Camilla Serck-Hanssen
Fighting Achilles

Introduction

In the first book of the *Transcendental Dialectic* we find Kant's criticism of four syllogistic proofs about the soul. On his view the proofs are formally invalid. Hence, as opposed to the view of their proponents – the rational psychologists – the inferences do not establish: (i) that the soul is a substance; (ii) that the soul is simple; (iii) that the soul is a person; (iv) that the soul can exist independently of material beings. Kant calls these fallacious proofs transcendental paralogisms.

The second paralogism, as presented and discussed by Kant in the A-edition of the *KrV*, is particularly intriguing and difficult. It is particularly intriguing because it is the most powerful and convincing of the paralogisms. It is to wit “the Achilles of all the dialectical inferences of the pure doctrine of the soul, nothing like a mere sophistical play that a dogmatist devised in order to give his assertions a fleeting plausibility, but an inference that seems to withstand even the sharpest testing and the greatest scruples of inquiry.” (*KrV* A351) It is particularly difficult because Kant's dismissal of it is notoriously complex and suggests that the rational psychologist makes a number of quite different kinds of mistakes. Thus it is unclear how Kant's critique of the second paralogism in the A-edition is supposed to fit with his general description of a transcendental paralogism in the same edition. The aim of this paper is to show that one can nevertheless extract a line of criticism against the rational psychologist, which shows that the rational psychologist involves himself in a formal fallacy that accords with the general description of a transcendental paralogism.

My argument will be developed in five steps. In the first step I argue that both of the two standard ways of rendering the logical form of the second paralogism in the A-edition are problematic since they overlook that the middle term of the proof involves a negation. Since for Kant negation can be of two kinds: finite or infinite, I present the hypothesis that the second paralogism is a formally invalid inference because of an ambiguity pertaining to the logical operation of negation. In the second step, I present the historical background of infinite negation and Kant's version of it. In both versions, the crucial difference is that only mere negation (i.e., negative judgment in Kant's parlance) is compatible with empty terms. In the third step, I use the distinction between negative and infinite judgment to present four different possible interpretations of the second paralogism. In the fourth step, I argue that according to Kant's reconstruction of the rational psychologist's argu-

ment he (the rational psychologist) in fact helps himself to a valid form of the proof that has an infinite judgment in the Minor. I then suggest that Kant's critique of the rational psychologist is that he is not *justified* in using this valid form. Thus, contrary to his own assumption, the rational psychologist has at his disposal only the formally invalid form of the inference with a negative Minor. In the fifth step, I first show that my reading accords with Kant's general description of a transcendental paralogism. I then argue that the error which underlies and indeed conceals the syllogistic fallacy, is that the rational psychologist assumes but has not shown that the soul as an object of pure rational psychology exists.

Step one: The ambiguous form of the Second Paralogism in the A-edition

My aim in this part is to suggest a new way of understanding the ambiguity inherent in the second paralogism in the A-edition of the *KrV*.

The second paralogism reads:

That thing whose action can never be regarded as the concurrence of many acting things,
is **simple**

Now the soul, or the thinking I, is such a thing

Thus etc. (*KrV* A351)

Adding the words Kant left to be filled out we get the full inference:

That thing whose action can never be regarded as the concurrence of many acting things,
is **simple**

Now the soul, or the thinking I, is such a thing whose action can never be regarded as the
concurrence of many acting things

Thus the soul, or the thinking I is simple

In the commentary literature there are basically two approaches to the logical form of this inference. Some readers argue that it is clearly a valid three term categorical syllogism:¹

Every S is P

a is S

a is P

¹ Cf. Van Cleve 1999: 173; Bennett 1974: 72–73.

Others argue that despite appearances the second paralogism Kant holds that contains an ambiguous middle term.² As such its form is really an invalid four term syllogism:

Every S is P
 a is R
 a is P

Both readings can be challenged. The main problem with the first reading is that it goes against Kant's explicit claim that the second paralogism *qua* being a paralogism is a formal *fallacy* (KrV A402). The second reading seems more promising, since it complies with Kant's claim that the syllogistic inference is fallacious. However, since on this reading the fallacy is construed as an ambiguity in the *meaning* of the middle term, it is still not clear why Kant would describe this as a *formal* fallacy. Meaning, it seems, would have to do with the conceptual content of the terms rather than form. My main reason for suggesting a third way of analyzing the second paralogism is however that both approaches overlook an interesting feature of the premises, namely that the middle term is a compound that contains a negation: "that thing whose action can *never* be regarded as the concurrence of many acting things".³ What complicates the question concerning the logical form of the inference is that according to Kant there are two different kinds of negation, represented respectively by the operations of infinite and negative judgment. As we will see in more detail below, the truth of a negative judgment is compatible with an empty subject term, while the truth of an infinite judgment is not. So which one does the rational psychologist apply? Or does he maybe apply both?

On this basis I present the following hypothesis: The second paralogism is a formally invalid inference because of an ambiguity pertaining to the logical operation of negation in the middle term. This hypothesis can already be supported by the following considerations: I) Since their truth conditions are different, an inference that fails to distinguish the two forms of negation can be formally invalid; II) Since both kinds of negation are logical operations, it would be proper to describe their equivocation as a formal fallacy and not simply as a conflation of meaning.

² Cf. Grier 2001: 163n.; Allison 2004: 335–337. Admittedly, with respect to the formal fallacy both Grier and Allison only discuss this explicitly in relation to the first paralogism. Nevertheless, their analysis follows the same pattern for all the paralogisms. Cf. also Wuerth 2010: 248–249; Wunderlich 2001: 178.

³ This negative character of the middle term in the paralogism is overlooked by almost all commentators. An exception is Toni Kannisto, who makes this point in his 2012.

Step two: Two kinds of negation

The idea that there are two kinds of negation, one that negates propositions and another that negates terms, stems from Aristotle. The application of infinite negation became widespread from the 12th century onwards, and its logic became a topic of much discussion. Medieval logicians attempted to formulate principles of conversion by contraposition for infinite negation, for instance, the principle that you can change a proposition from a negative to an affirmative if you change the predicate term from finite to infinite. However, as Buridan showed, many of these principles failed in cases of empty subject terms.⁴ E.g. “Some S is not P” does not equal “Some S is non-P” if S is empty. To give an example: “It is not the case that the greatest prime number is even” does not equal “The greatest prime number is non-even”. While the first is true – since the greatest prime number is nothing at all, it is also not even – the latter is patently false by implying that the greatest prime number – a non-entity – is odd.

Kant’s distinction between negative and infinite judgment seems to be indebted to this medieval discussion, and in particular to Buridan’s restriction on the use of empty terms in infinite negation. Just how it was handed down to Kant and in what version, I do not know, but it is striking that in the canonical logic book of his time, *the Port Royal Logic*, there is no discussion of empty terms or the problems they can cause in relation to negation.⁵ But let us now turn to Kant’s own discussion of the two forms of negation.

In the *KrV*’s table of judgments Kant presents three judgmental forms under the moment of quality: affirmative, negative and infinite judgment. In adding infinite judgment as “a special member of the classification”, his table seems to deviate as he says from general logic (*allgemeine Logik*), which counts the infinite form simply as an affirmative.⁶ In his commentary to the table Kant gives us the reason for this deviation.

The first point to notice is that Kant does not seem to have anything against the ordinary practice of general logic as such. That is, as long as the logicians consider only the pure form of thinking in general and abstract from the content of the concepts and their relation to objects, infinite judgments can be counted among the affirmatives.

⁴ For a helpful presentation of this see Parsons 2012.

⁵ Cf. Ashworth 1978: 147. Referred to in Parsons 2012.

⁶ The same holds for the third member of the moment of quantity, since logically speaking there is no difference in assigning a predicate to a subject concept whose extension is one single object and assigning it to all members of the class of subjects (Cf. *KrV* A71–3 / B97–8).

Nevertheless, from the point of view of transcendental logic the picture is different. Transcendental logic does not and indeed should not abstract from questions concerning the content of thought and the relationship between thoughts and objects. To the contrary, its task is precisely “to expound the elements [the intuitions and concepts] and principles without which no object can be thought at all” (*KrV* A62 / B87). This means that a theory that attempts to examine logical functions with a view to their involvement in transcendental logic – which arguably is what Kant is up to in the table of judgment – must do so by taking into consideration distinctions that are of crucial importance beyond the domain of mere general logic.

This explains why infinite judgment *cannot* be taken simply as an affirmative judgment in Kant’s table of judgment. For the application conditions of this kind of judgment are weaker than for the affirmative type, which requires access to determinate positive properties or predicates of the object. Nevertheless, an infinite judgment does require a relation to an object (the extension or referent of the subject term of the judgment). To see what this means, let us however first look at the second judgmental form in this group (i.e. quality), namely *negative* judgment.

In the *Jäsche Logic* Kant tells us: “[I]n the negative [judgment] the subject is posited outside the sphere of the [predicate]” (*Logik* 9:104). That is, a negative judgment tells us what anything falling under the subject concept is not, but this is compatible with the object not belonging to any domain at all. As such Kant fully agrees with the view that became dominant from late medieval times, namely that negatives are compatible with empty subject terms. His point is however not the logical argument for why this must be the case, namely that it is required to keep the square of opposition coherent.⁷ His focus is instead on the logical operations involved in judgment. Both in the *Jäsche Logic* and in the *Dohna-Wundlacken Logic* Kant claims that in a negative judgment the negation affects the copula (*Logik* 9:104n.; *Log-DW* 24:764). This means that the truth of a negative judgment taken in isolation (i.e., if one has no other grasp of the object *x* than it not having some property *F*) severs itself from any metaphysical questions concerning the possible existence of *x*. In doing so it is significantly different from infinite judgment.

⁷ The argument for why this must be true is the following:

Suppose that ‘*S*’ is an empty term; it is true of nothing. Then the I form: ‘Some *S* is *P*’ is false. But then its contradictory E form: ‘No *S* is *P*’ must be true. But then the subaltern O form: ‘Some *S* is not *P*’ must be true. But that is wrong, since there aren’t any *Ss*.

As is easily seen, the problem disappears the moment one changes the truth conditions for O so that it is vacuously true when *S* is an empty term.

This argument is given e.g. in Kneale / Kneale 1962: 55–60, and cited in Parsons 2012.

First, both in the *Jäsche Logic* and in the *Dohna-Wundlacken Logic*, we are told that in an infinite judgment it is not the copula but the predicate that is affected by the negation (*Logik* 9:104n.; *Log-DW* 24:764). As such Kant follows the Aristotelian and medieval view that infinite judgment involves a negation of a term or predicate, rather than the negation of a proposition. Second, in the case of infinite judgment there is not only an act of removing but also indeed an act of positing the subject *in* a sphere, albeit the infinite one outside the sphere of the predicate that the judgment negates. In other words, as opposed to negative judgment infinite judgment is an act that preserves the claim that *x* is something. It says in effect that (all, some or the) *x* is non-*F*. Thus the employment of this form commits one to an act of positing the subject in a domain (outside of *F*). Consequently, just like Buridan taught us, infinite judgment does not allow empty subject terms.

Step three: Four possible interpretations of the second paralogism

As already argued above, the standard logical interpretations of the second paralogism share one fault, they overlook that the inference contains a negation in connection with the middle term:

That thing whose action can *never* be regarded as the concurrence of many acting things, is **simple**

Now the soul, or the thinking I, is such a thing whose action can *never*
be regarded as the concurrence of many acting things

Thus the soul, or the thinking I is simple

Since in transcendental logic negation can be taken in two different ways, namely as negative or as infinite judgment, the question is how we are to understand the middle term. It can be either be a compound where the negation pertains to the copula (operation of negative judgment). This yields the following reading of the middle term:

- a) is not a thing whose action can be regarded as the concurrence of many acting things

or it can be a compound where the negation pertains to a – in this case complex – term or predicate (operation of infinite judgment). This yields reading:

- b) non-“thing whose action can be regarded as the concurrence of many acting things”

When read through the lenses of first order logic⁸ the second paralogism can therefore at least in principle be read in four different ways via different combinations of a) and b) in the major and minor premise.

I) a) in the Major and a) in the Minor

For all things, if it is not a thing whose action can be regarded as the concurrence of many acting things, then it is simple

Now the soul, or the thinking I, is not a thing whose action can be regarded as the concurrence of many acting things

Thus the soul, or the thinking I is simple

II) b) in the Major and b) in the Minor

For all things, if it is non-thing whose action can be regarded as the concurrence of many acting things, then it is simple.

Now the soul, or the thinking I, is non-thing whose action can be regarded as the concurrence of many acting things

Thus the soul, or the thinking I is simple

III) a) in the Major and b) in the Minor

For all things, if it is not a thing whose action can be regarded as the concurrence of many acting things, then it is simple.

Now the soul, or the thinking I, is non-thing whose action can be regarded as the concurrence of many acting things".

Thus the soul, or the thinking I is simple

8 By "through the lenses of first order logic" I mean that I read it with the help of first order logic, which brings out that there can be four possible readings. Still I do not attempt to formalize the inferences in first order logic. This would only lead to confusion since this system has a different understanding of the universal quantifier and also does not operate with infinite negation. If one prefers to remain in the system of syllogistic logic there are however only two possible readings. Since in that case the middle term in the Major does not contain a copula, the negation in the Major can only be read as negating the complex "thing whose action can be regarded as the concurrence of many acting things". The Minor would however still lend itself to the two readings of the middle term. One would therefore have:

(II)* All non-S are P

a is non-S,

a is P

(IV)* All non-S are P

a is not S

a is P

IV) b) in Major and a) in the Minor

For all things, if it is non-thing whose action can be regarded as the concurrence of many acting things, then it is simple.

Now the soul, or the thinking I, is not a thing whose action can be regarded as the concurrence of many acting things.

Thus the soul, or the thinking I is simple

But which of these forms does the second paralogism have according to Kant's analysis? To decide on that issue we need to look into Kant's reconstruction of the rational psychologists arguments as well as his subsequent criticism.

Step four: What is the form of the second paralogism?

Most commentators have noticed that Kant does not really discuss the major premise of the syllogism. Its form can however be found by considering the way Kant constructs the indirect proof in support of the second paralogism. As has been noticed by Van Cleve,⁹ Kant replaces the original Major of the paralogism with the contrapositive proposition “if something is non-simple, its action can be regarded as the concurrence of many acting things”. This shows that Kant understands also the original Major as employing the infinite kind of negation. We are thus left with two possibilities which both use the infinite form in the Major:

II)

For all things, if it is non-thing whose action can be regarded as the concurrence of many acting things, then it is simple.

Now the soul, or the thinking I, is non-thing whose action can be regarded as the concurrence of many acting things

Thus the soul, or the thinking I is simple

or

IV)

For all things, if it is non-thing whose action can be regarded as the concurrence of many acting things, then it is simple.

Now the soul, or the thinking I, is not a thing whose action can be regarded as the concurrence of many acting things.

Thus the soul, or the thinking I is simple

⁹ Van Cleve 1999: 176.

Notice that while II) is valid, IV) is not, for from a negative premise¹⁰ no affirmative conclusion can follow (Cf. *Logik* 9:124, rule 1 and 5). But which of the two presents the correct reading of the second paralogism? Recall that according to Kant, the rational psychologist certainly *believes* that he is using a logically valid form. But Kant still holds that the rational psychologist is involved in an invalid inference. Kant's point is hardly that his opponents believe that an affirmative conclusion can follow from a negative premise. The problem I surmise must rather be that the rational psychologist's *arguments* for his inference only suffice for the invalid form IV), nevertheless he helps himself to II). To support that reading, I first turn to the rational psychologist's argument behind the second paralogism to tease out more precisely what kind of assumption is at stake in the rational psychologist's line of reasoning.

Kant presents the argument on the form of a *reductio*:

Every composite is an aggregate of many, and the action of a composite... is an aggregate of many actions or accidents... [S]uppose that the composite were thinking; then every part of it would be a part of that thought, but the parts would first contain the whole thought only when taken together. Now this would be contradictory. For because the representations that are divided among different beings (e.g. the individual words of a verse) never constitute a whole thought (a verse), the thought can never inhere in a composite as such. This is only possible in one substance, which is not an aggregate of many, and hence it is absolutely simple. (*KrV* A352)

For our present purpose the only thing we need to focus on in this argument is its logic. As I have already pointed out, Kant replaces the Major of the paralogism with the contrapositive proposition “if something is non-simple, its action can be regarded as the concurrence of many acting things”. The rational psychologist then proceeds to prove the Minor, via a *reductio*: The supposition that the non-simple (i.e. the composite) is thinking is contradictory. Hence, the Minor is true: the soul, or the thinking I, is such a thing whose action *cannot* be regarded as the concurrence of many acting things. He then moves by means of *modus tollens* to “the soul is not non-simple” which he identifies with “the soul is (absolutely) simple”:

Major: If something is non-simple, its action *can* be regarded as the concurrence of many acting things.

Minor: The soul or thinking I *cannot* be regarded as the concurrence of
many acting things

Conclusion: The soul is simple.

10 Recall that logically speaking, infinite judgment is an affirmative, so that form of negation does not lead to the same problem.

In the argument presented above, it is assumed that the affirmative judgment “the soul is simple” can be derived by means of *modus tollens*. This is however only possible if “soul” is not an empty term, for only in that case is “the soul is not non-simple” equivalent “the soul is simple”. Although it is possible that Kant takes his opponents not to be aware of this restriction and thus commit a logical error, I believe that his presentation and discussion of the argument behind the second paralogism suggest that he rather takes the rational psychologist to assume that “soul” is not an empty term.

If that is the correct diagnosis, it shows that if the rational psychologist cannot justify his assumption that “soul” is not an empty term, his argument would fail to establish the truth of the conclusion. But it also shows something more. Since infinite judgments are not compatible with empty subject terms, the rational psychologist would only be justified in using the invalid form of the second paralogism namely IV) which employs a negative Minor.

Step five: Why the proof for the simplicity of the soul is a transcendental paralogism.

An easy and elegant way of supporting my reading would be to argue that it fits because Kant takes himself to have shown already that the soul is not a genuine object but only a transcendental illusion.¹¹ As such, despite appearances, “soul” is no better off than “largest prime number”, for they are both empty. The emptiness of “soul” would require the negative form of the Minor. Hence, despite the hopes and beliefs of the rational psychologist the argument at his disposal would be the fallacious version IV). Moreover, on that form the second paralogism would indeed itself serve as a counterexample. For if the soul is nothing but an illusion, the premises would be true¹² and the conclusion false.

Nevertheless, I do not believe that an attempt to reconstruct Kant’s criticism along these lines is the way to go. The reason is that I do not think that Kant does show in the first part of the Dialectic (or elsewhere for that matter) that the soul cannot be anything but an illusory object. It falls beyond the scope of this paper to show why I reject such a reading. What I do need to show however is that

¹¹ Such a reading would seem to fit with Grier’s view (taken over by Allison) that the soul is only an illusory object of the corresponding idea of reason.

¹² Recall that if “soul” is empty, then the Minor in IV is vacuously true. If the soul is nothing at all, it is also true that it is not the case that it can be regarded as the concurrence of many acting things.

Kant's diagnosis of what goes wrong in the second paralogism can be read the way I suggest, without assuming that he takes for granted that he has already shown that "soul" is an empty term.

Let me begin to support my reading of the fallacy by showing that it fits with Kant's general description of a transcendental paralogism in the A-edition according to which it is (i) a *sophisma figurae dictionis* (KrV A402); (ii) "in which the major premise makes a merely transcendental use of the category, in regard to its condition, but in which the minor premise and the conclusion, in respect of the soul that is subsumed under this condition, make an empirical use of the same category" (KrV A402f).

I have argued that in the Major of the second paralogism the negation is of the infinite kind that pertains to a (complex) term. In the Minor, the rational psychologist is however only justified in using negative judgment, which is the negation of the copula. If this is correct, there is indeed a kind of *sophisma figurae dictionis*, although not of the common type where the problem is the ambiguity in the meaning of the middle term. The ambiguity lies instead in the *logical operation of negation*. In the Major it is the term that is affected, in the Minor, however, it is the copula.

One might want to argue against this reading of the ambiguity by pointing out that it does not appear to agree with Kant's use of "*sophisma figurae dictionis*" in the *Jäsche Logic*, according to which it is an inference in which "the *medius terminus* is taken in different meanings (*Bedeutung*)" (*Logik* 9:135). Notice however that in the KrV Kant formulates himself quite carefully, and says only that "if one wants to give a logical title to the paralogism... then it can count as a *sophisma figurae dictionis*" (KrV A402). This strongly suggests that he uses a well-established term to describe a new phenomenon to which he believes the term nevertheless can be appropriately applied.

My understanding of the ambiguity in the proof of the second paralogism also fits very well with Kant's more detailed second description of a transcendental paralogism: "the major premise makes a merely transcendental use of the category, in regard to its condition, but in [...] the minor premise and the conclusion, in respect of the soul that is subsumed under this condition, [one] make[s] an empirical use of the same category" (KrV A402–403). The point is not that the category (or rather the term that correlates to it) is an *ambiguous middle term*.¹³ In fact, in connection with the first paralogism Kant says explicitly that it is one and

¹³ This reading creates a problem since "simple" is not the middle term of the syllogism. The problem is discussed among others by Wunderlich 2001: 177. Since on my reading the category is not supposed to be the ambiguous middle term, this issue need not be addressed here.

the same concept (substance) that is used in the paralogistic proof (*KrV* A403).¹⁴ Rather, what is ambiguous is the *use of a concept in regard to its condition*. Now the condition in the Major is the middle term. As we have seen, this condition must be taken as the infinite form of negation, which implies that the corresponding domain is not empty. Since there is no other constraint on these objects than that they have the negative property: non-thing whose action can be regarded as the concurrence of many acting things, the category (i.e. the concept simple¹⁵) is applied to objects in general in the Major, i.e., it is applied transcendently.¹⁶

However, according to Kant, the problem is that:

[I]n the minor premise and the conclusion, in respect of the soul that is subsumed under this condition, [one] make[s] an empirical use of the same category. (*KrV* A402f.)

At first this diagnosis of the fallacy looks odd. If indeed the Major is true of all things in general, it is certainly also true of empirical objects, which is a subset of the extension of “thing in general”. Hence, if the soul is correctly subsumed under the condition of the Major, the application of ‘simple’ to the soul must be warranted. However, a few lines later Kant says that the empirical use of the category is in fact illicit (*unzulässig*) (*KrV* A403).¹⁷ This accords well with the reading I argue for in this paper, since it takes Kant to hold that although the rational psychologist *in fact* applies the form of the Minor that would make a valid proof, he is nevertheless not *justified* in using that form, for he has not shown that “soul” is not an empty term. And if it is an empty term, its purported object soul would not in fact appear in the domain of objects in general or any of its subsets. In this case, the conclusion that the soul is simple will not follow.

Now Kant’s claim that there is an illicit empirical use of the category contains a clue, which adds to our understanding the sense in which the rational psychologist lacks justification. It is a necessary condition of the legitimate empirical use of the category that the object to which it is applied exists. The phrase “illicit empirical use” therefore strongly suggests that the underlying error of the rational psychologist is that he assumes but somehow fails to support the assumption that soul exists. Notice that if we could grant him this existential assumption he

¹⁴ For this point and the reference I am indebted to Toni Kannisto.

¹⁵ Admittedly it is not clear from the discussion in the paralogism itself why or how ‘simple’ corresponds to the category of reality. Kant claims that he can only show this in the Antinomies (*KrV* A404n.). But that need not concern us here.

¹⁶ Cf. A238: “The transcendental use of a concept in any sort of principle consists in its being related to things *in general* and *in themselves*”.

¹⁷ Unfortunately the Cambridge Edition makes a mistake here and translates “*unzulässig*” with “unreliable” (i.e. conflates it with “*unzuverlässig*”) whereby Kant’s point is lost to view.

could move validly to his substantive metaphysical conclusion that the soul is simple. Obviously if the soul exists, “soul” is not an empty term. This suggests that the vulnerable heel of Achilles lies in the rational psychologist’s unjustified assumption that the soul exists. In what remains of the paper I shall attempt to substantiate this suggestion.

The first piece of evidence in favor of this thesis¹⁸ is found in some recent historical studies of the rational psychologists. As has been shown by Corey Dyck, the point of departure of Kant’s actual opponents – namely Wolff and his followers – was indeed some very general *empirical* claims about the experience we actually have of our self as an inner object.¹⁹ Such claims would arguably include claims about the existence of a self. The problem is not that such claims need to be seen as unjustified in the sense that they are simply taken for granted. Indeed the “father” of rational psychology, Wolff, argues for existence of the I (or we) by means of the following syllogism:

Whoever is conscious of himself and of other things, is. We are conscious of ourselves and other things. Therefore, we exist. (*DM* § 6)²⁰

As Dyck has convincingly shown, the problem is however that for Kant no empirical starting point of this kind can ever yield genuine metaphysical knowledge about the soul. Rational psychology proper must restrict itself, for

[i]f the least bit of anything empirical in my thinking, any particular perception of my inner state, were mixed among the grounds of cognition of this science, then it would no longer be a rational but rather an *empirical* doctrine of the soul. (*KrV* A342 / B400)²¹

From this it also follows that the actual rational psychologists were in fact wrong in thinking that their arguments and approach sufficed to justify the assumption that the object of rational psychology proper exists. Therefore, their arguments also failed to show that “soul” (in the sense required by rational psychology proper) is not empty.

To undercut the possibility of a pure science of the soul Kant must however go beyond a critique of actually presented proofs, he must indeed show that there are *no other possible alternatives*. If my reading is correct, this second and more

18 The idea that the fundamental problem of the paralogisms lies in their assumption that the soul exists is to my knowledge first made by Toni Kannisto, Cf. Kannisto 2012, part III. On this point my reading is very much indebted to his original analysis.

19 Cf. Dyck 2009: 249–75.

20 Quoted in Dyck 2009: 252 (*DM* = Wolff, *Vernünfftige Gedanken*, 1720).

21 Also quoted by Dyck 2009: 249.

ambitious line of criticism must include a line of argument that shows that even the arguments that the rational psychologist could have launched would have failed to justify the assumption that the soul as an object of pure rational psychology exists. Kant's criticism of the argument for the second paralogism is notoriously complex. Nevertheless I believe we can extract a line of argument, which supports my suggestion that at least part of what goes wrong is that the assumption that the soul exists cannot be justified. So let us now return to the argument for the second paralogism and Kant's criticism of it.

Recall that the argument for the second paralogism is presented by Kant as follows:

Every composite is an aggregate of many, and the action of a composite... is an aggregate of many actions or accidents... [S]uppose that the composite were thinking; then every part of it would be a part of that thought, but the parts would first contain the whole thought only when taken together. Now this would be contradictory. For because the representations that are divided among different beings (e.g. the individual words of a verse) never constitute a whole verse, the thought can never inhere in a composite as such. This is only possible in one substance, which is not an aggregate of many, and hence it is absolutely simple. (KrV A352)

According to Kant the *nervus probandi*, i.e., the crux of the argument is “that many representations have to be contained in the absolute unity of the thinking subject in order to constitute one thought” (KrV A352). Kant argues that there are only three alternative grounds by means of which the rational psychologist could know the proposition in question and they all fail (KrV A353).²² The attempt to prove the proposition from mere concepts, i.e. analytically, is ruled out since the proposition aims to say something about that which is causally responsible for the thinking,²³ an aim that can never be reached by analyzing a mere concept. The synthetic a priori approach is also ruled out since the case at hand does not comply with the conditions of the possibility of such proofs. Finally, the experiential route is ruled out since it cannot yield results for rational psychology proper.

For my present purpose I neither need to discuss whether Kant is right in seeing these as the only alternatives, nor do I have to discuss whether his way of dismissing them is acceptable. What I need for my argument to work is to note *that* he dismisses all of the alternatives and to show that his explanation for why

²² As Allison notes, in line with his critical method Kant's attack is aimed at the *grounds* of the proof of this proposition, rather than at the proof itself, cf. Allison 2004: 342n. Hence, Kant himself can assume simply an agnostic position about the existence of the soul and its properties.

²³ In reading it as a causal claim I follow Allison 2004: 343.

pure rational psychology must fail is indeed connected to its inability to show that its object (the soul) exists. I now turn to support the latter claim.

In the general introduction to the paralogisms we are told that rational psychology proper is “a putative science, which is built on the single proposition *I think*” (KrV A342 / B400). Moreover, the I of the “I think” is merely the *form* of consciousness and not even a concept of an object (KrV A346 / B404). When we turn to Kant’s diagnosis of what goes wrong in the proof for the second paralogism we find the same fundamental idea. What Kant tells us is not only that the three alternative routes (outlined above) for supporting the *nervus probandi* fail, but also that since they all fail,

the formal proposition of apperception **I think**, remains the entire ground on which rational psychology ventures to extend its cognition; this proposition is of course obviously not an experience, but rather the form of apperception, on which every experience depends... (KrV A354)

In other words, since the three possible ways of proving the *nervus probandi* fail, the kernel of truth in the proposition “that many representations have to be contained in the absolute unity of the thinking subject in order to constitute one thought” (KrV A352) is reduced to the principle of apperception. Now in that principle the simplicity of the I (or call it soul if you like) is already contained, i.e., to say that this I is simple does not amount to additional knowledge about “a thinking being in general” (KrV A354), for “**I am simple** signifies no more than that this representation I encompasses not the least manifoldness within itself, and that it is an absolute (though merely logical) unity” (KrV A355). If we believe otherwise, it is because we *unjustly* (*mit Unrecht*) make the I think “which must be regarded as the **merely subjective condition** [of possible cognition in general] into a **concept** of a thinking being in general” (KrV A354).

Hence, as several interpreters have noted, Kant’s point is clearly that the principle of apperception does not yield any metaphysical insight about the real properties of that thing which thinks.²⁴ For the purpose of my argument, however, the most important point is that as a formal principle it also does not show that there even *exists* a thinking I or soul. This seems to be at least part of what Kant is after when he claims:

This much is certain: through the I, I always think an absolute but logical unity of the subject (simplicity), but I do not cognize the *actual* simplicity of my subject. (KrV A356, trans. modified, italics mine)²⁵

²⁴ Cf. among others Bennett 1974: 86; Grier 2001: 167; Kitcher 2011: 7; Rosefeldt 2000: 95.

²⁵ Unfortunately the Cambridge Edition, 420, translates *wirklich* as real instead of actual.

The use of "actual" points to the modal category of existence and strongly suggests that Kant's point is that one cannot move directly, i.e., analytically, from a truth about the logical unity of the I to a truth about the simplicity of an actual, i.e., an *existing* soul. This would of course also fit well with his more general point that 'existence' cannot be contained in another concept, e.g. 'I' (KrV A599 / B627).²⁶ Moreover, the putative existence of this formal I cannot be proven by any experience either, as it belongs to the form "on which every experience depends". Thus, the I of "I think" *qua* form of apperception is clearly not the kind of self that could be proven to exist e.g. by Wolff's syllogism presented above, since that inference was precisely grounded in experience.

Nevertheless, presumably because the rational psychologist in Kant's eyes conflates the logical and formal I of apperception with the actual self of which one is conscious when one has occurrent thoughts and experiences, he also illicitly moves from the logical unity of the former to the actual simplicity of the latter. But although one may "pretend to know that the thinking I, the soul (a name for the transcendental object of inner sense) is simple; nevertheless on this account this expression has no use at all that reaches to *actual* objects, and hence it cannot extend our cognition in the least" (KrV A361, trans. modified, *italics mine*).²⁷ Again the point seems to be not only that one has no experience that allows one to apply the concept of simplicity to soul in a way that would yield cognition of the properties of the object soul,²⁸ but also that one lacks the grounds for asserting that there indeed is, in the sense of exists, an (actual) soul.

So while the rational psychologist can get 'simplicity' from the purely formal "I think" of apperception, this I does not suffice for the other required element, namely that this *I exists as an object* (call it soul). When the rational psychologist nevertheless believes (according to Kant's reconstruction) that he has obtained the metaphysical knowledge that the soul exists as simple, it is presumably because he uncritically lets the existence claim he extracts from empirical knowledge claims about himself ride piggyback on the insight into the purely formal simplicity of the I think, now understood as the I of the principle of apperception. But such a transfer of 'existence' is unjustified.

What I have offered in support of my reading is clearly not a full interpretation of Kant's notoriously complicated discussion of the second paralogism. Nevertheless, I believe it suffices to support my claim that on Kant's view, the

²⁶ For the same point see Rosefeldt 2000: 215.

²⁷ Again the Cambridge Edition, 422, translates *wirklich* as real instead of actual.

²⁸ For the view that the problem is (only) that we cannot apply 'simple' (or any of the other categories) to soul in any metaphysically significant way because we lack the appropriate application conditions for using the schematized categories see among others Wuerth 2010: 236–37.

second paralogism is formally invalid because the rational psychologist assumes, but has not proven, that the soul exists. Hence, for all we know, “soul” could still be an empty term.

Conclusion

I have argued that the only form of the second paralogism (A-edition) that the rational psychologist is justified in applying is one that is formally invalid. In the Major the rational psychologist applies the infinite form of judgment, while in the Minor only the negative form stands at his disposal. Despite the confusing complexity in Kant’s discussion and criticism of the second paralogism, I thus take myself to have shown that it is indeed a formal fallacy, and that it turns on the ambiguous use of the logical operation of negation, not in an ambiguity in meaning or content. Moreover if my reading is correct, the ground of the fallacy is the unjustified assumption that the soul – in the sense of the object of *pure* rational psychology – exists. This existence assumption makes the rational psychologist *believe* that “soul” is not an empty term and hence he also believes that the formally valid form of the syllogistic proof is available to him. Thus the vulnerable heel of Achilles is precisely this underlying existence assumption.

In analyzing the failure of the rational psychologist in this way, I am not ascribing to Kant the view that the second paralogism is itself a counterexample with true premises and a false conclusion. My claim on his behalf is the weaker one, namely that he takes the rational psychologist’s ground for using the infinite form in the Minor to be an *unjustified* assumption that the soul (*qua* object of pure rational psychology) exists. This assumption could, however, still be true. Nevertheless Kant appears to think that this assumption could never be justified, at least not by any theoretical means. In this paper I have taken no stance on whether Kant’s view on this is correct or not. But this much is certain: if the rational psychologist were to find a way to support his assumption that the soul exists, a route which did not comply with the tenets of transcendental idealism, this success would indeed “be a stumbling block to the whole Critique” (KrV B409).

References

- Allison, Henry, 2004, *Kant's Transcendental Idealism*, New Haven / London: Yale University Press.
- Ashworth, E. J., 1978, "Existential Assumptions in Late Medieval Logic", *American Philosophical Quarterly*, 10, 141–47.
- Bennett, Jonathan, 1974, *Kant's Dialectic*, Cambridge: Cambridge University Press.
- Dyck, Corey W., 2009, "The Divorce of Reason and Experience: Kant's Paralogisms of Pure Reason in Context", *Journal of the History of Philosophy*, 47, 249–75.
- Dyck, Corey W., 2014, *Kant and Rational Psychology*, Oxford: Oxford University Press.
- Grier, Michelle, 2001, *Kant's Doctrine of Transcendental Illusion*, Cambridge: Cambridge University Press.
- Kannisto, Toni, 2012, *From Thinking to Being – Kant's Modal Critique of Metaphysics*, Ph.d. dissertation, University of Oslo.
- Kitcher, Patricia, 2011, *Kant's Thinker*, Oxford: Oxford University Press.
- Kneale, William / Kneale, Martha, 1962, *The Development of Logic*, Oxford: Oxford University Press.
- Parsons, Terence, 2012, "The Traditional Square of Opposition", in: Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition),
URL = <<http://plato.stanford.edu/archives/fall2012/entries/square/>>.
- Rosefeldt, Tobias, 2000, *Das Logische Ich: Kant über den Gehalt des Begriffes von sich selbst*, Berlin: Philo.
- Van Cleve, James, 1999, *Problems from Kant*, New York: Oxford University Press.
- Wolff, Christian, *Vernünfftige Gedanken von Gott, der Welt und der Seele des Menschen, Auch allen Dingen überhaupt, Den Liebhabern der Wahrheit mitgetheilet*, Halle, 1720.
- Wuerth, Julian, 2010, "The Paralogisms of Pure Reason", in: Paul Guyer (ed.), *The Cambridge Companion to Kant's Critique of Pure Reason*, Cambridge: Cambridge University Press.
- Wunderlich, Falk, 2001, "Kant's Second Paralogism in Context. The *Critique of Pure Reason* on Whether Matter Can Think", in: Wolfgang Lefèvre (ed.), *Between Leibniz, Newton, and Kant: Philosophy and Science in the Eighteenth Century*, Dordrecht: Kluwer Academic Publishers, 175–88.