

Critique of Pure Reason the Dialectic

Immanuel Kant

Copyright ©2010–2015 All rights reserved. Jonathan Bennett

[Brackets] enclose editorial explanations. Small ·dots· enclose material that has been added, but can be read as though it were part of the original text. Occasional •bullets, and also indenting of passages that are not quotations, are meant as aids to grasping the structure of a sentence or a thought. Each four-point ellipsis indicates the omission of a brief passage that seems to present more difficulty than it is worth. Longer omissions will be reported between square brackets in normal-sized type. This version follows (B) the second edition of the *Critique*, though it also includes the (A) first-edition version of the Paralogisms of Pure Reason. Undecorated marginal numerals refer to page-numbers in B; ones with an 'A' in front refer to A, and are given only for passages that don't also occur in B. The likes of ..356 in the margin mean that B356 (or whatever) started during the immediately preceding passage that has been omitted or only described between square brackets. These marginal numerals can help you to connect this version with other translations, with the original German, and with references in the secondary literature. Cross-references to other parts of this work include the word 'page(s)', and refer to numbers at the top-right corner of each page.—The Transcendental logic divides into the Transcendental analytic, which started on page 45, and the Transcendental dialectic, which starts here.

First launched: January 2008

Contents

| | |
|--|------------|
| Introduction | 155 |
| 1. Transcendental illusion | 155 |
| 2. Pure reason as the seat of transcendental illusion | 157 |
| Book 1: The concepts of pure reason | 163 |
| 1. The ideas in general | 163 |
| 2. The transcendental ideas | 166 |
| 3. System of the transcendental ideas | 171 |
| Book 2: The dialectical inferences of pure reason | 174 |
| Chapter I, The paralogisms of pure reason (first edition) | 175 |
| First paralogism: Substantiality | 178 |
| Second paralogism: Simplicity | 179 |
| Third paralogism: Personhood | 182 |
| Fourth paralogism: ideality (in regard to outer relation) | 184 |
| The paralogisms of pure reason (second edition) | 197 |
| Chapter 2: The antinomy of pure reason | 206 |
| 1. System of cosmological ideas | 207 |
| 2. Antithetic of pure reason | 211 |
| First antinomy | 213 |
| Second antinomy | 216 |
| Third antinomy | 219 |
| Fourth antinomy | 222 |
| 3. What's at stake for reason in these conflicts | 225 |
| 4. The transcendental problems of pure reason, considered as downright <i>having</i> to be soluble | 230 |
| 5. A sceptical look at the cosmological questions raised by the four transcendental ideas | 233 |
| 6. Transcendental idealism as the key to sorting out the cosmological dialectic | 235 |
| 7. Critical solution of reason's cosmological conflict with itself | 238 |
| 8. Applying the regulative principle of pure reason to the cosmological ideas | 242 |
| 9. Putting the regulative principle of reason to work empirically, in connection with the cosmological ideas | 244 |

| | |
|--|------------|
| Chapter 3: The ideal of pure reason | 263 |
| 1. The ideal in general | 263 |
| 2. The transcendental ideal | 264 |
| 3. Speculative reason's arguments for the existence of a supreme being | 269 |
| 4. There can't be a successful ontological argument for the existence of God | 272 |
| 5. There can't be a successful cosmological argument for the existence of God | 276 |
| 6. There can't be a successful physico-theological argument for the existence of God | 283 |
| 7. Critique of all theology based on speculative principles of reason | 287 |
| Appendix to the transcendental dialectic | 292 |
| 1. The regulative use of the ideas of pure reason | 292 |
| 2. The final purpose of the natural dialectic of human reason | 302 |

Book 2:

The dialectical inferences of pure reason

..396 The object of a purely transcendental idea can be said to be something of which we have no concept, despite the idea's being something that reason is compelled by its own inherent nature to produce. That 'can be said', and it's true: Of an object that satisfies the demands of reason it is indeed impossible for us ever to form a concept of the understanding, i.e. a concept that could be exhibited and intuited in experience. Still, it might be better, and less likely
397 to mislead, if we said instead that although we can't have any knowledge of the object that corresponds to such an idea, we do nevertheless have a *problematic* concept of it.

The transcendental (subjective) reality of the pure concepts of reason depends on our having been led to such ideas by a necessary inference of reason. There will therefore be inferences of reason, having no empirical premises, through which we infer from something we know something else of which we have no concept though an inevitable illusion leads us to regard it as objectively real. Because of the conclusions they come to, these movements of the mind would be better called 'sophistical' [*vernünftelnde*] rather than 'inferences of reason' [*Vernunftschlüsse*; note the similarity of the two words—one might translate the former as 'fooling around with reason'], though their origin gives them some claim to the latter title, since they aren't fictitious and have arisen not by chance but from the very nature of reason. They are sophistries [*Sophistikationen*] not of men but of pure reason itself, and not even the wisest of men can free himself from them. If he works hard at it he may be able to guard himself against actual error; but he'll never be able to free himself from the illusion, which unceasingly mocks and teases him.

So there we have it: there are exactly three kinds of dialectical inferences of reason—just as there are three ideas in which their conclusions result.

(1) I call the first kind of inference of reason the transcendental *paralogism*. In it I conclude from the transcendental concept of the subject, which contains nothing manifold, the absolute unity of this subject itself, though in doing this I have no concept whatsoever of this subject. [Kant will explain this later. *Very* briefly and sketchily, the thought is this: The transcendental concept of *myself* is what's involved in every thought I have of the sort 'I now experience x', 'I now think about y'. It is 'transcendental' in the sense that it isn't the concept of thinking-being-with-such-and-such-characteristics; I can attribute to myself various properties, but when I do that, the transcendental concept is the concept of the *I* that does the attributing, not the *I* to which the properties are attributed. In that sense, then, my transcendental concept of myself doesn't reflect *any* of my complexity, i.e. 'contains nothing manifold'. And I commit a paralogism = invalid-inference of pure reason when I go from that •premise about the total uncomplexity of the transcendental *I* to a •conclusion about my not being in any way complex.]

(2) I shall call the state reason is in when conducting the second kind of sophistical inference the *antinomy* of pure reason. It involves the transcendental concept of the absolute totality of the series of conditions for any given appearance—e.g. the series of *all* the causes of a given event. [Note that whereas 'paralogism' is a label for a certain kind of *inference* that reason conducts, 'antinomy' here is the name of the *state that reason is in* when it conducts a certain kind of inference—a state of **conflictedness**, in which has two conflicting but equally bad ways of looking at something. Kant switches to calling individual pairs of conflicting propositions 'antinomies' = **conflicts** only when he gets to 'Comment on the first antinomy' on page 215.] When I think about my concept of the unconditioned synthetic unity of the series in one of the two ways, I find the concept to be

398

self-contradictory, so I take it in the other way, inferring that *that* is the truth of the matter, though in fact I have no ·acceptable· concept of that either. [A rough, quick example: When I try to think about all-the-causes-of-event-E on the assumption that every one of those causes also had a cause (so that the chain of them had no first member), I get into intellectual trouble; so I rush to the conclusion that some causes were not themselves *caused*, but were rather exercises of *freedom*; and that turns out to be intellectually problematic too.]

(3) Finally, in the third kind of sophistical inference, from •the totality of conditions for thinking of objects as such that I could be confronted with I infer •the absolute synthetic unity of all the conditions for things to be possible. That is, from things that I don't know (because I have merely a transcendental concept of them) I infer a *Being of all beings*, which I know even less through any transcendental concept, and of whose unconditioned necessity I can form no concept whatsoever. I'll label this dialectical inference of reason the *ideal* of pure reason.

Chapter I

The paralogisms of pure reason (1st edition)

399

A •logical paralogism is an inference of reason that is fallacious in form, whatever its content is. It counts as a •transcendental paralogism if there's a transcendental basis for the formal fallacy. A fallacy of this sort is based on the nature of human reason; the illusion it gives rise to can't be avoided, though it may be rendered harmless.

A concept that wasn't included in the general list of transcendental concepts must yet be counted as belonging to that list. I'm talking about the concept (or the judgment, if

you like) 'I think'. It's easy to see that this is the vehicle of all concepts: ·the only way for the concept C to come before me or enter into my scheme of things is for it to be the case that *I think* C·; and that includes transcendental concepts. So *I think* must itself count as transcendental. But it can't have any special label, because all it does is to bring forward, as 400 belonging to consciousness, *any* thought that one has; and that's why its omission from the initial list doesn't mean that the list was defective. Although it's not an empirical concept, it belongs on one side of a certain distinction that can be drawn empirically: the distinction between •myself considered as a thinking being, a soul, an object of inner sense, and •myself as a body, an object of outer sense. ·Obviously, the transcendental *I* belongs on the mental/soul/inner side of that divide·. I label as the 'rational doctrine of the soul' the kind of psychology whose subject-matter is expressed purely through the transcendental concept *I*. It is 'rational'—in the sense of having-to-do-only-with-reason·—because in it I don't try to learn anything about the soul from experience. In the •empirical doctrine of the soul I appeal to experience ·through inner sense·, and get specific detailed information about my soul; but in the rational doctrine of the soul I let all those details go, set aside *all* empirical input, and restrict myself to what I can learn about my soul considered just as something that is present in all thought.

So we have here something purporting to be a science built on the single proposition *I think*. How good are the grounds for thinking that there *is* such a science? That's the question we have to address now. You might want to object:

'The proposition *I think*, which expresses the perception of oneself, contains an inner experience. So the ·supposedly· rational doctrine of the soul built on this proposition is never pure—it is always to that extent based on an empirical principle.'

400

401

[Kant replies, at unhelpful length, that *this* ‘inner perception’ involves no details, doesn’t serve to mark off oneself from other things, and is simply a necessary accompaniment of *all* thought and experience; so that it shouldn’t be regarded as empirical knowledge. Then:] If to this all-purpose representation of self-consciousness we added the slightest object of perception (even if it’s only pleasure or unpleasure), that would immediately transform •rational psychology into •empirical psychology.

Thus, *I think* is rational psychology’s sole text, from which its whole teaching has to be developed. Obviously, if this thought is to be *about* something (myself), it can involve only transcendental predicates of that something, since the slightest empirical predicate would destroy this science’s rational purity, its independence from all experience.

What we have to do here is follow the guidance of the categories, with just one difference. In the transcendental logic I have always taken the categories in the order

•quantity, quality, relation, modality;

and I stand by that ordering considered as an aspect of the theory of categories, but in our present context I have to vary it by adopting the order

•(1) relation, (2) quality, (3) quantity, (4) modality.

That’s because our starting-point here is a given *thing*—*I* as a thinking being—so we must start with the category of *substance* (which is one of the categories of relation). Starting from there, we’ll be going through other classes of categories in reverse order [not strictly true!]. Thus, the topic [= ‘logical geography’] of the rational doctrine of the soul, from which everything else that it contains must be derived, is this:

(1)

The soul is *substance*

(2)

In quality it is *simple*

(3)

Through the different times when it exists, it is *one*, i.e. unity and not plurality

(4)

It relates to *possible* objects in space

All the concepts of pure psychology can be assembled out of these elements, with no other source being called upon. Here is how: 403

- this substance, merely as an object of inner sense, yields the concept of *immateriality*;
- as simple substance, it yields the concept of *incorruptibility* [here = ‘indestructibility’],
- its being a thinking substance that lasts through time yields the concept of *personhood*;
- all three of those combine to yield the concept of *spirituality*; and
- the substance’s relation to objects in space yields the concept of *causal interplay with bodies*, which in turn
- leads us to represent the thinking substance as the source of life in matter, i.e. as soul (*anima*), and as the basis of *animality*. Finally,
- animality, when combined with spirituality, yields the concept of *immortality*.

Out of all this there arise four paralogisms of a transcendental psychology that is wrongly regarded as a body of knowledge about the nature of our thinking being—knowledge that we acquire through pure reason. The only basis we can find for it is the simple, intrinsically empty representation *I*; 404

and this doesn't even qualify as a concept; it's merely a bare consciousness that accompanies all concepts. All that is represented through this *I* or *he* or *it* that thinks is a transcendental subject of thoughts = *x*. [In adding '= *x*' Kant wants to convey that this item is characterless, empty, a sort of place-holder, rather than something with a describable character of its own.] It is known only through the thoughts that are its predicates; apart from them we can't have any concept of it; any attempt that I make to characterize my transcendental *I* will use my representation of it in thoughts of the type: 'I conclude/think/see/believe/suspect/know that I am F'—where the first 'I' is the transcendental one—so that the attempt to describe it must revolve in a perpetual circle. There's no escape from this, because consciousness as such isn't a representation that picks out one object as distinct from others; rather, it is a form of representation in general. . . .

405 It must at first seem strange that something that is a pre-condition for *my* thinking—i.e. something that is merely a property of myself as a thinking subject—also holds for everything that thinks. That is the strangeness of the thesis that we can use a seemingly empirical proposition as the basis for a necessary and universal judgment, namely the judgment that anything thinks must be constituted in the way that the voice of self-consciousness declares that I am constituted. But although it is strange, it is also true, and here is why: It is *a priori* necessary that I attribute to a thing all the properties that are preconditions of my having any thought about them. Now, I can't have the slightest representation of a thinking being through any outer experience; I have to get it through inner self-consciousness; which means that I get my thoughts about thinking beings other than myself by transferring my consciousness to them. [Kant's next sentence is long and hard to follow. Its gist

is this: When I want to think about (for example) *you* as a thinking being, and so 'transfer my consciousness' to you, I am not mentally transferring to you any of my individual qualities. The transferable 'I think' that is involved here isn't what Descartes took it to be (when he argued from it to 'I exist'), namely a perception of an existent thing. And the use I am making of it is merely problematic; i.e. I'm using it only to ask some questions—I want to know what can be inferred from such a simple proposition, whether or not its 'I' stands for something that actually exists. Then:]

If our knowledge-from-pure-reason of thinking beings in general were based on

- more than the *cogito*, i.e. the inevitable, always-present, empty 'I think'
- our observations of how our thoughts come and go, and the natural laws of the thinking self that we derived from these observations,

that would give rise to an **empirical** psychology, a theory about the workings of inner sense. Perhaps it could explain the appearances of inner sense; but it couldn't ever reveal properties that don't in any way belong to possible experience (e.g. properties that something has because it is *simple*), or yield any knowledge of absolutely necessary truths about the nature of thinking beings as such. So it wouldn't be a **rational** psychology. 406 A348

Since the proposition 'I think' (taken problematically) contains the form of every single judgment of the understanding, and accompanies all categories as their vehicle, it is obvious that when we draw conclusions from that proposition we must be using our understanding only in a *transcendental* manner. [Why 'understanding' rather than 'reason'? Presumably because these would be inferences from a single premise, whereas Kant defines 'reason' in terms of inferences from two or more premises.] Since using the understanding in this way keeps

out any admixture of experience, and in the light of what I have already shown, we can't have much optimism about what we are going to achieve in this way. Well, let's keep a critical eye open as we follow this procedure through all the basic concepts of pure psychology.

From here until page 197 the material all comes from (A) the first edition of the *Critique*; the second-edition (B) version begins at page 197.

First paralogism: Substantiality

- If our representation of something x is the *absolute subject* of our judgments, so that x can't be used as determination of something else, x is *substance*.
- I, as a thinking being, am the *absolute subject* of all my possible judgments, and this representation of myself can't be used as predicate of anything else.
- Therefore I, as thinking being (soul), am *substance*.

Critique of the first paralogism of pure psychology

In the analytical part of the Transcendental Logic I showed that pure categories—one of which is the concept of substance—
 A349 •have no objective significance except when they are brought to bear on an intuition, and •are applied to the complex web of intuition as unifiers. In the absence of this web, they are merely forms of a judgment, without content. I can say of any thing that because it is a thing it 'is substance', in the sense that I am distinguishing it from mere predicates and states of things. •And from that I get something like the paralogism:

In all our thought, the *I* is the subject, in which thoughts inhere only as states; and this *I* can't be represented as the state of something else. So everyone *must* regard himself as substance, and regard

•his• thinking as merely properties that he has, states that he is in.

But what *use* am I to make of this concept of a substance? I certainly can't infer from it that I as a thinking being *persist* for myself and don't in any natural manner either *arise* or *perish*. But there's no other use I can make of the concept of the substantiality of myself as a thinking subject; if I can't use it to infer my permanence, I can't use it for anything. [Recall that in the Analytic Kant treated permanence, or never-going-out-of-existence, as the essence of the empirically usable category of substance.]

To see how far we are from being able to deduce permanence from the pure category of substance, consider how we have to proceed when we want to use the concept in an empirically useful way: to do this we must, at the outset, have an object that is *given in experience* as permanent. In contrast with that, in •the paralogism's inference from• the proposition *I think* we don't take any experience as our basis; rather, we infer a conclusion merely from the concept of the relation that all thought has to the *I* as the common subject that *has* the thought. . . . The *I* is indeed in all thoughts, but this representation doesn't contain the slightest trace of intuition, distinguishing the *I* from other objects of intuition. So we can indeed perceive that this representation keeps turning up in all thought, but not that it is an abiding intuition of something that continues in existence while its transitory thoughts come and go.

Conclusion: transcendental psychology's first inference of reason, in putting forward •the constant logical subject of thought as being knowledge of •the real subject in which the thought inheres, is palming off on us something that is a mere pretence of new insight. We don't and *can't* have any knowledge of any such subject. It's true that consciousness is needed if our representations are to be *thoughts*, which

A350

A351 implies that we'll encounter our perceptions only in the transcendental subject, .i.e. in the framework provided by 'I think'; but beyond this logical meaning of the *I*, we know nothing about the subject in itself that underlies this *I* as substratum, as it underlies all thoughts. We can allow the proposition '*The soul is substance*' to stand, as long as it's recognised that this concept of the soul as substance doesn't carry us an inch further, and so can't yield us any of the usual deductions of the pseudo-rational doctrine of the soul. . . .i.e. if we recognise that this concept signifies a substance only in idea, not in reality.

Second paralogism: Simplicity

- If something *x* is such that its action can never be regarded as the upshot of several things acting in concert, then *x* is *simple*.
- The soul or the thinking *I* is such a being.
- Therefore, the soul or the thinking *I* is simple.

Critique of the second paralogism of transcendental psychology

This is the Achilles [here = 'the strong man', 'the chief pusher-around'] of all the dialectical inferences in the pure doctrine of the soul. It's not a mere sophistical trick that a dogmatist [see note on page 15] has rigged up to give superficial plausibility to his claims; rather, it's an inference that seems to withstand even the keenest scrutiny and the most scrupulously exact investigation. Here it is ·with the details filled in·:

A352 Any *composite* substance *x* is an aggregate of several substances; anything it does (or any property that it has) is an aggregate of several actions (or properties), each belonging to one or other of the several substances. Now an effect *can* be the upshot of the working together of many acting substances (as

the motion of a body is the combined motions of all its parts). There's no difficulty in thinking about such compositeness when it concerns things that are *external* to the mind. But it's different when we come to thoughts—*internal* episodes belonging to a thinking being. For suppose that a thinking thing is composite; then every part of it would contribute a part of its thought, and its *whole* thought would have to come from all of its parts taken together. But this is ·covertly· self-contradictory. [From here to the end of this indented passage, this version expands on what Kant wrote, in ways that the ·small dots· convention can't easily indicate.] The movement of a composite body is the upshot of movements of all its parts, and they are conceptually unified as a single movement through someone's perceiving the body as a unity. Similarly, the thought of a composite thinker would have to be the upshot of thoughts of its parts; but how are those sub-thoughts to be conceptually united as a single thought? (Must they be so united? Yes. Consider a parallel case: the thought of a line of poetry. I think of hounds while you are thinking of spring, your brother is thinking of winter, and your sister is thinking of traces—but this state of affairs doesn't constitute anyone's thinking 'The hounds of spring are on winter's traces'. *That* thought has to be had by *someone*.) Any thought of a composite thinker has to be the thought of *someone*; it can't be the thought of that very composite thinker, because *every* thought of such a thinker is an upshot of many sub-thoughts, which means that we can never get down to a thought that is inherently and absolutely *one*, from which we might get going on conceptually unified composites. So a thought can't possibly be had by something that

is essentially composite; it must be had by a *single* substance, one that isn't an aggregate of substances, i.e. one that is absolutely simple.

A353 The core of this argument lies in the proposition that if many representations are to form a single thought they must be contained in the absolute unity of the thinking subject. But this can't be proved from *concepts*. The proposition

P: *A thought must be an effect of the absolute unity of the thinking being*

can't be treated as analytic. There's no conceptual contradiction in the supposition that ·P is false, i.e. that· a thought consisting of many representations might come from •the collective unity of different substances acting together (like the motion of a body coming from the motions of all its parts), rather than coming from •the absolute unity of the subject. So the necessity that (P) a composite thought must come from a simple substance can't be demonstrated through the principle of identity—i.e. can't be proved by showing that its contradictory is inconsistent·. Might P be known synthetically and completely *a priori* from mere concepts? You won't want to suggest *that* if you have understood my account of what makes it possible for synthetic propositions to be known *a priori*!

Nor will *experience* show us (P) that every thought must involve an absolutely single subject. Experience can't tell us about the necessity of *anything*, and anyway the concept of *absolute unity* is completely out of reach of experience. Well, then, what about this proposition P on which the whole psychological inference of reason depends—where can we get it from?

A354 It's obvious •that if anyone x wants to represent a thinking being y to himself he has to put himself in y's place, as it were *substituting* his own subject for y's, . . . and •that the reason why we insist that anyone who has a thought must be

absolutely unitary ·= partless = simple· is just that otherwise we couldn't have the 'I think' . . . For although the whole of the thought could be split up and distributed among many subjects, the subjective 'I' can't be split up and distributed, and it's this *I* that we presuppose in all thinking.

As in the first paralogism, so here too the formal proposition of self-awareness, *I think*, remains the only basis that rational psychology can rely on when it sets out to enlarge its knowledge. But this proposition is not itself an experience—it is the *form* of the self-awareness that belongs to and precedes every experience. Given that that's its status, its bearing on any possible item of knowledge is only that of a *merely subjective condition* of that knowledge; and we go wrong when we transform it into a condition—an *objective condition*—of the possibility of a knowledge of objects, i.e. into a *concept* of thinking-being-as-such. ·We don't and can't have any such concept·: the only way we can represent to ourselves *thinking-being-as-such* is by putting ourselves, along with ·the *I think* which is· the formula of our consciousness, in the place of every other thinking being.A355

So the famous psychological proof is based merely on the indivisible unity of a representation *I*, and all that *that* does is to govern the verb *think* in its relation to a person. It's obvious that in attaching *I* to our thoughts we refer to the thought-haver only transcendently; we aren't saying anything about any quality that it has; indeed we aren't acquainted with, and don't know anything about, any qualities that it may have. All the *I* refers to is a transcendental subject—a *something in general*. There is nothing determinate [here = 'detailed'] in it, which is one reason why it has to be simple. . . . But this simplicity of *the representation of a ·thinking· subject* is not knowledge of the simplicity of *the subject itself*

A356

So this much is certain: through the *I*, I always have the thought of myself as ‘simple’ in the sense of having an absolute but merely *logical* unity; but this doesn’t involve me in knowing anything about the actual simplicity of myself as a haver of thoughts. Just as the proposition ‘I am substance’ involves only the *pure* category of substance, which I can’t make any use of empirically, so here I can legitimately say: ‘I am a simple substance’, i.e. a substance the representation of which never involves a pulling together of several different elements, but. . . this proposition tells me *nothing* about myself as an object of experience, because the concept of substance is used here in a way that •doesn’t involve any underlying intuition and therefore •doesn’t have an object. . . . Now let us *test* the supposed usefulness of this proposition ‘I am a simple substance’.

The only reason why anyone has cared about the assertion of the simple nature of the soul is as a way of distinguishing this •thinking subject from all •matter, thus enabling the soul to escape from the dissolution to which matter is always liable. [That was one of Descartes’s two arguments for the immateriality of the soul: all matter is divisible, no soul is divisible, therefore etc.] That’s why the proposition in question is usually expressed as ‘The soul is not corporeal’. Well, now, suppose

A357

we

- take this top proposition of rational psychology, in the meaning that is appropriate to a judgment of pure reason derived solely from pure categories, and

- allow it full objective validity, so that it becomes the •fact–stating proposition that *everything that thinks is a simple substance*;

•even with this grotesque self-indulgence we *still* can’t get the top proposition to throw any light on the question of whether or how the soul differs from matter. That is what

I am about to show; and that will be tantamount to sidelining this supposed psychological insight, relegating it to the domain of mere ideas without the grip on actuality that would give it an objective use.

In the Transcendental Aesthetic I conclusively proved that bodies are mere •appearances of our outer sense, not •things in themselves. So we’re entitled to say that our thinking subject isn’t corporeal: it is represented by us as an object of *inner* sense, so it can’t be an object of *outer* sense, i.e. an appearance in space, •as bodies are. This amounts to saying that we can’t find thinking beings—*as thinking beings*—among outer appearances; i.e. that their thoughts, consciousness, desires and so on can’t be outwardly intuited because they all belong to inner sense. This argument seems to be so natural and so popular that even people with only average intellectual abilities have relied on it as a reason for the age-old view that souls are quite different from bodies.

A358

•Here, as so often, a genuine truth has to be watched so that it doesn’t purport to say more than it does. It is true that extension, impenetrability, cohesion, and motion—in short, everything that outer senses can give us—are different from and don’t contain thoughts, feelings, desires, or decisions, because these are never objects of outer intuition. But •let’s not let that run away with us. There is

(1) the *Something* that underlies outer appearances, affecting our sense in ways that give it representations of space, matter, shape etc.;

and there is

(2) the *Something* that is the subject of our thoughts.

And the above argument for saying that the soul is not a body doesn’t conflict with the view that (1) is identical with (2)—i.e. that what *underlies* outer appearances is the same noumenon (or, better, the same transcendental object) as what *underlies* or *has* our thoughts. It’s true that the way our

outer sense is affected by the Something doesn't give us any intuition of representations, of will, or the like, but only of space and space-related properties; but the Something itself isn't extended or impenetrable or composite, because those predicates have to do only with sensible intuitions that we have through being affected by certain objects that we know nothing about in any other way. In saying that the Something is 'not extended' etc., we aren't expressing any knowledge about what kind of an object it is, but only acknowledging that considered in itself—apart from any relation to the outer senses—it's not something to which those predicates of outer appearances can be applied. But there's nothing about it that is inconsistent with the predicates of inner sense, representations and thought. Thus, even if we allow that the human soul is simple in nature, that doesn't distinguish it from the substratum of matter—if matter is considered (as it should be) as mere appearance.

A359

If matter were a thing in itself (·and if the soul were also a thing in itself·), then matter as composite would have to be different from the soul, which is simple. But when we take matter to be mere outer appearance of Something that can't be known through any predicate that we can assign to it, we have to admit that this Something might be simple, even though it affects our senses in such a way as to give us the intuition of something extended and therefore composite. Nor is there any obstacle to supposing that •the substance that appears to our outer sense as extended *has thoughts*, and that •it can represent these thoughts by means of its own inner sense. If that were how things stood, a single thing would be (taken one way) corporeal while also being (taken another way) a thinking thing whose thoughts we can't intuit though we can intuit their signs in the domain appearance. And then we'd have to give up the thesis that only **souls** think, taking souls to be substances of a particular kind; we

would have to replace that by the commonplace statement that **men** think, i.e. that the very same thing that as outer appearance is extended is also (in itself) internally a simple subject of thoughts.

A360

[Kant now re-states the view he has been expressing, in several ways that aren't sufficiently different to throw much new light. Then:]

..A361

Thus the collapse of rational psychology's main support brings the whole thing crashing down. It's as true here as it is elsewhere that we can't hope to extend our knowledge through mere concepts—let alone through the consciousness that is the merely subjective form of all our concepts—in the absence of any relation to possible experience. And in our present case there is an extra reason for that general result. The basic concept of a *simple nature* can't be fitted to anything we encounter in experience, so that there's no way it can function as an objective concept.

Third paralogism: Personhood

- Anything that is conscious of •the numerical identity of itself at different times—i.e. of being the very same individual thing at different times·—is to that extent a *person*.
- The soul is conscious of the numerical identity of itself at different times, .
- Therefore the soul is a person.

Critique of the third paralogism of transcendental psychology

If I want to know through experience the numerical identity of an external object, I shall focus on the permanent element in the appearance—the element that is the subject x such that everything else in the appearance is a state of x—and I shall note *its* identity throughout the time in which the states come and go. Now, I am an object of inner sense, and all

A362

time is merely the form of inner sense. Consequently, I relate each of my successive states to the numerically identical self in all •time. . . . This being so, the proposition that the soul is a person has to be regarded not as something I *infer* but rather as an identical [here = 'trivially analytic'] proposition about consciousness of oneself in time—which is what makes it valid *a priori*! For all it says, really, is that in the whole time in which I am conscious of myself I am conscious of this time as belonging to the unity of myself. I can say

- this whole time is in me, as individual unity, or that
- I am to be found as numerically identical in all this time,

and it makes not the slightest difference which I say.

In my own consciousness, therefore, identity of person is unfailingly met with. But if I view myself from the standpoint of someone else (as an object of *his* outer intuition), it is this external observer who first represents

- me as in time;

because really all I get from my self-awareness is a representation of

- time in me.

A363 Although this observer admits the *I* that accompanies. . . .all representations at all times in *my* consciousness, he won't infer from this that I am something objectively permanent. For just as the time in which he places me is the time not of •my sensibility but of •his, so the identity that is necessarily bound up with •my consciousness is not therefore bound up with •his identity. . . .

The identity of the consciousness of myself at different times is therefore only a formal condition of my thoughts and their coherence, and in no way proves the numerical identity of myself as a thinking subject. Despite the logical identity of the *I*, there may have been a change that rules out a continuing identity. It could be that one thinking

subject is replaced by another, that by a third, and so on, while the same-sounding *I* is used all through, because each outgoing thinking subject hands over its state to its immediate successor.⁴

A364

Consider the dictum of certain ancient schools, that everything in the world is *in a flux* and nothing is *permanent*, nothing lasts. This can't be reconciled with the thesis that there are substances, because they are by definition permanent things; but it isn't refuted by the unity of self-consciousness, because our own consciousness doesn't tell us whether as souls we are permanent or not. Since we count as belonging to our identical self everything we are conscious of, we have to judge that we are one and the same throughout the whole time of which we are conscious. [Kant wrote '*only* what we are conscious of', but that was presumably a slip, because 'everything that we are conscious of' is what's needed for his line of thought.] But we still can't claim that this judgment would be valid from the standpoint of an outside observer. Here is why: What we encounter in the soul is not any permanent appearance, but only the representation *I* that accompanies and connects all the inner appearances; so we can't prove that this *I*, a mere

⁴ An elastic ball that collides with another similar one in a straight line passes on to the other its whole motion, and therefore its whole state (that is, if we take account only of the positions in space). If, then, in analogy with such bodies, we postulate substances of which one passes on to another its representations along with the consciousness of them, we can conceive a whole series of substances of which the first transmits its state together with its consciousness to the second, the second to the third, and so on down the chain, with each substance handing over all its own states and those of its predecessor. The last substance would then be conscious of all the states of all the substances that had been switched into and out of the series, and would be conscious of them *as its own*, because they would have been transferred to it along with the consciousness of them. Yet it wouldn't have been one and the same person in all these states.

A365 thought, isn't in the same state of flux as the other thoughts that are strung together by means of it.

[Kant now offers a horribly difficult paragraph, about the order in which we *do* argue for

- 'the soul is permanent',
- 'the soul is a substance', and
- 'the soul is a person',

and the order in which we *could* argue for them if things were different in certain ways. The details are cloudy, and the paragraph seems not to be needed for a grasp of the main lines of Kant's thoughts about the paralogisms. He then continues:]

A366 Just as we have kept the concept of *substance* and of the *simple*, it's also all right for us to keep the concept of *person*; but we must give it its merely transcendental status as something that concerns the unity of the subject—the thinking subject about which we don't know anything else, but whose states are thoroughly inter-linked by self-awareness. Taken in this way, the concept is good enough for practical use; but we mustn't parade the proposition 'The soul is a person' as •adding something to our self-knowledge through pure reason, and as •exhibiting to us, from the mere concept of the identical self, an unbroken continuance of the subject. Why? Because if we look to this concept for leverage on any question that aims at synthetic knowledge, it will just keep spinning on its axis, giving no help. We don't know what *matter* may be as a thing in itself, but because it is represented •to us• as external •to us•, we can observe its permanence as appearance. But if I want to observe the mere *I* in the change of all representations, I have no other correlate to use in my comparisons except again myself, with the universal conditions of my consciousness. [Kant means: In empirically identifying matter as substantial, I compare some of my intuitions with others, comparing •the subset of them that do pertain

to matter with •the subset that don't. But when I come to the question of whether *I* am a substance, all I can appeal to is the omnipresent *I* that accompanies absolutely all my mental states, so that I can't show my substantial status by comparing some of my intuitions with others.] So if someone else raises the question of whether I am a continuously existing person, the only answers I can give are tautological ones in which I . . . take for granted that which the questioner wants to know. •That is, I answer his question about what I am in the only way I *can* tackle such a question, namely by reporting on my own inner states and events; but I have to report these as *mine*, with the *I* running all through my account; that makes my subjective *I* deputise for the questioner's objective concept of substance, and so has the effect of presupposing an answer to his question without throwing any light on it.

Fourth paralogism: ideality (in regard to outer relation)

- If the only basis for believing in x's existence is an inference to x as a cause of given perceptions, then it is open to question whether x does exist.
- The existence of outer appearances is never immediately perceived; our only basis for believing in their existence is an inference to them as causes of given perceptions.
- Therefore it is open to question whether any objects of the outer senses really exist.

367

My label for this uncertainty—this open-to-question-ness—is 'the ideality of outer appearances'; and the doctrine of this ideality, •expressed in the conclusion of the fourth paralogism•, is called **idealism**. The opposing doctrine, which says that we *can* have certainty about •the real existence of• objects of outer sense, is called **dualism**.

Critique of the fourth paralogism of transcendental psychology

Let's start with the premises. This paragraph and the next will give a sympathetic statement of the lines of thought that lie behind the premises of the fourth paralogism. We're justified in contending that we can't immediately perceive anything that isn't in ourselves, and that for me the only object of a mere perception—i.e. the only thing that I immediately perceive—is my own existence. So the existence of an actual object outside me. . . . is never given directly or immediately in perception. Perceiving something is having one's inner sense in a certain state; and the only way to bring an outer object *x* into the story is by thinking of *x* as the outer cause of the inner state, and thus *inferring* the existence of *x*. . . . Obviously what is external to me isn't in me; so I can't encounter it in my self-awareness or, therefore, in any perception, because the right way to see perceptions is as mere states of our self-awareness.

..A368

So I'm not in a position to *perceive* external things, but can only infer their existence from •my inner perception, taking •this as an effect of some external immediate cause. Now, the inference from a given effect to a definite cause is always uncertain, because the effect may be due to more than one cause. Thus, when we are thinking about the causes of perceptions, it always remains doubtful—open to question—whether the cause is internal or external; i.e. whether all the so-called •outer perceptions aren't a mere play of our •inner sense, or whether they are related to actual external objects that cause them. Anyway, the existence of outer objects is only inferred, and is vulnerable to all the troubles that an inference can run into, whereas the object of inner sense (I myself with all my representations) is immediately perceived, and there can't be any doubt that *it* exists.

So it's wrong to think of an 'idealist' as someone who

denies that there are any external objects of the senses. An idealist, properly so-called, is someone who won't admit that the existence of such objects is known through immediate perception, from which he infers that there couldn't be any experience that made us completely certain of the reality of external objects of the senses.

A369

Before exhibiting our paralogism in all its deceptive illusoriness, I should first remark that we must distinguish •transcendental idealism from •empirical idealism. [Kant will stay with this and related distinctions for about four pages. He won't again refer explicitly to •the fourth paralogism, but his discussion of types of idealism constitutes a critique of •it.] By *transcendental idealism* I mean this doctrine:

Appearances are *all* to be regarded as mere representations, not as things in themselves, so that time and space are only •sensible forms of our intuition, not •states given as existing by themselves and •not conditions of objects viewed as things in themselves.

To this idealism there is opposed a *transcendental realism* that regards time and space as given in themselves, independently of our sensibility. The transcendental realist thus interprets outer appearances (taking for granted that they are real) as things-in-themselves, which exist independently of us and of our sensibility, and are therefore outside us—taking the phrase 'outside us' in its most radical sense. It's this transcendental realist who afterwards plays the part of empirical idealist: after wrongly supposing that if objects of the senses are external they must have an existence by themselves, independently of the senses, he finds that from this point of view all our sensuous representations are inadequate to establish the reality of those objects.

A370

The transcendental idealist, on the other hand, can be an empirical realist—or a *dualist*, as he is called. That is, he can grant the existence of matter without •going outside his mere self-consciousness or •assuming anything more than the certainty of his representations. . . . For he regards the facts about what matter there *is*, and even about what there *could be*, as facts merely about appearance; and when appearance is separated from our sensibility it is nothing. For him, therefore, matter is only a species of representations (intuition); and these representations are called ‘external’ not because they relate to objects that are *in themselves external* (they don’t), but because they relate perceptions to the space in which all things are external to one another, although the space itself is in us.

[It may be useful to have a brief restatement of the main theses of the preceding two paragraphs: Kant has distinguished

(1) two transcendental theses about matter, i.e. two views about the meanings or metaphysical status of propositions about matter:

(a) idealism: such statements are really complex statements about our states of mind;

(b) realism: such statements are entirely independent of facts about our minds—they don’t imply such statements and aren’t implied by them.

And he has distinguished

(2) two empirical theses about the status, for us, of the proposition that there is matter in the world:

(a) idealism: we can’t have certainty that the proposition is true;

(b) realism: we can be perfectly certain that the proposition is true.

One natural pairing, Kant is saying, is

(1b) transcendental realism and (2a) empirical idealism.

Because the proposition that there is matter has a status that puts it out of our reach, we can’t be sure that there is any matter. The other natural pairing is

(1a) transcendental idealism and (2b) empirical realism.

The proposition that there is matter is a special kind of proposition about our own mental states; that puts it within our reach, enabling us to be quite sure that it is true.]

Right from the outset I have declared my acceptance of transcendental idealism; and that clears the way for me to accept the existence of matter on the unaided testimony of my mere self-consciousness, taking it to be proved in the same way that I prove ·to myself· the existence of myself as a thinking being. ·Here’s how it goes·: I am conscious of my representations; so these representations exist, and so do I, the subject that *has* them. External objects (bodies) are mere appearances, so they are only one kind of representation that I have, and representations of that kind aren’t *of* anything beyond the representations themselves. Thus •the existence of external things is as secure as •my own existence, because I know *both* from the immediate testimony of my self-consciousness. The only difference is that the •representation of myself as the thinking subject belongs to **inner** sense only, whereas the representations that signify extended things belong also to **outer** sense. [Note ‘belong *also*’: outer sense is just a *part* of inner sense.] I don’t need *inference* to establish the reality of outer objects, any more than I need inference to establish. . . .the reality of my thoughts. In both cases, the objects are nothing but representations, the immediate perception (consciousness) of which is automatically a sufficient proof of their reality.

. . . .Transcendental realism, on the other hand, inevitably runs into trouble, and finds that it has to allow empirical idealism. Here is why: It regards the objects of outer sense as distinct from the senses themselves, taking mere appearances as self-subsistent beings that exist outside us. On that view, however clearly we are conscious of our representations of these things, it’s still far from certain that if the representations exist then the corresponding objects also exist. In my

A371

A372 system, on the other hand, these external material things are. . . nothing but mere appearances, i.e. representations of whose reality we are immediately conscious.

So far as I know, all psychologists [here = 'philosophers of mind'] who adopt empirical idealism are transcendental realists; and they have certainly been consistent in parading empirical idealism as 'setting' an important problem from which human reason can't easily extricate itself. For if we regard outer appearances as representations produced in us by their objects, and if these objects are things existing in themselves outside us, it's impossible to see how we could come to know the existence of the objects other than by inferring causes from effects; and the conclusions of such inferences are *always* doubtful, even when the cause in question is *in us*. Perhaps our outer intuitions are indeed caused by something that is (in the transcendental sense) 'outside' us; but, if so, this cause isn't the kind of object we have in mind when we talk about 'matter' and 'bodies'. . . . What we are talking about is not this •transcendental object that we don't know about either through inner or through
A373 outer intuition. Rather, we are speaking, of the •empirical object, which is called an *external* object if it is represented *in space*, and an *inner* object if it is represented only *in its time-relations*. And space and time are to be found only *in us*.

The phrase 'outside us' is thus unavoidably ambiguous: sometimes it refers to

(1) something which *as a thing in itself* exists apart from us,

and at other times it refers to

(2) something belonging solely to outer *appearance*.

The psychological question about the reality of our outer intuition involves (2), and we need an unambiguous way of saying this. So I shall distinguish (2) *empirically external*

objects from (1) ones that may be said to be transcendently external, by labelling (2) as 'things that are to be found in space'.

Space and time are indeed *a priori* representations that reside in us, as forms of our sensible intuition, *before* any real object has acted on our senses through sensation and enabled us to represent the object in terms of its spatial and/or temporal relations. But the material or real element, the Something that is to be intuited *in space*, necessarily presupposes perception; in the absence of perception, no power of imagination can invent and produce that Something. So it is sensation that indicates a reality in space or in
A374 time. . . . (When a sensation is taken to be *of* something, though without giving any details about it, we call it 'perception'.) Once a sensation has been given, its internal variety enables us to picture in imagination many objects that have no empirical place in real space or time. There's no room for doubt about this: it's perception that provides the raw materials we need if we are to have thoughts about objects of sensible intuition. This holds equally for 'inner perceptions of' pleasure and pain and for the sensations of the outer senses, such as colours, heat, etc., but just now my topic is the 'outer' part of the story. *This perception represents something real in space*, and here is my three-part reason for saying so. (1) Just as space is the representation of a mere *possibility* of coexistence, perception is the representation of a *reality*. (2) This reality is represented •to outer sense, i.e. •in space. (3) Space is itself nothing but mere representation. And so we get the 'double' result:

•Only what is represented in space can count as real in space.⁵

⁵ We must take careful note of the paradoxical but correct proposition that there's nothing in space but what is represented in it. 'Why is it true?' Because space itself is nothing but a representation,

A375 •*Everything* that is represented through perception as given in space is real in it. . . .

So all outer perception provides immediate proof of something real in space—or, rather, it is itself what is real. This puts empirical realism beyond question—there *does* correspond to our outer intuitions something real in space. Of course space and all its appearances are representations, which means that they are only *in me*, but ·that doesn't abolish the distinction between inner and outer·: what is real, i.e. the material of all objects of outer intuition, is given in this space as actual and independent of all imaginative invention. And it's impossible for anything that is (in the transcendental sense) *outside* us to be given in *this* space, which is nothing apart from our sensibility. Thus, even the strictest idealist can't require a proof that our perception has a corresponding object that is 'outside' us in the strict ·transcendental· sense. If there were any such object, it couldn't be represented and intuited as outside us; because this would involve *space*, a mere representation, containing no reality that isn't in perception. . . .

A376

Knowledge of objects can be generated from perceptions, either by mere play of imagination or by means of experience. [See note on page 155 regarding 'knowledge']. And in the course of this there can indeed arise illusory representations, ones with no corresponding objects, the deception being attributable sometimes to the imagination's playing tricks (in dreams) and sometimes to the judgment's going astray (in so-called 'sense-deception'). To avoid such deceptive illusion, we have to steer by the rule:

and nothing can be in it except what is contained in that representation. . . . It must indeed seem strange to say that a thing can exist only in the representation of it, but the sense of strangeness evaporates in our present context, where the things in question are not things in themselves but only appearances, i.e. representations.

Anything connected with a perception according to empirical laws is actual.

But such deception, as well as this shield against it, has as much to say to idealism as to dualism. I'm talking about ·transcendental idealism, i.e.· our present concern with the *form* of experience. I needn't re-introduce empirical idealism because I have already refuted that and its mistaken challenge to the objective reality of our outer perceptions [and Kant briefly repeats his arguments to that effect. Then:]

A.377

·A new distinction between kinds of idealism needs to be drawn now·. On the one hand we have

(1) the dogmatic idealist, who *denies* the existence of matter.

He must base this denial on supposed contradictions in the thought of there being such a thing as matter at all. I haven't needed to discuss this so far, but I shall do so: the difficulty will be removed in the next chapter [which starts on page 206] on dialectical inferences, where I'll display reason as being at odds with itself regarding the concepts it makes for itself. . . . On the other hand we have

(2) the sceptical idealist, who *doubts* the existence of matter, thinking that it can't be proved to exist.

·While it's appropriate to brush the dogmatic idealist aside as being wholly wrong·, the sceptical idealist, ·though also in error·, is a benefactor of human reason! All he does is to challenge our basis for asserting that matter exists; we thought we could base it on immediate perception, but he criticises that as inadequate. This challenge compels us to be constantly on the watch—even in the smallest advances of ordinary experience—to ensure that we don't treat as a well-earned possession something that we may have obtained only illegitimately. Now we can see clearly the value to us of these ·sceptical·-idealist objections. [Kant goes on to say *what* this 'value' is: it turns out to consist in our being forced

A378

by sceptical idealism to keep in mind and stay true to the tenets of transcendental idealism. If we treat outer objects as things in themselves, our situation is as bad as the sceptical idealist says, and indeed even worse. So we *have to* adopt the only alternative, namely the thesis that outer things are mere representations. He continues:]

A379

The question then arises: ‘In the philosophy of mind, is dualism the only tenable position?’, and our answer has to be: ‘Yes indeed, but only when “dualism” is understood in the empirical sense.’ That amounts to taking dualism as saying that in the interconnected web of experience

- matter, as substance in the [domain of] appearance really is given to outer sense,

just as

- the thinking *I*, also as substance in the [domain of] appearance, is given to inner sense.

Further, inner and outer appearances must be connected with each other according to the rules that this category of substance brings to our perceptions—inner as well as outer—enabling them to constitute one experience. But if we try (and people often do) to extend the concept of dualism and take it in the transcendental sense, we’ll arrive at something for which there isn’t the slightest basis. Why? Because we’ll have misapplied our concepts, taking •the difference between two ways of representing objects (which, as regards what they are in themselves, still remain unknown to us) as •a difference in these things themselves. (And this fault is present not only in transcendental dualism but also in the two opponents to it—*pneumatism* on one side and *materialism* on the other.) [Pneumatism is the thesis that the soul is immaterial.] The *I* represented through inner sense in time is a specifically quite distinct •appearance from objects in space outside me, but these two shouldn’t be construed as different •things. The *transcendental object* that underlies

outer appearances is not matter; the *transcendental object* that underlies inner intuition is not a thinking being. Rather, each of these is a ground (to us unknown) of the appearances that supply us with empirical concepts of matter and of mind. . . .

A380

Consideration of pure psychology as a whole, in view of these paralogisms

If we compare **(1)** the *doctrine of souls* as the physiology [= ‘empirical’] study of inner sense, with **(2)** the *doctrine of body* as a physiology of the object of the outer senses, we find that while there’s a lot of empirical knowledge to be gained in both of them, they are notably unlike in what can be learned *non-empirically* through them. In **(2)** there is much *a priori* synthetic knowledge to be had from the mere concept of an extended impenetrable being, whereas in **(1)** there’s no comparable knowledge from the concept of a thinking being. Here is why. Although both kinds of being are appearances, the **(2)** appearance to outer sense contains something fixed or lasting, which supplies the underlying *thing* which all the transitory states are states *of*. That enables it to present a synthetic concept, namely the concept of space and of an appearance in space. In contrast with that, time—which is the sole form of our inner intuition—doesn’t contain anything lasting, so it provides knowledge only of the change of states, not of any object that they are states *of*. In the ‘soul’ (as we call it) everything flows and nothing stays still except the *I*, which is simple solely because the representation of it has no content and thus no qualitative complexity. . . .

A381

A382

What would it take for us to have, through pure reason, knowledge of the nature of a thinking being as such? We would need the *I* to be an **intuition** which, being presupposed in all thought as such (prior to all experience), could

yield *a priori* synthetic propositions. But the *I* that we have is no more an •intuition than it is a •concept of an object! Rather, it is the mere **form** of consciousness, which can accompany the two kinds of representation (•inner and outer•) but can't elevate them to the level of items of knowledge unless **something else** is given in intuition—something that provides material for a representation of an object. Thus the whole of rational psychology, as a science surpassing all the powers of human reason, collapses. The most we can do is to •study our soul under the guidance of experience, and •confine ourselves to questions that stay within the limits of what might possibly be answered by inner experience.

But although rational psychology is useless as a way of extending our knowledge, and when so used is entirely made up of paralogisms, it undeniably has considerable negative value as a critical treatment of our dialectical inferences, those that arise from common and natural reason.

A383

What leads us to resort to a doctrine of the soul based on nothing but pure principles of reason? No doubt we are primarily aiming to secure our thinking self against the danger of materialism. This sense of danger takes the form of the fear that

(1) if all matter went out of existence, all thought—and even the very existence of thinking beings—would be destroyed.

But *that* fear is dealt with by the pure concept of our thinking self that I have been presenting. What we get from it, far from the fear (1), is a clear proof that

(2) if the thinking subject went out of existence, the whole corporeal world would necessarily also vanish, because that world is nothing but an appearance in the sensibility of our •thinking• subject, a way in which its representations occur.

Admittedly this doesn't tell me anything more about the properties of this thinking self, e.g. giving me insight into whether it is permanent or not. It doesn't even throw light on whether the thinking self exists independently of the transcendental substratum of outer appearances (supposing there is one); because that substratum is as unknown to me as is the thinking self. [Kant will now speak of •other-than-speculative reasons for hoping for something. He is speaking of •practical reasons—ones connected with morality rather than metaphysical theory—for *hoping* that one's soul can survive through into an after-life.] Still, I may come to have a non-speculative reason to hope for an independent and continuing existence of my thinking nature, throughout all possible changes of my state. In that case it will be a great help if, while freely admitting my own ignorance, I can repel the dogmatic assaults of a speculative opponent, showing him that just as I can't support clinging to my hope by appeal to any knowledge of the nature of the self, he can't bring such knowledge to support his denial that the hope can be realized.

A384

The real goal of rational psychology lies in three other dialectical questions that •are also based on this transcendental illusion in our psychological concepts, and •can't be settled except through the inquiry I have been conducting. They concern

- (1) the possibility of interaction between a soul and an organic body,
i.e. the question of what it is to have an animal nature and of how the soul fits into human life;
- (2) the beginning of this interaction,
i.e. the question of the soul in and before birth; and
- (3) the end of this interaction,
i.e. the question of the soul in and after death (the question of immortality). [In the foregoing, 'interaction' translates Kant's *Gemeinschaft*, often translated as 'community'.]

A385 Some people think that these questions involve difficulties that they can use as dogmatic [see note on page 15] objections to certain beliefs or hopes concerning the soul. They want to be admired for having a deeper insight into the nature of things than the general run of us can claim to have! Well, I maintain that what they have is merely a *delusion* in which they *hypostatise* something that exists merely in thought—that is, they treat it as a real object existing. . . .outside the thinking subject. [Kant is going to use ‘hypostatise’ quite a lot, often in its basic sense of ‘treat as a thing or substance’, but occasionally in the different though related sense of ‘treat as being real independently of the mind’.] In other words, they regard extension, which is nothing but appearance, as a property of outer things that exist quite apart from our sensibility, and claim that motion is due to these things and really occurs in and by itself, apart from our senses. This is a delusion because matter, whose interaction with the soul causes so much fuss, is a mere *form*, a particular way of representing an unknown object through the kind of intuition that is called ‘outer sense’. Perhaps there really is outside us something corresponding to this appearance that we call ‘matter’; but even if there is, in its role as appearance it isn’t outside us; it is only a thought *in* us, although this thought represents it as existing outside us because it comes to us through outer sense. So ‘matter’ doesn’t refer to a kind of substance that is utterly unlike the object of inner sense (the soul), but only to the distinctive nature of certain appearances of objects. The objects are in themselves unknown to us, but we call our representations of them ‘outer’ as compared with those that we count as belonging to ‘inner’ sense, although these outer representations belong only to the thinking subject, as do all thoughts. There is indeed something deceptive about them: representing objects in space, they seem to detach themselves from the soul, so to speak, and to hover outside it.

And yet the very space in which they are intuited is nothing but a representation, and what it’s a representation of can’t be found outside the soul. So we drop the question about the interaction between the soul and other known substances of a different kind outside us; and we’re left with a question about how representations of inner sense are connected with states of our outer sensibility—the question of how these can be so inter-linked according to settled laws that they hang together in a single experience. A386

As long as we hold inner and outer appearances together in our minds as mere representations in experience, we won’t see anything absurd or strange in the thought of their interaction—the interaction between these the two kinds of senses. We run in trouble over interaction only when

we hypostatise outer appearances, come to regard them not as representations but as things existing by themselves outside us, with the same qualities that they have in us, and think of them as acting on our thinking subject in the way they (as appearances) act on one another.

We get into difficulties then, because the efficient causes outside us—material things colliding with one another—have a character that can’t be squared with their effects in us. That’s because the cause relates only to outer sense, the effect to inner sense—and although these senses are combined in one thinking subject they are extremely unlike each other: the only outer effects are changes of place, and the only forces are drives that result in changes of place; whereas within us the effects are thoughts, which don’t have any spatial features—no locations, motions, shapes etc. That’s why, when we try to trace outer causes through to their effects in inner sense, we get lost. But we should bear in mind that bodies are not objects-in-themselves that are present to us, but a mere appearance of some unknown A387

object; •that motion is not an effect of this unknown cause, but only the appearance of its effect on our senses; •that bodies and motions are not something outside us, but mere representations in us; and •that, therefore, the motion of matter doesn't *produce* representations in us, because the motion is a representation only. . . . The bottom line is that this whole self-inflicted problem boils down to this:

How and through what cause are the •representations of our sensibility so interconnected that the •ones we call 'outer' intuitions can be represented according to empirical laws as objects outside us?

This question doesn't in any way involve the supposed difficulty of explaining how our representations could be effects of utterly different efficient causes outside us. That difficulty arose from our taking the appearances of an unknown cause to be the cause itself outside us—a mistake that is bound to lead to confusion.

A388 When a judgment involves a misapprehension that has taken deep root through long custom, one can't, straight off, correct it as clearly as one can correct mistakes that aren't conceptually confused by inevitable illusion. So my freeing of reason from sophisticated theories can hardly have yet the clarity that is needed for its complete success. But I think the following comments will be a move towards complete clarity.

Objections are of three kinds: **(1)** dogmatic, **(2)** critical, and **(3)** sceptical. A **(1)** dogmatic objection is directed against a *proposition*, a **(2)** critical objection is directed against the *proof* of a proposition. To make a **(1)** dogmatic objection to proposition P about some object x, one needs an insight into the nature of x that will entitle one to maintain the opposite of what P says about x. The objection counts as 'dogmatic' because it claims to know more about how x is constituted than does the proposition it is opposing. [Re 'dogmatic', see

note on page 15.] A **(2)** critical objection doesn't say anything about whether the proposition P is any good, so it doesn't presuppose. . . . fuller knowledge concerning the nature of the object x; all it does is to attack the proof that has been offered for P. If a critical objection succeeds, it shows only that P is unsupported, not that it is wrong. A **(3)** sceptical objection sets up P and not-P as equally matched opponents, treating each—turn about—as asserted dogmatically and objected to dogmatically by its opponent. This conflict, seemingly dogmatic on both sides, implies that all judgment on the topic in question is completely null and void. So dogmatic and sceptical objections both lay claim to as much insight into their object as they need for their assertion or denial. But a critical objection confines itself to •pointing out that an assertion presupposes something that's empty and merely imaginary, thereby •overthrowing the asserted theory by pulling its supposed foundation out from under it, without trying to establish any rival view about the nature of the object.

A389

When we bring the ordinary concepts of our reason to bear on the question of the interaction between our thinking subject and the things outside us, we are dogmatic, regarding outer things as real objects existing independently of us (in line with a certain *transcendental dualism*, which doesn't assign these outer appearances to the subject as representations, but completely separates them from the thinking subject, placing them outside us while still giving them the properties they are given in our sensible intuition. This switch is the basis of all the theories about the interaction between soul and body; they all accept without question the objective reality of outer appearances. . . . The three standard theories about this are in fact the only *possible* theories: that of **(1)** physical influence, that of **(2)** predetermined harmony, and **(3)** that of supernatural intervention. [At Kant's time

..A390

'physical' and its cognates in other European languages didn't imply any restriction to items that *we* would include in 'physics'. It comes from a time-honoured trilogy—logical (what *must be*), physical (what *is*), and ethical (what *ought to be*). So item **(1)** is simply the view that bodies and minds genuinely causally affect one another.]

The accounts **(2)** and **(3)** of the relations between the soul and matter are based on an objection to **(1)** the view of common sense. The objection is this: what appears as matter can't by its immediate influence be the cause of •representations, because •these are too different in kind from matter. [Kant goes on to say that this objection would be meaningless if the objectors regarded matter, in the way Kant does, as a mere representation produced by unknown outer objects. Then:] If their objection were to square with my principles, it would have to say that the true (transcendental) object of our outer senses can't be the cause of the representations (appearances) that we label as 'matter'. But no-one is entitled to say *anything* about the transcendental cause of our representations of the outer senses; so the objection in this form of it is entirely groundless. So we'll have to take these objectors against **(1)** the doctrine of physical influence to be •sticking to the ordinary outlook of transcendental dualism, and •supposing that matter is a thing-in-itself rather than the mere appearance of some unknown thing. So the aim of their objection will be to show that outer objects of this kind, which don't exhibit among themselves any causality except the causing of movements, can't possibly be efficient causes of representations; so that a third entity must intervene to establish if not •interaction then at least •correspondence and harmony between the two. But this objection of theirs *starts* with a basic untruth, namely •their view about• physical influence, which is built into their dualism. Thus, what their objection really refutes is not •the thesis of• natural influence between soul and

body but rather •their own dualistic presupposition. . . . A392

So the commonly accepted doctrine of physical influence can't be effectively opposed by a **dogmatic** objection. •There are two bases from which someone might try to launch a dogmatic objection•. •He could flagrantly hypostatise representations, setting them outside himself as real things. I have shown how untenable this is, and that there's no alternative to transcendental idealism. Well, then, •he could •accept that alternative•, agree that matter and its motion are mere appearances and therefore mere representations, and object to **(1)** on the grounds that the unknown object of our •outer• sensibility couldn't possibly be the cause of representations in us. But he can't justify this, because no-one is in a position to work out what an *unknown* object can or can't do!

However, a sound **critical** objection can be made against the ordinary version of **(1)** the doctrine of physical influence •between soul and body•. The supposed interaction between two kinds of substances, the thinking and the extended, is based on a crude dualism; it turns extended substances (which are really nothing but mere representations of the thinking subject) into things that exist by themselves. . . .

Let's take the notorious question of the interaction between the thinking and the extended, filter out from it any fictitious ingredients, and see what we are left with. It is simply this: A393

How is it *possible* for a thinking subject—*any* thinking subject—to have outer intuition, i.e. an intuition of space and of the filling of space by shape and motion? Which is a question that none of us can possibly answer. This gap in our knowledge *can't* be filled. The most we can do is to mark its place by referring to 'the transcendental object that causes representations of this •outer• type, though we can never know anything about it or even have a concept of it'.

We don't *need* such a concept when dealing with problems arising in the •domain of experience, for then we treat these appearances as objects in themselves, without worrying about the ultimate basis for their possibility as appearances. But we *would* need the concept of a transcendental object if we were to pass the limits •of this •domain•.

These reminders of what the inter-relation is between thinking beings and extended beings suffice to settle all the arguments about the state of the thinking nature before this inter-relation begins (i.e. prior to life) or after it ends (in death). Take the opinion that the thinking subject was able to think before becoming connected a body. This becomes the thesis that

•Before the start of the kind of sensibility through which something appears to us in space, the transcendental objects that do in fact appear •to us• as bodies could have been intuited in an entirely different manner.

And the opinion that after the end of the soul's connection with the corporeal world it could still go on thinking becomes the thesis that

•A stoppage of the species of sensibility through which transcendental objects. . . appear to us as a material world wouldn't automatically create a stoppage of all intuition of transcendental objects. It's quite possible for those same unknown objects to go on being known by the thinking subject, though not of course now intuited as bodies.

Now, no-one can give this the faintest support from any speculative principles. Even the *possibility* of what is asserted can't be established, but only assumed. But it's equally impossible to bring any valid dogmatic objection against it. None of us knows anything about the absolute, inner cause of outer corporeal appearances; so none of us can justify

claiming to know what the outer appearances in our present state (that of life) really rest on; or to know that when this state ends (in death), that will bring the end of all outer intuition or even of the thinking subject itself.

A395

Thus all strife about the nature of the thinking being and its connection with the corporeal world is sheerly a result of plugging a gap in our knowledge with paralogisms of reason, treating our thoughts as things and hypostatizing them. This gives rise to a 'science' that is entirely imaginary, on both sides of each debate, because both sides •suppose they have knowledge of objects of which no human being has any concept, or •treat their own representations as objects, and so whirl around in a perpetual circle of ambiguities and contradictions. This dogmatic delusion keeps many people in bondage to theories and systems, by tempting them with thoughts of an imagined happiness. [It's not clear whether Kant is referring to the happiness of believing in such a theory or system, or to the happiness of life after death.] And the only way of getting free from this bondage is through a sober critique that is both strict and fair, and that confines all our speculative claims to the domain of possible experience. It doesn't do this by stale scoffing at ever-repeated failures, or pious sighs over the limits of our reason, but by effectively *fixing* these limits in accordance with established principles, inscribing its 'go no further' on the Pillars of Hercules. [These marked the two sides of the straits of Gibraltar, regarded by the ancients as the furthest limit of sea voyaging.] Nature herself has erected these, so that the voyage of our reason shan't be extended further than the continuous coastline of experience lets us go—a coast we can't leave without venturing on a shoreless ocean which, after alluring us with deceptive promises, eventually compels us to abandon as hopeless all this vexatious and tedious endeavour.

A396

A394

* * *

I still owe you a clear general account of the transcendental and yet natural illusion in the paralogisms of pure reason, and also a justification of my classifying them in a way that runs parallel to the table of the categories. If I had tried to provide these at the start of this chapter, I would have risked writing obscurely, or clumsily getting ahead of myself. I'll now try to provide what I owe.

Here's an account of *illusion* in general: it consists in treating the *subjective* condition of thinking as being knowledge of the *object*. That covers the •illusions of the senses that sometimes occur in special cases—e.g. being led by your blurred vision of something to think that it has a furry surface—but that isn't relevant to our present topic of dialectical •illusion of pure reason. *That* has to involve subjective conditions of *all* thinking, not just of some special cases. What *are* these universal conditions? Well, in the Introduction to the Transcendental Dialectic I showed that pure reason concerns itself solely with the totality of the synthesis of the conditions for a given conditioned [page 167], and there will therefore be only three cases of the dialectical use of pure reason:

A397

- (1) The synthesis of the conditions of thought as such.
- (2) The synthesis of the conditions of empirical thinking.
- (3) The synthesis of the conditions of pure thinking.

In each of these, pure reason is concerned only with the absolute totality of this synthesis, i.e. with the condition that is itself unconditioned. This •trio of kinds of synthesis gives rise to the •trio of transcendental illusions (corresponding to the •three chapters of the Dialectic) and to the •trio of 'sciences' of pure reason—(1) transcendental psychology, (2) transcendental cosmology, and (3) transcendental theology. Here we are concerned only with the first.

[Back at page 171 Kant arrived at those three pretended sciences through three logical forms of inference, based on three logical types of proposition: (1) subject-predicate ('categorical'), (2) if-then ('hypothetical'), (3) either-or ('disjunctive'). He doesn't work hard at showing that or why the two trios coincide, though we'll see right away what he thinks ties the two versions of (1) together.]

[This next paragraph expands Kant's very compressed one in ways that the •small dots• convention can't easily handle. The expanded version is offered with fair confidence that it correctly presents Kant's thought.] Thought *as such* doesn't involve any object (whether of the senses or of the pure understanding). That is, it isn't essential to thought that it be *about* something. So the synthesis of conditions of thought—i.e. the stitching together of items that must be involved in *any* thinking—can't involve any relations with objects; it's just a stitching together of •the thought with •the thinker, the person whose thought it is. And objects enter the picture only through the thinker's *mistaken* view that he has a synthetic representation of himself as an object.

It follows from this that (1) the dialectical inference to the conclusion that there is an unconditioned item that is a condition of all thought doesn't commit a •material error (for it abstracts from all content, all objects); rather, it is defective in •form alone, which is why it must be called a 'paralogism'.

A398

Furthermore, since the *only* condition that accompanies all thought is the *I* in the universal proposition 'I think. . .', *that's* what reason has to identify as the unconditioned condition. It is indeed a condition of all thought, but only a •formal condition, securing the •logical unity of every thought, with no *object* having any role in it. . . .

If you ask me 'What is the constitution of a thing that thinks?', you'll want a synthetic reply. (An analytic answer might explain what is meant by 'thought', but that's

all; it can't tell you anything about what makes thought possible.) Now, I have no *a priori* knowledge on which to base a synthetic reply, because that would require an appeal to intuition, but your highly general question shuts intuition out because it concerns thought *as such*, including thought that doesn't involve intuition. . . . Yet it *seems* as if I could reply to you on the basis of the proposition that expresses self-consciousness—*I think*. For this *I* is the primary subject—so it is substance, it is simple, and so on. [Kant then says something impenetrably obscure about what should lead me to have suspicions about this plausible answer to your question, though it doesn't diagnose *what* is wrong with it. Then:]

But I can learn what has gone wrong if I dig deeper into the origin of the attributes that I ascribe to myself as a thinking being as such. [Kant then provides the diagnosis that we have already met: concepts such as 'is a substance', 'is simple' and so on aren't fit for expressing any item of knowledge except in a context where there can be intuitions supporting the distinction between what is and what isn't a substance, what is and what isn't simple. He continues:] If I call a thing in the ·domain of· appearance *simple*, I mean that the intuition of it, though it is a part of the appearance, can't in its turn be divided into smaller parts. ·And in our present context, where our concern is with thinking beings *as such*, thinking beings *in general*, no role can be given to intuitions.· If I know something as *simple in concept* but not as *simple in the ·domain of·* appearance, then this isn't an item of knowledge about the object but only about the concept that I form of a 'something' that can't be intuited. My only ground for saying in this case that I think something as completely simple is that I really don't have anything to say about it except merely that 'it is something'.

Now the bare self-awareness, *I*, is in concept substance, in

concept simple, etc.; and in this sense all those psychological doctrines are unquestionably true. But this doesn't give us the knowledge of the soul that we are looking for. Why not? Because none of these predicates can be applied to anything given in intuition, so they can't have any consequences that hold for objects of experience, so they are entirely empty. The concept of substance doesn't teach me that the soul endures by itself, or that it is a part of outer intuitions that cannot itself be divided into parts, and therefore can't arise or perish by any natural alterations. These are properties that would make the soul known to me in the context of experience and might tell me something about its origin and future state; they're the kind of thing that brings the *schematised* concept of substance into play [see page 92]. But if I say, in terms of the mere ·unschematised· category, 'The soul is a simple substance', it is obvious that since the bare concept of substance (supplied by the understanding) contains nothing beyond the requirement that a thing be represented as being subject in itself, and not in turn predicate of anything else, nothing follows from this as regards the permanence of the *I*, and the attribute 'simple' certainly doesn't aid in adding this permanence. Thus, from this source, we learn nothing whatsoever as to what may happen to the soul in the changes of the natural world. If we could be assured that the soul is a *simple part of matter*—a *physical atom*—we could use this knowledge, with the further assistance of what experience teaches about such things, to deduce the permanence, and (with its simple nature thrown into the mix) the indestructibility of the soul. But of all this, the concept of the *I*, in the psychological principle 'I think', tells us nothing.

[The remaining couple of pages of the first-edition treatment of the Paralogisms are brutally difficult, and are probably not worth the trouble, given that Kant is going to re-do the

whole thing in the second edition.]

[The treatment of the Paralogisms as re-stated in the second edition of the *Critique* starts here, picking up from page 178. It is about half the size of the A-material that it replaces; it will run to page 204.]

The paralogisms of pure reason (second edition)

[Back on page 178 Kant wrote of the need to 'keep a critical eye open as we follow this procedure through all the 'four' basic concepts of pure psychology'. He now says that he'll present the material in a 'continuous' way, meaning that he won't (as he did in the first edition) deal at length with each of the four paralogisms separately, though he does start by giving them a paragraph each. That's after a preliminary paragraph which he says may help us to sharpen our wits in the remainder of the chapter. This extremely difficult paragraph, which isn't needed for what follows, is omitted from this version.]

(1) Here's a proposition that is absolutely necessary—indeed, it's analytic:

•In all judgments I am the *determining* subject of the relation that constitutes the judgment.

It has to be granted that the *I*, the *I* that thinks, can always be regarded as *subject*, and as something that doesn't occur in the thought in a merely *predicate*-role. But this doesn't mean that *I*, as *object*, am for myself a *self-subsistent* being or *substance*. The latter statement goes very far beyond the former, and demands for its proof data that aren't to be met with in thought. . . . [In the background of this is Kant's thesis, expounded in the metaphysical deduction of the categories, that our concept of *substance* is the concept of something that figures in our thought as a subject and never as a predicate.]

(2) Here is another proposition that is analytic because it merely states something that is already contained in the very concept of thought:

•The *I* of self-awareness, and therefore the *I* in every act of thought, is *one*, and can't be resolved into a plurality of subjects, and consequently signifies a logically simple subject.

But this doesn't mean that the thinking *I* is a simple *sub-* 408
stance. That proposition would be synthetic. The concept of substance always relates to intuitions that in my case •have to be sensible, and therefore •lie entirely outside the domain of the understanding and its thought. [Recall that •for Kant 'sensible intuition' means 'intuition in respect of which the person is **pas-**
sive'; he holds that all human intuition is like that (hence 'in my case'); and that •he ties 'understanding' tightly to '**active**'.] But it is of this thought—and not of anything intuitive or sensible—that we are speaking when we say that the *I* in thought is simple. •A comment on these first two paragraphs: It's very hard work to find out which of the things that intuition presents us with are substances, and which of them are simple; so it would be astonishing if results about substantiality and simplicity were just handed to me, as though by revelation, in the poorest of all representations—the mere bare empty *I think*.

(3) A third proposition that is implied by the concepts that it uses, and is therefore analytic:

•Through all the variety of which I am conscious •through time• I am identical with myself.

But this identity of the subject, of which I can be conscious in all my representations, doesn't involve any intuition of the subject that would present it as an object; so it can't signify the identity of the *person*, i.e. the . . . identity of one's own substance, as a thinking being, in all change of its states. To establish *that*, we would need various synthetic judgments, 409
based on intuition, that come to us, not a mere analysis of the proposition *I think*.

(4) A fourth analytic proposition:

- I distinguish my own existence as that of a thinking being from other things outside me—among them my body.

This is analytic because *other* things are ones that I think of as *distinct* from myself. But this proposition doesn't tell me whether this consciousness of myself would be possible if there were no things outside me giving me representations, or therefore whether I could exist merely as thinking being (i.e. without existing in human form, ·equipped with a body·).

So we see that the analysis of my consciousness of myself in thought in general, ·thought as such·, contributes *nothing* to my knowledge of myself as object. ·Those who have fallen for the paralogisms have· mistaken •the logical exposition of thought in general for a •metaphysical account of the nature of the object.

Suppose that we *could* prove *a priori* that all thinking beings are in themselves simple substances, so that personhood is inseparable from them and that they are conscious of their existence as separate and distinct from all matter. That would be a great stumbling-block—indeed *the* great stumbling-block—in the way of my whole critique. Why? Because in conducting such a proof we would have stepped outside the world of sense and entered the domain of noumena; and no-one could then deny our right to advance yet further into this domain, indeed to settle there and—with luck—stake a claim to permanent possession. The proposition

Every thinking being is, just because it is a thinking being, a simple substance

is a synthetic *a priori* proposition. •It's synthetic because it goes beyond the concept from which it starts, adding to the concept of a thinking being its way of existing. •And it's *a priori* because the predicate (namely simplicity) that it adds

to the concept of the subject can't be given in any experience. It would then follow ·from the supposition at the start of this paragraph· that *a priori* synthetic propositions are possible and admissible, not only (as I have said) in relation to objects of possible experience and indeed as principles of the possibility of this experience, but that they are applicable to things in general and to things in themselves—a result that would make an end of my whole critique and force me to go along with the status quo. But if we look closer we'll find that there is no such serious danger.

The whole procedure of rational psychology is dictated by a paralogism that is exhibited in this inference of reason:

- Anything that can't be *thought* otherwise than as subject doesn't *exist* otherwise than as subject, and is therefore substance. 411
- A thinking being, considered merely as such, can't be thought otherwise than as subject.
- Therefore a thinking being exists also only as subject, i.e. as substance.

Why is this a paralogism? What's wrong with it? The answer is that there is an ambiguity in the middle term, the one that occurs in both premises, namely: 'can't be thought otherwise than as subject'. The major premise uses this term unrestrictedly: it speaks of things that 'can't be thought other than as subject' *however* they are being mentally engaged with, including their being presented in intuition. But the minor premise concerns something that 'can't be thought otherwise than as subject' when thinking about itself as a subject of thought and the unity of consciousness and *not* when it is confronting itself through inner sense as something given in intuition. So the conclusion is reached invalidly, through a fallacy of ambiguity. 412

[In a long and very difficult paragraph Kant reminds us of claims he has defended in two parts of the *Analytic*; these,

he says, confirm that he is right in ‘resolving this famous argument into a paralogism’. Then:]

Refutation of Mendelssohn’s proof of the permanence of the soul

..413 The usual argument for the soul’s permanence ·or immortality· takes it that the soul is a simple being and argues that it therefore can’t go out of existence by *dissolution*; ·i.e. it doesn’t have *parts*, so it can’t be destroyed by being taken *apart*·. The acute Mendelssohn soon noticed that this argument doesn’t prove that the soul can’t go out of existence, because it might be supposed to go out of existence ·not by falling apart but· by *vanishing*. In his *Phaedo* he tried to plug that gap by arguing that the soul can’t undergo such a process of vanishing, which would be a true annihilation. [Kant’s point in that last clause is just that when a thing is merely dismantled that isn’t a true annihilation because its parts stay in existence, whereas what’s at issue now is a complete annihilation with nothing left behind.] His tactic was to argue that a thing that is *simple* can’t cease to exist. His argument goes like this (not a quotation):

414 The soul has no parts, so there is no plurality involved in it. So it can’t be diminished or lessened in any way, which means that it can’t *gradually* lose something of its existence, *gradually* going out of existence. If it could go out of existence, therefore, this would have to happen absolutely *suddenly*, with no time between a moment when it exists and a moment when it doesn’t—which is impossible.

But what Mendelssohn overlooked was this: even if the soul is simple, meaning that it doesn’t have parts that are external to one another, and thus doesn’t have **extensive** magnitude, we can’t deny that it like every other existing thing has **intensive** magnitude, i.e. a *degree* of reality in respect of its

faculties and indeed of all that constitutes its existence; and this degree of reality can diminish through all the infinitely many smaller degrees. In this way the substance whose permanence is at issue here might be changed into nothing not by dissolution but by gradual loss of its powers—by fading away to nothing, so to speak, ·rather than shrinking down to nothing·. For consciousness itself always has a degree, which can always be diminished; and the same must also be true of the power of being conscious of the self and likewise of all the other faculties. [In a footnote here, Kant seeks to allay an obscure worry about how the notion of •degrees of consciousness relates to the notion of •clarity of thought. The cure of the trouble is also obscure, and the footnote is omitted here.] Thus the permanence of the soul, regarded merely as an object of inner sense, remains unproved and indeed unprovable. Its permanence during life is, of course, evident in itself, because a human being is not only something that *thinks* but also something that is an object of the outer senses. But this won’t satisfy the rational psychologist, who sets out to prove from mere concepts the soul’s absolute permanence beyond this life.

· A LONG FOOTNOTE ·

[At this point Kant has a very long footnote, which it is more convenient to lift up into the main text. Here it is:]
Some philosophers think they have done enough to show that some new scenario of theirs is *possible* by defying everyone to prove that it contains a contradiction. (For example, those who think they can see the possibility of thought even after this life has stopped—although all they know about thought comes from empirical intuitions of our human life!) But those who argue in that way can be brought to a puzzled halt by the presentation of other ‘possibilities’ that are no less bold. [Kant wrote ‘no more bold’, but we’ll see below that this must have been a slip.] An example would be **(1)** the ‘possibility’ that

a simple substance might divide into several substances, and conversely **(2)** the ‘possibility’ that several substances might fuse together to form one simple substance. ·I shall comment on these in turn·.

(1) Something is divisible only if it is composite, but it might be a composite not of substances but only of degrees of the various powers of a single substance. We can certainly make sense of this thought:

All the powers and abilities of the soul, even that of consciousness, are reduced to one half in such a way that the substance still remains.

Well, there’s no *contradiction* in the thought:

All the powers and abilities of the soul are reduced to one half, with the half that the soul loses staying in existence outside it.

In that scenario, everything in the soul that is real and therefore has a degree—in other words, its entire existence, nothing omitted—has been halved; so another separate substance would come into existence outside it, ·to possess the half of everything lost by the original soul·. The many items that have been divided all existed before the division. They didn’t exist as many substances, but as many items that contributed to the reality of the substance, i.e. to how much existence it had. So its being *one* substance was therefore only a mode of existence, which in virtue of this division has been transformed into a plurality of subsistence. [That last sentence conservatively translates what Kant wrote. He may mean something like this:

What we are fundamentally talking about here are the many thoughts and powers etc. that have been split into two groups. The fact that they were all possessed by what we call ‘one substance’ is just a fact about how *they* existed—one might say that ‘they existed one-sub-ly’; whereas now, after the split-up, they exist two-sub-ly.

That seems to be the most plausible reading of the passage, though it doesn’t explain Kant’s using first ‘existence’ and then ‘subsistence’.]

(2) Several simple substances might be fused into one, with nothing being lost except the plurality of *things*, because the one substance would contain the degree of reality of all the former substances together. And perhaps the simple substances that appear to us as *matter* might produce the souls of children, i.e. producing them through a division of the parent souls considered as *intensive* magnitudes, with the parent souls making good their loss by fusing with new material of the same kind. (This division of the parent souls wouldn’t be mechanical or chemical, but rather would involve a causal influence unknown to us, of which mechanical and chemical influences were only appearances.)

I’m *not* saying that these fantasies are useful or valid; and the principles of my Analytic have warned us against using the categories (including that of substance) in any way except empirically. But if

the **rationalist** is bold enough to construct a self-subsistent being out of the mere faculty of thought, with no help from any permanent intuition through which an object might be given, doing this merely on the ground that the unity of self-awareness in thought can’t be explained in terms of something composite; instead of admitting, as he ought to do, that he can’t explain the possibility of a thinking nature *·at all·*,

why shouldn’t the **materialist**, though he can’t appeal to experience in support of his ‘possibilities’ either, be justified in being equally bold and using his principle to establish the opposite conclusion, while still preserving the formal unity *·of self-awareness·* upon which his opponent has relied?

·END OF LONG FOOTNOTE·

416 [Now we have three impossibly difficult paragraphs, which won't be paraphrased here. Here are three points that may help you to wrestle with this material yourself.

[(1) Kant is here using 'analytic' and 'synthetic' as labels not for •propositions but for •procedures, giving them senses that are *totally* different from what they have in the rest of this work. These new-to-us meanings were in fact quite standard before Kant's time (and perhaps afterwards, though his main use of 'analytic' and 'synthetic' as labels for propositions seems to have grabbed the limelight and pushed the other out of sight). The two procedures differ in *direction*: a •synthetic study of a body of doctrine starts with its final output, the theories that constitute it, and works backwards to the reasons for those theories, the reasons for those reasons, and so on back to the ultimate basis for the whole thing; whereas an •analytic study starts with what is epistemically basic—what one knows at the start—and proceeds from there to consequences, then consequences of those consequences, and so on to the final theories. [In a footnote on page 173 we saw Kant using 'analytic' and 'synthetic' in different though related senses, namely as labels for methods of exposition.]

[(2) In this present passage, Kant describes a •synthetic' approach to rational psychology as expressed in the displayed quartet of propositions on page 176 above: the approach starts from 'All thinking beings are substances' and works its way 'backwards' through the other three members of the quartet, and eventually back to the all-purpose *I think* that assures of our existence. He doesn't *make* much of this, except to remark that it commits rational psychology to 'problematic idealism', i.e. to the thesis that 'Is there a material world?' is a permanently open question.

[(3) He also describes an •analytic' approach that starts with *I think* and works its way forward to the simplicity and substantiality of the soul. He uses this to generate

another displayed quartet (which is elegant rather than helpful), and to make the claim that we now see rational psychology straining to give information about the nature of thinking beings, and being thwarted because it can't adopt the *materialist* view that thinking things are material things, or the *spiritualist* view that there are only immaterial thinking things. Then:]

..421

So there's no informative *doctrine* of rational psychology, but only a *discipline*. This sets impassable limits to speculative reason, limits that keep us from •throwing ourselves into the arms of a soulless materialism or, on the other side, •losing ourselves in a spiritualism that must be quite unfounded so long as we remain in this present life. Without providing any positive doctrine, rational psychology reminds us that we should regard

reason's refusal to give a satisfying answers to our inquisitive questions about things that are beyond the limits of this present life

as being

reason's hint that we should divert our self-knowledge from fruitless and extravagant speculation to fruitful practical use.

Though in such practical use reason is never directed to anything but objects of experience, it gets its principles from a higher source, and sets us to behave as though our destiny reached infinitely far beyond experience, and therefore far beyond this present life.

All this makes it clear that rational psychology owes its origin simply to a misunderstanding, in which •the unity of consciousness that underlies the categories is mistaken for •an intuition of the ·thinking· subject as an object, and is then brought under the category of substance. This unity is really only unity in *thought*, and on its own it doesn't present any object; so the category of substance can't be applied

422

to it, because that category always presupposes a given *intuition*. Therefore, this ·thinking· subject can't be known. The subject of the categories cannot, just by thinking them, acquire a concept of itself as an object of the categories. [The unity of consciousness 'underlies' the categories, and the thinking self is the 'subject' of them, in the sense that the only way to use any category is in a thought that one has, an *I think*, and in this context the 'subject' is the being that *has* the thought.]. . . .⁶

There's a desire to get knowledge that will extend beyond the limits of possible experience while also furthering the highest interests of humanity; speculative philosophy has claimed to satisfy it; and we can now see that the claim is based on deception. Still, the severity of my critique has

⁶ As I have already said, the *I think* is an empirical proposition, and contains within itself the proposition 'I exist'. But I can't say 'Everything that thinks exists', because that would imply that the property of thought makes everything that has it a necessary being. So I can't regard my existence as **inferred from** the proposition 'I think', as Descartes maintained; because if it were inferred from that premise there would also have to be the premise 'Everything that thinks exists'. Rather, the proposition that I exist is **identical with** *I think*. The *I think* expresses an indeterminate empirical intuition, i.e. perception. . . . But the *I think* precedes the experience that would be needed to apply the category to the object of perception. ·That doesn't make trouble for the *I exist* that is identical with *I think*, because· the existence involved in this isn't yet a category. A genuine category can't be applied to an indeterminately given object, but only to one of which we have a concept and are asking whether it exists outside the concept. In our present context an indeterminate perception simply points to something real that is given—given to thought as such and therefore not given •as appearance or •as a thing in itself (*noumenon*) but simply •as something that actually exists and is tagged as such in the proposition, *I think*. When I called the proposition *I think* an 'empirical proposition', I didn't mean that the *I* in this proposition is an empirical representation. On the contrary, it is purely intellectual, because it belongs to thought in general, ·i.e. to *all* thought, including thought that has no empirical content·. . . .

rendered reason good service by proving that it's impossible to arrive dogmatically at any results—concerning any object of experience—that lie beyond the limits of experience. ·Why is that a 'service'?· Because it secures reason against any possible assertion of an opposing view. The defence against the opposing view can be seen as having four stages: **(1)** try to prove that the proposition one is defending is necessarily true; **(2)** find that this can't be done; **(3)** explain why it can't be done, namely because of the unavoidable limits of our reason; then **(4)** make the opponent back down because he too has been trying to infringe the limits.

But this doesn't take anything away from the right, indeed the necessity, of believing in a future life in accordance with the principles of the •practical use of reason, which is closely bound up with its •speculative use. The merely speculative proof has never had the slightest influence on ordinary common-sense. It stands on the tip of a hair, so precariously that even the schools can stop it from falling only by keeping it spinning around like a top; so even they can't see it as providing an enduring foundation on which something might be built. My critique doesn't at all lessen the value of the proofs that work for the world at large; indeed it increases their clarity and natural force by stripping away those dogmatic pretensions. Here is why:

Those arguments place reason in its own special domain, namely, the order of ends ·or purposes·, which is also an order of nature. Now, because reason is in itself not only a theoretical but also a practical faculty, it isn't tied down to natural conditions and can legitimately expand the order of ends—and with it our own existence—beyond the limits of experience and of life.

·And here is *how* it does that·. When dealing with the *analogy with the nature* of living beings in this world, reason has

to accept the principle that no organ, no faculty, no impulse, no *anything* is either superfluous or disproportioned to its use, so that everything is exactly conformed to the end or purpose that is destined for it. Now, if we were to judge things on the basis of this kind of analogy, we would have to regard man. . . . as the only creature who is excluded from this order of ends. [Although he hasn't said so, Kant must here be thinking of the natural 'order of ends' as the way every feature of any organism is fitted for its 'end' of its own survival and flourishing.] Think about man's natural endowments, not merely his talents and the impulses to enjoy them, but above all the moral law within him. These go far beyond any benefit or advantage he could get from them in this present life—so far beyond that they teach him to prize the mere consciousness of a righteous attitude as being supreme over all other values, quite apart from any advantage it might bring him and apart even from the shadowy reward of posthumous fame. They make him feel an inner call to fit himself, by his conduct in this world and by renouncing many of its advantages, for citizenship in a better world that he has in his idea. [Here, as always in the Dialectic, 'idea' is a Kantian technical term, meaning 'concept of reason'.] This powerful and incontrovertible proof is reinforced by our ever-increasing knowledge of •purposiveness in everything we see around us, and by contemplation of the •immensity of creation, and therefore also of a certain •limitlessness in how far our knowledge might be extended and in our •drive to extend it accordingly. All this still remains to us after the critique has done its work; but we must give up all hope of grasping the necessary continuance of our existence merely from our theoretical knowledge of ourselves.

Concluding the solution of the psychological paralogism

The dialectical illusion in rational psychology arises from the confusion of an idea of reason—the idea of a pure intelligence—with the completely featureless concept of a thinking being as such. I think *myself*—in the all-purpose *I think*—for the sake of a •possible experience, at the same time abstracting from all •actual experience; and from my ability to do this I infer that I can be conscious of my existence even apart from experience and its empirical conditions. In doing this I am confusing •the possible *abstraction* from all the empirical details of my existence with •a supposed consciousness of a possible *separate existence* of my thinking self, and that leads me to think I have *knowledge* that what is substantial in me is the transcendental subject. But really all that I have in thought is the unity of consciousness. . . .

The task of explaining how the soul relates to the body doesn't properly belong to •the psychology I'm discussing here, because •it aims to prove the personhood of the soul even when it is not related to the body (i.e. after death), so that •it is *transcendent* in the proper sense of that term. It does indeed occupy itself with an object of experience, but only in the aspect of it in which it ceases to be an object of experience. My doctrine, on the other hand, does supply a sufficient answer to this question about how the soul relates to the body, including the question of whether and how they could act on one another causally. It's generally recognised that what makes that problem especially difficult is the belief that

(1) the object of inner sense (the soul)—the formal condition of its intuition = **time** only, is basically unlike

(2) the objects of the outer senses (·bodies·)—the formal condition of their intuition = **space as well as time**.

428 But the two kinds of objects differ from one another not intrinsically but only in so far as (2) *appears* externally to (1); whatever thing-in-itself underlies (2) the appearance of matter may after all not be so radically unlike (1) the thinking subject. When you bear that in mind you'll find that this difficulty vanishes. The only question that remains is this:

•How is it possible for *any* two substances to interact causally?

But that question lies outside the domain of psychology; and you won't hesitate to agree, in the light of what I have said in the Analytic regarding basic powers and faculties, that the question lies outside the field of all human knowledge.

Moving across from rational psychology to cosmology

The proposition *I think* or *I exist thinking* is an empirical one. So it is based on empirical intuition, and thus on how the object of the intuition—which in this case is the *I*, the thinking subject—presents itself as an appearance. It seems to follow on my theory •that the soul, even in its thinking, is completely transformed into appearance, and •that in this way our consciousness itself, as being a mere illusion, must amount to nothing.

[Kant next discusses a different *I think*—the all-purpose one that is involved in *any* thought, in thought *as such*, the *I think* that has been the focus of the Paralogisms. *This* is a logical puller-together [*logische Funktion*, 'logical function'] of whatever variety of elements intuition may present me with; it's something that I actively *do*, not something that I sensibly

= passively *encounter*. So it doesn't exhibit ·this *I*·, the subject of consciousness, as an appearance; it doesn't exhibit me *as* anything at all; it is involved in *all* the intuitions that I have, both sensible (passive) and intellectual (active), so it can't itself have any features that would tie it to one or other of those two kinds of intuition. In this *I think* I don't, of course, represent myself to myself *as I am in myself*, but nor do I represent myself to myself *as I appear to myself*. And if represent myself as (1) a subject of thoughts or as (2) a ground of thought, I am not here using the categories of (1) substance or of (2) cause. That's because the categories are operators on materials supplied by our sensible intuition, but the *I think* that we're discussing here isn't among those materials—it's the *doer* that pulls the materials together under a single consciousness. Kant ends this amazingly difficult paragraph thus:] In the consciousness of myself in mere thought I am the *being itself*, without providing any facts about myself for me to think about.

The proposition *I think*, understood as amounting to *I exist thinking*, is no mere logical puller-together; it says something *about* the subject (which in this case is also the object) regarding its existence; it requires an inner-sense intuition that presents the object not as a thing-in-itself but merely as an appearance. So here we have not simply 430 (1) activity of thought but also (2) ·passive· receptiveness of intuition—i.e. we have (1) the thought of myself applied to (2) the empirical intuition of myself. Let's pretend that I want information about how my thinking subject goes about its pulling-together work when it is applying the categories of substance, cause and so on. How can I go about enquiring into this? •I need more than merely the all-purpose *I* that accompanies all thought, because I'm looking for *information about* myself as an actively thinking subject, and the all-purpose *I* contains no information. •And my inner-sense

intuition of myself as an appearance won't do the job either. Why not? Because there would be a kind of circularity in trying to learn about the pulling-together work of the *I* from an intuition that is itself an appearance pulled together by the *I*. So· what I need to carry out this pretended inquiry is an intuition of myself that enables me to know myself as a *noumenon*. And that's impossible, because the only intuition that I have of myself is a sensible one providing only data of *appearance*. . . .

Suppose that the following were true ·as indeed it is·:

We eventually discover—not in experience but in certain *a priori* laws of the pure use of reason (laws that are not merely logical rules but concern our existence)—grounds for regarding ourselves as *legislating* completely *a priori* in regard to our own *existence*, settling what sort of things we are. This reveals in us a *spontaneity* through which we determine our reality with no need for the conditions of empirical intuition. And we also become aware of something else. Although our existence can't be *thoroughly* determined other than through sensibility, we become aware that the consciousness of our existence contains *a priori* something that can—by virtue of a certain inner •power—serve to determine our existence in its relations to a non-sensible intelligible world.

This wouldn't contribute *anything* to the project of rational psychology. In this marvellous •power that my consciousness of the moral law first revealed to me, I would have for the determination of my existence a principle that is purely intellectual. But what predicates would I use in doing this? They would have to be just the ones that are given to me in sensible intuition; which means that as regards rational psychology I would be exactly where I was

before ·practical reason was brought into the story·. I can't have knowledge of myself unless I can make use of my concepts-of-understanding such as substance, cause, and so on; and I can't give meaning to *them* except with help from sensible intuitions; and sensible intuitions can never help me to move beyond the domain of experience. Still, in my practical thinking (which is always directed to objects of experience), it is all right for me to apply these concepts to freedom and the subject that has the freedom, giving them meanings that are analogous to the meanings they have when used theoretically. In doing this, however, I would be using these concepts merely to capture the •logical functions of

•subject and predicate, ground and consequence,
·and not the full-fledged schematised concepts of

•substance and property, cause and effect·.

That would enable me to think of the acts that I perform 432
in conformity to ·moral· laws as always capable of being explained in terms of the laws of nature and the categories of substance and cause, although they come from an entirely different source. I needed to make these points so as to head off any misunderstanding of my doctrine about our *appearing* to ourselves in self-intuition. I'll revert to these matters later.