self from any other thing (A346/B404). For that reason, the proposed cognition 'humans are thinking selves' is impossible, because it purports to explain how subjects of thought exist through their inherence in human beings – but there are no properties to be explained. Kant's distinctive and devastating objection to any claim of the form 'X's are thinking selves' is not that this or that proposed *explanans* fails, but that the *explanandum* is empty. That is why thinking beings cannot be 'thought' otherwise than as subject, because thinking beings are associated with no intuitive or conceptual properties. Although I take this to be Kant's insight, he would see it as available to his contemporaries who recognized the absence of the intuition of a self and the fact that consciousness supplies no properties of its own; it is 'diaphanous'. Given these widely acknowledged facts, his contemporaries should also know that the thinking self cannot be a determination of something else.

Yet, 'thinking being' or 'I' must be used as the subject of a judgement to indicate the necessary unity of self-consciousness that is required for judgement. As Kant says in the note, 'thought' is ambiguous, because in the major it refers to thought about things, whereas in the minor, 'thought' indicates the requirement of thought for a subject (the thinking being). It seems to me that this way of construing the argument not only explains why Kant could take the minor premise to be known and necessarily true but also makes the confusion seem reasonable even for sophisticated thinkers: the thinking being cannot be conceived as a property of something else, yet the thinking being must be able to be used as a subject in judgements, so it seems to fit the definition of a substance.

5. Why the illusions are inevitable

I conclude with a further *desideratum* for interpreting the First Paralogism. As Proops notes, as with all transcendental illusions, the Paralogism about the soul as substance is supposed to be inevitable due to the demands of reason for completeness (pp. 92–3). One way to appreciate why Kant believed that the Paralogisms posed an ever-present danger to theorizing about the self is to recognize that the transcendental psychology

https://doi.org/10.1017/S1369415424000128 Published online by Cambridge University Press

of the Deduction has important similarities to Wolff's project of Rational Psychology. Wolff tries to derive the real essence of the soul from the operations of mental powers through the intermediate step of reducing the operations to the fewest basic faculties. Kant tries to explain the possibility of empirical cognition in terms of the fewest basic faculties needed to account for its manifest features of locating objects in space and time, of judging them under concepts, and of making further judgements through reasoning. Beyond the overlap of their projects, Wolff was seeking a single basic power that would explain all mentality (including cognition); Kant argues that all cognition can be traced to the 'root' power of apperception (A114). In the B edition, he characterizes the principle of the unity of apperception as the supreme principle governing all cognition. In a passage already cited, Kant notes that thinkers are intelligences who are conscious solely of their powers of combination; in the Second Analogy, he observes that activity is taken to imply a force or power – and so substance (A204-5/B250). Drawing these claims together, the transcendental psychology of the Deduction would establish something that looks quite similar to what Wolff was seeking: All cognition is based on a single fundamental power, which is governed by a single principle, and which therefore might be thought to reside in a substance.

For Kant, the paralogistic inferences are inevitable because the job of philosophy is to investigate the necessary conditions for the possibility of cognition and reason's demand for complete intelligibility misleads philosophers about what they have discovered. Although Kant has found the most basic power in the sense of the power with which all other powers must be in accord for cognition to be possible, he has not found a power that explains the existence of all the other powers. Although he has found a principle that governs all cognition, he has not found a substance whose existence is the ground of the operation of that principle. Not understanding the difference between a 'dissection' of the power of cognition into its necessary elements (cf. A65/B70) and an explanation through the natures of entities, Rational Psychologists fall into the error of believing that they have met the demand of reason for complete intelligibility. Thus, they move from Kant's modest conclusions to extravagant and baseless claims about simple substances. Rational Psychology and the Transcendental Deduction are complementary contexts for understanding the Paralogisms, because the comparison between what scientific metaphysics can achieve (presented in the Deduction) and the goals and methods of Rational Psychology is necessary for appreciating the nature and allure of the Paralogistic claims.

Proops cites an important passage in the *Prolegomena* where Kant reflects on a different way that the idea that the self is a substance might arise:

> It has long been observed that in all substances the true subject – namely that which remains after all accident have been removed – and hence the substantial itself, is unknown to us . . .
>
> Now it does appear as if we have something substantial in the consciousness of ourselves (i.e., in the thinking subject), and indeed have it in immediate intuition; for all the predicates of inner sense are referred to the **I** as subject, and this **I** cannot again be thought as the predicate of some other subject. It therefore appears that in this case completeness in referring the given

https://doi.org/10.1017/S1369415424000128 Published online by Cambridge University Press

> concepts to a subject as predicates is not a mere idea, but that the object, namely, the **absolute subject** itself, is given in experience. But this expectation is disappointed. (4: 333–4)

I agree with Proops about why a cognizer might believe herself to have found a true substance in her own mind: she is aware of states through inner sense that must be grounded in some substance and there is no constant thing in which the states might inhere, so their ground must be the *substantiale*, the propertyless bearer of properties. I also agree with Proops about the source of the difficulty: there is no intuition of the I, but merely an 'intellectual consciousness' of both our activity and our existence. As Proops notes, this intellectual consciousness has 'no official place in [Kant's] ... epistemology' (p. 84).

The difficulties of representing the I that thinks are also on full display in the B Deduction when Kant says that, in the original unity of apperception,

> I am not conscious of myself as I appear to myself, nor as I am in myself, but am conscious only that I am. This representation is a thinking, not an intuition. (B157)

The problem arises because of how we cognize the unity of apperception. As Kant puts it in A:

> We are conscious a priori of the thoroughgoing identity of ourselves in regard to all representations that can ever belong to our cognition, and are conscious of it as a necessary condition for the possibility of all representations. (A116)

Proops is right that an intellectual consciousness does not fit into Kant's classification of representations as either intuitions or concepts. Since it has no intuitive and no conceptual content, the representation 'thinker' is empty and cannot function in the predicate position of a judgement of the form 'X's are thinkers'. That seems to me a more genuinely Kantian explanation of why 'I' cannot appear as a predicate than its being a singular representation. Given Proops's criticisms of Kant's attempts to fashion a representation of the 'I think', perhaps he would agree.

Notes

1 *The Fiery Test of Critique. A Reading of Kant's Dialectic* (Oxford University Press 2021); citation throughout by page number only.

2 Although that is changing recently. See e.g. Marcus Willaschek, *Kant and the Sources of Metaphysics* (2018).

3 Sellars notes that the arguments of the Paralogisms share:

> many of the obscurities of the transcendental deduction categories – to which it is closely related – and has been subject of almost as many controversies and misunderstanding. Thus, the concept of the 'transcendental unity of apperception' plays a key role in both the Deduction and the Paralogisms, and ... is, perhaps, the central concept of the *Critique* ... (Sellars 1970: 6)

https://doi.org/10.1017/S1369415424000128 Published online by Cambridge University Press

Strawson introduces his discussion of the Paralogisms by noting that:

> It will be helpful first to remind ourselves of the significance of the idea of the necessary unity of consciousness, or the unity of apperception, as this emerged from the discussion of the Transcendental Deduction. What was the transcendental unity of consciousness required for, and what did it require? (Strawson 1966: 163).

4 Translations from CPR are from Pluhar (Kant 1996).
5 Translation from Ameriks and Naragon (Kant 1997).
6 Translation from Pluhar (Kant 1987).

References

Ameriks, Karl (1982/2000) *Kant's Theory of Mind*. Second Edition. Oxford: Clarendon Press.
Bennett, Jonathan (1974) *Kant's Dialectic*. Cambridge: Cambridge University Press.
Carl, Wolfgang (1989) *Der Schweigende Kant: Die Entwürfe zu einer Deduktion der Kategorien*. Göttingen: Vandenhoeck & Ruprecht.
Dyck, Corey W. (2014) *Kant and Rational Psychology*. Oxford: Oxford University Press.
Kant, Immanuel (1987) *Critique of Judgment*. Tr. Werner Pluhar. Indianapolis IN: Hackett.
—— (1996) *Critique of Pure Reason: Unified Edition*. Tr. Werner Pluhar. Indianapolis IN: Hackett.
—— (1997) *Lectures on Metaphysics*. Trans. and ed. Karl Ameriks and Steve Naragon. Cambridge: Cambridge University Press.
Sellars, Wilfrid (1970) "'. . . this I or he or it (the thing) which thinks'". *Proceedings of the American Philosophical Association*, 44, 5–31.
Strawson, P. F. (1966) *The Bounds of Sense*. London: Methuen.
Willaschek, Marcus (2018) *Kant and the Sources of Metaphysics*. Cambridge: Cambridge University Press.

Cite this article: Kitcher, P. (2024). Understanding the First Paralogism: A Friendly Disagreement. *Kantian Review* **29**, 289–298. https://doi.org/10.1017/S1369415424000128

https://doi.org/10.1017/S1369415424000128 Published online by Cambridge University Press