Kant's analysis of the Paralogism of Rational Psychology in *Critique of Pure Reason* Edition B

One third of the transcendental dialectic in Kant's *Critique of Pure Reason* is devoted to demolishing the pseudo-science of rational psychology. In this part of his work Kant attacks the idea that there is an ultimate subject of experience - the 'I' or Self - which can only be investigated and understood intellectually. The belief that such a study is possible, is natural to human reason; but it is based on demonstrable error. Kant tries to exorcise our minds from falling prey to this mistake.\(^1\)

An important part of Kant's case against the rationalist dualist is contained in his presentation of a syllogism and his exposure of it as a paralogism - a piece of fallacious reasoning. I think that this section forms a main element in Kant's argument against the mind as subject; and I propose to examine its logical and historical character and its philosophical force.\(^2\) The argument is presented most clearly and succinctly in the Second Edition (B410-13), and characteristically the heart of it is contained in a footnote. It takes the form of exposing as fallacious the rationalist's argument for his conclusion that the self is the ultimate subject. We need to understand the structure of this argument, and the precise nature of the fallacy which Kant detects in it. Only when this has been done, will we be in a position properly to evaluate how effective is Kant's claim here to have demonstrated that rational psychology is a misconceived exercise.

Kant constructs a syllogism which proceeds thus:

*Major Premiss*: That which can only be thought as a subject cannot exist other than as a subject.

*Minor Premiss*: A thinking being, considered as such, can only be thought as a subject.

*Conclusion*: Therefore it can only exist as a subject.

Analogous syllogisms can be used to prove the other main theses of rational psychology - that the thinking being, or soul, is simple, enduring, and exists

---


2 Of modern discussions of Kant's argument, the one which approaches closest to my own is Patricia Kitcher, *Kant's Transcendental Psychology* (Oxford University Press, 1990), especially pp.181-94. Much less satisfactory is Andrew Brook, *Kant and the Mind* (Cambridge University Press, 1994), pp.113-14. Jonathan Bennett, *Kant's Dialectic* (Cambridge University Press, 1974), pp. 72-6, indicates a preference for Kant's A edition version of this material over the B edition, but he does not give adequate reason for this unlikely strategy.
necessarily. In all cases the outer terms - the soul, and the manner of existing - are connected through the middle term - being able to be thought only in a certain way.

On first appearance, the syllogism is valid and its premisses are convincing. Let us consider, first, the truth of the premises. The major premiss may be faulted for its verificationism; the fact that we can only think of things in a certain way, need not entail that they really are so. But Kant would take such criticism to be a sign of transcendental realism, and I agree with him that we must accept as true what we cannot conceive as being otherwise.3

In the minor premiss we have a specific commitment of the rationalist, rather than (as in the major premiss) a metaphysical principle of potentially more general appeal. But again the reasons for accepting the premiss are strong. The soul, as it presents itself to our experience, does have the characteristics which are specified in the various versions of the minor premise. For example, it does indeed seem to be the case that there is nothing else to which a soul could stand as attribute; it cannot coherently be supposed to be a state or condition of something which equally well might lack such an attribute. In the same way, the soul appears to be simple and unitary,4 and so on for the other fundamental claims of rational psychology.

So Kant turns to the other main way in which the syllogism might be rejected. Instead of attacking the truth of its premises, he concentrates on the validity of its movement from premises to conclusion. His charge is that the argument equivocates in its use of the middle term between the two premises, and that this equivocation takes the specific form of the fallacy of figura dictionis or figure of speech. We must now explore exactly what Kant's objection to the argument is, and whether it is justified.

The idea of the fallacy of figure of speech goes back to Aristotle's De Sophisticis Elenchis. It is a vice of argument that turns essentially on the language in which the argument is expressed (as opposed to those features of the argument which would remain however it is expressed); and in Aristotle's analysis the distinctive feature of this particular vice is that it confuses surface with deeper levels of grammar.5


For example, "cut" and "conjecture" and "dream" and "perceive" are all active verbs (in English). Many people would reject the idea that cutting, conjecturing, dreaming and perceiving are all activities. Let us not dispute the truth of that thought as regards the processes which these verbs refer to. Still there remains the question of whether at the linguistic and grammatical level it really is correct to group all these verbs under the same category which simple-minded attention to the surface would suggest. Perhaps some of these words are in fact passive verbs: "dream" and "perceive" are plausible candidates.

These hints are worked out in more detail as Aristotle explores the precise character of arguments which commit the specific fallacy of figure of speech. He defends the idea that these arguments involve a confusion, not so much over the meanings of individual words or more extended expressions, but rather over the grammatical role - or category - to which such linguistic items should be assigned in a correct understanding of their employment. An example, which will concern us shortly, is the distinction between internal and external objects of verbs. By "objects of verbs" here I mean, of course, to designate what are, at least primarily, linguistic items.

I shall argue that for those who wish to read Kant correctly, it is important to appreciate these background ideas in Aristotle. But before I come to that, a preliminary point is whether Kant actually understood Aristotle in the sense for which I have been arguing. Perhaps he got Aristotle wrong.

This charge might be fair in other contexts, but in this case it is not. Kant says that the middle term in the paralogism is "taken in totally different senses" in the two premises (B411 n.); and Aristotle notes the strong similarity between the argumentative faults of figure of speech and equivocation (S.E. 22, 178a24-8). Aristotle reports an alternative, erroneous diagnosis of the trouble with some paralogisms, as based on mistake of fact in one of the premises (178b10-23); the diagnosis fails because it is unable to handle all the various arguments which, Aristotle suggests, are intuitively to be grouped under the same heading. So Aristotle's diagnosis of this particular vice of argument - failure to distinguish different figures of speech - is clear and considered; and Kant appears to have understood him in the correct sense.6

---

6 Kant's brief comments on the sophisma figurae dictionis in the Hechsel Logic pp.110-11 and the Jäsche Logic section 90 (J.M. Young, Immanuel Kant: Lectures on Logic (Cambridge University Press, 1992) pp.410, 628) are sound and consistent with what he says here in the Critique of Pure Reason.
We have reason to think, then, that Kant's diagnosis of the specific fault in the
paralogism of rational psychology is accurate in the sense of being faithful to its
historical origins. Is it also sound as a piece of philosophy? Can those who would
otherwise be tempted by the paralogism, be rendered immune to its charms by
following Kant's remedy? I shall answer these questions in the affirmative; and a first
step is to identify the particular grammatical phenomenon which creates the difficulty
in the present dialectical context.

My diagnosis is that the phenomenon in question is the distinction between external
and internal objects of active verbs. Consider the activity of catching, the objects
which the activity may have, and the grammatical objects which may follow the verb
"catch". We may catch a ball, a brick, a box, and so on; or we may fail to catch such
things. Equally we may catch a catch - or drop it. A person who catches six balls and
drops three others, thereby catches six catches and drops three. On the other hand,
some things are such that they cannot be caught (by us, at least); examples are lightly
fried eggs, or grass cuttings.

These simple facts may suggest that the objects of catching - the things that can be
catched consist of various items, such as balls, bricks, boxes, catches - , while others
things in the world are excluded from being possible objects of catching. The world
of objects bifurcates into just two sets - the objects of catching and those things that
cannot be caught. But clearly this thought oversimplifies the matter. To be sure,
catches and balls are both things that can be caught; but surely they are so in
significantly different senses or ways? Our task now is to articulate and justify this
intuition.

One difference between catches and other objects of catching is this: any qualification
which attaches to the catching also attaches to the catch, but this is not the case with
the other objects. Thus if the activity of catching is agile or impressive, so also is the
catch; but while the catching may be agile, the ball that is caught is not agile. The
reason for this is clear. A catch is an internal object of catching. To catch is to catch a
catch; anyone who catches anything, thereby catches a catch. By contrast, we may
call such things as balls and boxes external objects of catching; for even though they
can be caught, the nature and manner of their existence is quite independent of the
catching.
For all that, catches can be dropped, just as balls and boxes can be; and this aspect of their logic can give some encouragement to the tendency to treat them as objects of the same kind as the external objects of catching. Therefore it becomes a task, even in this straightforward and philosophically insignificant case, to keep the distinction between internal and external objects of the activity firmly in view.

Catching provides a perspicuous example of a conceptual phenomenon which is widespread but less clear in other cases. Eating is an essentially object-related activity. To eat is necessarily to eat something. We eat potatoes, bread and food, but not nails or paint. "Food" in this list functions as an internal object of eating; it is what, as a matter of conceptual necessity, anyone who eats eats. But it is can easily seem to be external. For we surely want to say that bread and potatoes are food, while nails and paint are not; and bread and potatoes are external objects of eating.

Or consider the case of singing. We may sing "God save the Queen!" or "La donna è mobile". Such things are songs. Suppose that I sing more randomly in my bath. What results is still a song, although probably not a song to which we could give a name. In these cases we sing - and sing a song - , but there is no song that we sing. There may be an air of paradox about these truisms; but of course, they are all indeed true. How do we disarm the paradox? The best strategy is once again is to invoke our distinction between internal and external objects of activities - in these cases, eating and singing. "Song", like "food", is such an internal object. But notice also how readily these names can also be applied to the external objects ("God save the Queen!", potatoes). That is the justification for the claim that such terms are equivocal as between an internal and external use.\textsuperscript{7}

What is fairly easy to understand in the case of catches, although less so in the case of food and songs, is still more elusive in the case of thought. But the lesson should be the same. Thinking is an activity ("think" is an active verb) that takes two kinds of object. One kind is external. We can think of all sorts of things; but these things exist independently of our thought, and the manner of their existence does not - at least, directly - affect the nature of our thought.

\textsuperscript{7} For further discussion of these conceptual points, see G.E.M. Anscombe, 'The Intentionality of Sensation', in R.J. Butler (ed.), \textit{Analytical Philosophy: Second Series} (Oxford: Blackwell, 1968), pp.158-80. I have developed these ideas much further in J.D.G. Evans, 'Souls, attunements and variation in degree: \textit{Phaedo} 93-4', \textit{International Philosophical Quarterly} 34 (1994), 277-87, and 'Platonic Arguments', \textit{Aristotelian Society Supplementary Volume} 70 (1996), 177-93.
We also think thoughts. These are the internal objects of thinking. Any act of thinking must logically have such an internal object; and the nature of the thought correlates with the manner of the thinking. However we also find it natural to use the word "thought" to designate the external objects - the things about which we think. This habit is widespread among philosophers. Not only is it reflected in the classical tendency (found, for example, in Descartes, Locke and Berkeley) of calling all objects of experience "ideas". In using this term these philosophers are talking about what is external in the world, but they are also reflecting on the modes in which we experience such objects. The habit also survives in the more recent practice whereby philosophical investigations are characterised as conceptual. If a philosopher writes about *The Concept of Mind* or *The Mind - a conceptual analysis*, what is he doing but treating thoughts as things which it is specially appropriate for a philosopher to investigate?

This is the mental habit which, according to Kant's diagnosis, holds risks for the proper study of psychology. If the distinction between internal and external objects of thinking is not recognised and held clear, it is likely that insights which relate to our habits of thinking will (unthinkingly, we might say) be transferred to the world about which we think. Of course, the world which we seek to study must be amenable to our intelligence, if the study is to have any chance of success. It does not follow that the world should be construed on the model of an internal object of thought.

If we return to the syllogism in which Kant expresses the first paralogism in Second Edition B, we can gloss his comment, particularly in the footnote (B411), as a claim that in the major premiss the objects of thought are external objects, while in the minor premiss the thought is internal to the thinking of it. There is equivocation in the talk about thoughts as between the two premisses, and the key to exposing the equivocation lies in recognising the grammatical phenomenon of figure of speech.

In the two premisses we are presented with claims about the objects of our thought. The analysis which I have submitted - and particularly the comparison between objects of thinking, catching and eating - shows how there can be equivocation within the supposed middle term in these premisses, as Kant claims is the case in his B Edition analysis. But because our entire appraisal of this argument, as of any other, is inevitably intellectual and involves thinking, the analysis also makes clear how natural and ineluctable is the illusion which the argument generates. That is the other part of Kant's claim. I hope that with the help of Aristotle's account of faults in argument, I have done something to indicate the philosophical richness of the
diagnosis of mistaken reasoning which Kant seeks to unmask in this section of his work.

Prof. J.D.G. Evans, Queen's University of Belfast
January 1998