

On Strawson On Kantian Apperception¹

Dennis Schulting

University of Amsterdam
Department of Philosophy
Nieuwe Doelenstraat 15
1012 CP Amsterdam
The Netherlands
Email: d.schulting@uva.nl

Abstract:

Strawson famously argues that Kant's argument for the necessary conditions of experience can only be retained once freed from *a priori* synthesis. Strawson claims that a purely 'analytical connexion' between experience and the object of experience is conceptually inferable from a thoroughly analytic premise concerning the capacity for self-ascription of representations. In this paper, I take issue with the way in which Strawson construes the analyticity of the principle of self-ascription or what Kant calls the principle of transcendental apperception. More particularly, I shall argue that Strawson's unity argument, viz. his construal of the *unity* of consciousness, on which the principle of self-ascription depends, suffers from a modal fallacy. Whilst arguing this, I shall suggest that *a priori* synthesis is required even for *analytic* unity of consciousness to be possible.

P. F. Strawson's take on a core element in Kant's argument, in the Deduction of his *Critique of Pure Reason* (hereafter CPR),² for the unity of consciousness as a necessary condition for the possibility of experience in general is a familiar one. Strawson argues in his classic *The Bounds of Sense* (Strawson 1966) that what can be salvaged

¹ This article is an abridged and modified version of a chapter from Schulting, forthcoming. It was presented at the In Memory of P.F. Strawson Philosophy Conference, held at the University of Witwatersrand, Johannesburg, South Africa, 22-23 September 2007. I thank the organisers, especially Lucy Allais, for inviting me to give a presentation. I also thank the participants Quassim Cassam, Thaddeus Metz, Murali Ramachandran, Paul Snowden, and Scott Stapleford for their comments and discussion. An early version was read at a seminar meeting of the Instituto de Investigaciones Filosóficas UNAM (IIF) in Mexico City in November 2005. I hereby gratefully acknowledge the opportunity provided by the IIF through their fellowship scheme, which enabled me to draft the paper. Many thanks are also due to Steven M. Bayne for his very helpful comments on a rough version of the paper, to Manfred Baum who read a later version, and to Christian Onof for his invaluable suggestions for improving the penultimate draft, in particular in the matter of the logic. I am solely responsible for what remains obscure or faulty in it.

² Citation from Kant's *Critique of Pure Reason* is according to the standard abbreviation A/B followed by the respective page numbers in the original. Quotations in English are from the Guyer/Wood translation in *The Cambridge Edition of the Works of Immanuel Kant* (Cambridge: Cambridge University Press, 1997, corrected version 2003). References between brackets are to the Akademische Ausgabe of Kant's works in German (Berlin: de Gruyter, 1900-) by means of the abbreviation AA, followed by Roman volume number, page number(s) and line number(s).

from Kant's argument amounts to a certain self-reflexiveness between the unity of consciousness and the perception of an objective world that is perceived as being external to, and thus distinct from, oneself as the one perceiving. The self-reflexiveness basically consists in the reciprocal relation between perceiving and perceived. While there must be someone doing the perceiving, the perceiving points to something that there must be to be perceived. More in particular, the unity of consciousness, one's subjectively perceiving something to be the case, must be seen as necessarily dependent upon what is objectively perceived, namely the world of particulars in and through which the subject patterns a trajectory of experience. The self-reflexivity at issue is in essence tantamount to a conceptual connection; indeed, as Strawson (1966: 96) observes, it provides the 'direct analytical connexion' that one would wish to see in Kant's argument. The argument for this connection is generally presented as a transcendental argument.³

These insights, which Strawson believes lie hidden behind Kant's sometimes arcane reasoning and should be able to be extracted from it by virtue of scaling away the speculative excess, are crucial. They are so for two reasons. First, they shed light on the exigencies for a philosophically meaningful construal of the systemic thrust of Kant's argument in the Deduction. Secondly, they open up an avenue for bringing much needed clarity to the ongoing debates in epistemology on the relation between the beliefs that a self entertains apropos of the world and their answerability to the world, on the way that the conceptual form of judgment latches onto non-conceptual content, on issues regarding justification and truth, and so forth.

However, I am not engaging in the standard debate on the nature of transcendental arguments and their wider philosophical relevance. Nor do I reflect on the broader philosophical implications of Kant's arguments. My reason for considering Strawson's construal lies in the fact that it singularly directs our attention to an element of Kant's argument that is universally misunderstood (including, I shall argue, by Strawson himself) as a result of which, I contend, a proper understanding of Kant's main claim in the Deduction has not been possible up until now. This element concerns the precise meaning or sense of the unity of consciousness, and hence the meaning of the premise of the so-called transcendental argument. I believe that this is not just an interpretive issue in Kant scholarship, but bears on a philosophical problem in the analysis of knowledge that goes far beyond Kant exegesis. It is therefore worthwhile to cast some light on it.

I fully concur with Strawson's line of reasoning that an analytical connection and hence a reflexivity of sorts exists between, on the one hand, the subject of experience (and accordingly the premise of the putative transcendental argument) and, on the other, *that of which* one has experience, i.e., objects in the outer world (and accordingly, the conclusion of that argument). However, I take issue with Strawson's reconstruction of this reflexivity. My claim is that Strawson's reconstruction of Kant's argument suffers from a modal fallacy concerning the premise of the argument. Furthermore, I believe that the fallacy in Strawson's reasoning jeopardises the viability of constructing the kind of argument that Strawson wishes to extract from Kant's reason-

³ Notice that Kant does not talk about transcendental arguments as such, although like Strawson, I believe that Kant does have an argument that proceeds from the premise of self-consciousness to objectivity. I also agree with Strawson that the Transcendental Deduction is not merely an explanation of objective experience, but a proof too (cf. Strawson 1966: 88). I scrutinise Kant's own argument in Schulting, forthcoming. I shall use the epithet 'transcendental argument' in a broadly Strawsonian sense.

ing. I cannot of course, in the space of a paper, attend to every aspect of the problem under discussion. I shall therefore concentrate on Strawson's reconstruction of the premise of the argument. Much depends on the way the premise is construed.

In section 1, I survey Strawson's reconstruction of the transcendental argument. In section 2, I examine Strawson's reading of the *analyticity* of the principle of apperception or self-ascription and note some essential differences with Kant's own position. I subsequently point out that Strawson crucially fails to distinguish between two distinct ways of conceiving of the *unity of consciousness* of an experiencing subject (section 3); and that as a result Strawson does not heed the difference between two kinds of claim regarding the *necessary unity* or *unifiability* of representations (section 4). I thus address three closely related aspects of apperception: analyticity, unity, and necessity. In the course of pointing out the aforementioned failure regarding modality, I indicate that it affects the constructing of a transcendental argument in the way that Strawson does. I also suggest that *a priori* synthesis, which is disallowed by Strawson's argument for analyticity, avoids the problem regarding modality.

1. Strawson's Core Argument

Let me first review Strawson's core argument. In outline, Strawson's reconstruction of the 'direct analytical connexion'⁴, which is laid bare by the so-called transcendental argument (T), goes thus⁵:

- T1. Any experience is necessarily unifiable in a unity of consciousness, by virtue of the possibility of self-ascription, the condition of which is transcendental apperception, so as to show self-sameness of representation or a belonging to a single consciousness, viz., an *analytic* unity of consciousness
- T2. Unity of consciousness requires another unity, that is, an *objective* unity, the connectedness of which presupposes the employment of empirical concepts of the objective, these concepts being the rules that govern the connection of experiences so as to allow a differentiation of the objective from the subjective and hence to enable re-identification of one's variant representations (cf. Strawson 1966: 87)

Essentially, it is actual knowledge of the objective world, or in effect the concepts employed for such knowledge, that furnishes the material condition of the possibility of consciousness of 'my diverse experiences as one and all my own' (Strawson 1966: 107); that is to say, the unity of my consciousness derives ultimately from the unity of the world. It is in this way that my experience must be seen to depend on *that which* I experience, or put differently, *what* I experience enables my very experiencing. This concerns what Strawson calls the conceptualisability of experience, in that any experience gives rise to the *thought* of experience (see Strawson 1966: 107). The thinkability of experience consists in the ability to differentiate the experience, as an objective event, from the *experiencing*. The relation of dependency between the *experiencing* and its object thus manifests that there obtains a self-reflexiveness of experience. This, then, yields the following conclusions:

- 4 This means that the relation between the premises of T is one of deductive inference. Other than a standard logical inference or syllogism, however, T's premises are not independently had but are, in some way, *necessarily* related. T is a prototypical case of *a priori* conceptual analysis.
- 5 T is Strawson's *reconstruction* of Kant's argument; it is not necessarily *Kant's* argument. In fact, vital steps in Kant's argument are lacking in T. T is expounded in section 7 of Strawson's *The Bounds of Sense* (Strawson 1966: 97ff.).

T3. The self-reflexiveness of experience, viz., the differentiability between experience and that of which the experience is, shows that there exists an ‘analytical connexion’ between experience and objectivity (from T1, T2)

T4. Objectivity and the knowledge thereof are shown to be the necessary condition or the enabling condition of experience in general (from T3)

If Kant’s deductive argument is reconstructed in this way, it is clear that there is no need for an *a priori* synthesis of one’s identical representations so as to first ground rationally the concept of objective experience. What suffices is to demonstrate by means of inferential reasoning the incontrovertible truth of a unified objective world as the basic enabling condition of experience in general; and to show that we thus have knowledge, in virtue of the ability to form epistemically pertinent judgments, of what Strawson calls ‘objects in [the] weighty sense’ (1966: 73). At any rate, the concepts of the objective are not to be associated with the categories, for according to Strawson (1966: 87, 88) there is no intrinsic link between the epistemology of experience and the forms of logic. Furthermore, Strawson’s reading, which disallows *a priori* synthesis, is fully coherent with naturalism about the objective nexus of appearances, the existence of which must be presupposed so as to allow the *a posteriori* synthesis of representations in conscious experience (cf. Guyer 1980: 205, 208-209, 1987: 142; Hossenfelder 1978: 100-102). Transcendental apperception, which grounds unitary consciousness, is thus seen merely as a necessary conceptual tool in the explanation of natural phenomena and their connections and a subject’s experience thereof.

There are all sorts of interesting elements implicit in the argument’s chain from T1 to T4 that I find troublesome and therefore require clarification. I want to concentrate on a problem concerning premise T1. This regards a modal fallacy. The fallacy issues from a mix-up between two kinds of conceiving of the unity of consciousness, as well as from a certain (mistaken) understanding of analyticity that Strawson reads into Kant’s own position on the analyticity of the principle of apperception (CPR B135 [AA III: 110]). The mix-up I am referring to is evidently not a conflation of the two unities that Strawson clearly distinguishes, the analytic unity of consciousness and an objective unity, although in some way the latter distinction is related to the mix-up that worries me. Let me first turn to the aspect of analyticity. In section 3, I shall then come to speak of unity, after which, in section 4, the modality of the apperception principle will be addressed.

2. Analyticity

I want to argue that Strawson’s unity argument (T1) suffers from a modal fallacy. I believe this is partly due to a particular conception of analyticity, which wants to dispense with *a priori* synthesis. The premise of the transcendental argument that I outlined above, and which concerns the unity argument (T1), is presented as amounting to a strictly analytic principle. This is the principle of the self-ascription or apperception of representations. I argue that the way Strawson construes this principle is problematic. The link presumed between self-consciousness and objectivity (T2) is, as a result, vulnerable to the flaws of this particular construal of self-ascription or apperception. Whatever the case may be regarding the inferences drawn from T1, let us consider the principle of self-ascription in more detail, as a run-up to an assessment of the modal fallacy in Strawson’s unity argument further below.

Strawson apparently takes Kant’s premise, to wit the familiar proposition ‘The: I think must be able to accompany all my representations’ (CPR B132 [AA III:

108.19]), with which Kant launches his operative argument in the B-Deduction, to express the analytic principle PS

{PS=Principle of Self-ascription}: $(\forall x)(x \text{ is a representation and } x \text{ is being represented at time } t) \rightarrow \text{necessarily } [(\exists y)(y \text{ is a thinker 'I'}) \wedge (x \text{ is self-ascribed or at least self-ascribable under certain conceptual constraints by } y)]$

Call PS the definition for self-ascription or apperception, which explains the possibility of subjective experience or self-consciousness. There are grounds for thus formulating it in the text of Kant's Deduction.⁶ For Kant himself states regarding the possibility of self-ascription of representations:

This last proposition [viz., that unity of consciousness is only possible through synthesis, D.S.]⁷ is [...] itself analytic [...] for it says nothing more than that all my representations in any given intuition must stand under the condition, under which alone I can ascribe [*rechnen zu*] them as *my* representations to the identical self and thus grasp them together [*zusammenfassen*] as synthetically combined in one apperception through the general expression *I think*. (CPR B138 [AA III: 112.13-19]; translation slightly amended)

PS is routinely taken to explain a *de dicto* necessity: I must be able to ascribe to myself any representation that I have, I being the subject of any of a series of representations that I ascribe to myself, for which certain conceptual conditions for unification should be met (cf. CPR A122 [AA IV: 90.18-20]). By implication, *de facto* self-ascription establishes, *a posteriori*, a synthetic existential unity of all representations so ascribed as belonging to the unity of consciousness; however, nothing in the way of *a priori* synthesis seems thereby required.⁸

Presumably, the analyticity of PS has to do with self-ascription being criterionless, that is, immune to error through misidentification: one knows and cannot fail to know the conditions under which one ascribes one's representations to oneself (cf. Strawson 1966: 92, 93, 98, 165). Contrary to the application of concepts to objects, in regard to one's own representations ostensibly no identificatory criteria are required that first enable their self-ascription and no possibility of error exists: the concept of self applies

⁶ I disregard the putative differences between the account of apperception in the A- and B-Deductions.

⁷ Not surprisingly, this reference to an earlier section in the text is ignored by commentators who stress the analyticity of the self-ascription principle. Strawsonians, who insist on the analytic nature of the principle, namely fail to assess Kant's assertion, earlier at B135, that 'this principle of the necessary unity of apperception is, to be sure, itself identical, thus an analytical proposition, yet it *explains* [*erklärt*] as necessary a synthesis of the manifold given in an intuition, without which that thoroughgoing identity of self-consciousness could not be thought.' (CPR B135 [AA III: 109.19-23]; my emphasis, translation amended. Notice that for Kant the notions of necessity and apriority are inseparably linked [cf. CPR B4]). A good example of the difficulties that the Strawsonian runs up against in regard to Kant's analyticity claim as a direct result of this failure of assessment, is provided by Cassam's reflections on this score (Cassam 1987: 375ff.).

⁸ According to Guyer (1980: 212), the notion of apperception, as a consciousness of one's self-consciousness, can only be retained once freed from the 'encumbrance of a priori synthesis'. In other words, on the analytic reading of the principle expounded by Kant at B138, any synthetic unity that would be involved could only be *a posteriori*, viz., a unity of all *de facto* ascribed representations, not of all possible (i.e., past, present and future) representations (cf. Strawson 1966: 96). At any rate, this gainsays Kant's view, expressed in the omitted passage in the above quotation from B138, that the synthetic unity at issue is an *a priori* unity that is the very *ground* of any analytic unity of consciousness (see B135; cf. B134 [AA III: 110.9-11], where Kant speaks of 'synthetic unity of the manifold of intuitions, as given *a priori*' being 'the ground of the identity of apperception itself').

to one entity and one entity only, namely the self that I am when ascribing my representations to myself. More precisely, the extension of the concept of self consists of just one possible *particular* instance at any one time at which the concept is instantiated by some self who is self-consciously aware of her own representations. In any case of representing, I *a fortiori* know, *by way of* self-ascribing any representation that I have, that I am the one representing. As it appears, the analyticity of PS would thus concern the logically trivial truth that the ability to conceive of one's representations as one's own is reciprocal with the capacity to employ the indexical 'I' in all cases of such conceiving. That is to say, there is an analytic, conceptual relation between a representation and the agent of representation, which is the self-ascribing representer, or the thinking 'I'. Paul Guyer (1980: 209) puts it quite explicitly by contending that Kant holds that '[w]hatever is to count as a representation at all must be fit for self-ascription.⁹ [The 'I think'-proposition] asserts [...] that I cannot have a representation which is not subject to these conditions [i.e., the conditions for self-ascription, D.S.]. To put it bluntly, Kant asserts that I cannot have a representation which I cannot recognize as my own.'¹⁰

On this account, it seems that not only the conditions for representing (that is, for *having* representations) and the logical conditions for self-ascription of representations are conflated, but also the conditions for representing and *self-consciousness* given that, as Strawson (1966: 108) asserts, transcendental self-consciousness is the *a priori* form or condition of self-ascription. A representer could thus not be otherwise than an at least potential self-conscious representer. I believe that this view of the analyticity of the apperception principle is flawed, for it provides no ground for assuming that any agent of representation is *eo ipso*, even if only potentially, a self-conscious subject or that the subject who envisages his own future states of affairs has complete knowledge of future states of affairs as involving himself.¹¹

Notice that, as regards T1, Strawson (1966: 92) claims that experiences are necessarily unifiable in that they must satisfy the conditions of belonging to a single consciousness. This is a rather different claim from the one regarding the criterionless nature of self-ascription. It seems, then, that Strawson confuses two different arguments: one concerning the logical conditions governing the self-ascribability of one's *own* representations and another for the necessary unifiability of representations tout court, only the former of which would *prima facie* amount to a self-evident truth of logic, viz., the tautology that Strawson considers to be the nub of Kant's premise. This confusion relates to a modal confusion of which Strawson is guilty with respect to premise T1, namely with respect to the sense in which one should understand the *unity* of consciousness to which one ascribes one's representations. This is what I argue in sections 3 and 4. But let me first return to Strawson's understanding of the *analyticity* of self-ascription.

To illustrate the austere conception of apperception as a condition for representation in terms of PS, which dispenses with *a priori* synthesis, consider Malte Hossenfelder's interpretation of the 'I think'-proposition (Hossenfelder 1978). Having noted that there

9 Strawson speaks similarly of any representation's *potentiality* for self-ascription (Strawson 1966:101, 114, 117).

10 Notice that Guyer is critical of this view that he attributes to Kant.

11 Not all future representative states *need* be ones that I *self-consciously* represent, lest the conceptual condition for self-ascription be seen as concerning an ontological necessity, implying a necessary coexistence of representer and self-consciousness (cf. Ameriks 2000b: 249).

are ostensible intrinsic problems with Kant's appeal to *a priori* synthesis and assuming that the principle of self-consciousness is a tautological principle as previously defined (PS), Hossenfelder (1978: 100-101) attempts to cast light on the analyticity of the principle by suggesting that we substitute 'to represent' for the verb 'to think' in Kant's proposition 'The: I think must be able to etc.'. The proposition would then read 'The: I *represent* must be able to accompany all my representations.' Only in this way, Hossenfelder argues, can the analytic character of PS become explicit, for quite clearly its denial logically entails a contradiction. The premise of Kant's argument is then tantamount to nothing more than the unpacking of what is already contained in the concept of 'representation'. Hossenfelder thus reduces apperception to a conceptual principle of *representation* tout court.¹² We can translate Hossenfelder's substitution reading of PS as:

{PS'}: $(\forall x)(x \text{ is a representation and } x \text{ is being represented at time } t) \rightarrow \text{necessarily } [(\exists y)(y \text{ is a representer 'I'} \wedge (x \text{ is self-ascribed or at least self-ascribable by } y))]$

However, Kant himself does not regard the principle of self-consciousness as simply a principle of *representation*, so that the analytic (conceptual) relation obtains between represented and representer.¹³ But even if disregarding this historical point, it is not true that the said conceptual connection is *eo ipso* substitutable, in all possible cases, for the relation between a representation and a *self* who self-ascribes her representations to her identical self. First, a representer could just *be* representing without *self-ascribing* representations at all — this would amount to first-order representing without a second-order *representing* of one's representing by virtue of the self-ascription of representations to one's identical self (or to a perceiving without apperceiving)¹⁴. More intriguingly, a representer could be representing representations, or indeed ascribing representations to herself (through a self-reference of sorts), without however thereby self-ascribing them to a self in the strict sense, by which I mean the *same* self (*de re*) to which she also ascribes other representations (over time). It is possible even that a representer could effectively (*de re*) ascribe representations to 'others' when in fact she *believes* that she is ascribing them to herself (*de dicto*) (cf. e.g. Kant, CPR A363 [AA IV: 228.32-229.4] and A363-64 note [AA IV: 229]). (This involves problems concerning the metaphysical status of the identical self to which one ascribes representations, which I must leave aside for present purposes.)¹⁵

If we look at the cases of satisfaction of Kant's apperception principle, then we learn by analysis that apperception cannot be a condition of representing tout court (as on PS'). It is not at all the case that *all possible* representations (*sic*) are necessarily ac-

12 Hossenfelder's reconstruction of the 'I think' proposition is comparable to Reinhold's, of which Ameriks has shown that it does not fit Kant's view of apperception (Ameriks 2000b: 107ff.).

13 Notice that Kant's phrase at B131 (AA III: 108.20-21) continues: '...for otherwise something would be represented in me that could not be thought at all', which would make no sense on Hossenfelder's substitution proposal. Kant's suggestion is that in case the 'I think' would not accompany my representations something would still be represented (in me) but I would not *think* it, which is trivially true, but not in Hossenfelder's sense.

14 Cf. Kant, *Anthropology from a Pragmatic Point of View*, §5 (AA VII: 135).

15 I believe that the substitution by some of *de se* modality for the distinction of *de dicto/de re* in the case of self-consciousness glosses over the problems involved in attempts to determine the ontological status of the self underlying apperceptive self-consciousness and is therefore wholly stipulative. See in general Ameriks's authoritative account of Kant's metaphysics of the self in Ameriks 2000a.

accompanied by an ‘I think’, nor that all representations (*sic*) necessarily entail the (transcendental) unity of apperception; and nor do all of them effectively belong, necessarily, to the thoroughgoing identity of my self-consciousness (in the possessive¹⁶ sense). This can be demonstrated in a breakdown of the ‘I think’-proposition into its logical modalities.

Assume necessary possibility P₁: *de facto*¹⁷, ‘I think’ *accompanies all my representations*

if P₁, then, *ex hypothesi*, it must also be possible that:

P₂: *de facto*, ‘I think’ *does not accompany all my representations*

and/or:

P₃: *de facto*, ‘I think’ *does not accompany any other representations that happen to occur and are so occurrent in the mind at any time t at which the ‘I think’ is not instantiated*

and/or:

P₄: *de facto*, ‘I think’ *does not accompany any other representations that happen to occur and are so occurrent in the mind at any time t at which the ‘I think’ is not instantiated, and which are also interminably barred from being able to be so accompanied, i.e., such representations that evanesce immediately after having been prompted and leave no significant traces for possible retention and ‘taking up’ by an act of apperception (some representations may simply not be able to be retained or retrieved)*

P₂ is obviously spurious, for it is logically inconsistent for me, as the subject of thought, to assert that ‘I’ am thinking (*de facto*) — or to assent, whilst thinking, to the proposition ‘I am thinking’ — and yet not to accompany my representations that I am *thereby* thinking. The possessive pronoun ‘my’ in the predicate ‘all my representations’ refers rigidly. Those representations are *my* representations that *I* accompany as such by *effectively* thinking them.

In contrast, P₁ is true by definition; it expresses quintessentially the principle of identity, which is the first principle of discursive reason¹⁸. The totality of *my* representations that are occurrent share the *same* common mark ‘I think’, just in case I am accompanying them (as *my* representations ‘all together’ [*insgesamt*], as Kant puts it [cf.

16 I borrow this way of putting it from Ameriks 2000b: 281.

17 Notice that this adverbial phrase indicates that here an analysis, *ad oculos reflexionis*, of the possible cases of satisfaction of the ‘I think’-proposition in terms of its logical purport is concerned (cf. Deppermann 2001: 130); it is not suggested that an *actual occurrence* of *empirical* consciousness is at issue at this point in the argument (cf. Reich 1992: 27), even though Kant says elsewhere that the proposition itself is empirical (CPR B420 [AA III: 274:15-20]). By this I take Kant to mean that the existence *implied* in the cogito is necessarily (for us, human beings) in time and so reveals the requirement that an empirical manifold in inner sense is given *a posteriori* to thought. But this requirement does not affect the *logical thrust* of the cogito-argument.

18 Cf. Kant AA II: 294. Cf. also CPR B408 (AA III: 268.7-9). I must refrain here from further analysis.

CPR B132)]¹⁹ by means of the act of thinking, precisely *when* I am in the business of thinking (representing in a particular way).²⁰

P₃ reflects the case of a representer R representing any arbitrary occurrent representation x, y, or z. Whilst in this case P₁ is not satisfied, R would nonetheless be the representer of x, y, z, even if not aware of herself (*stricto sensu*) as in the business of representing and *a fortiori* being *self-aware (stricto sensu)* of doing so. (R does not accompany her representations in the transcendental way, but merely in the empirical way by just *having* them in any arbitrary array peculiar to her actual physio-psychological stance at a particular time. Strictly speaking, R does not *think*.) Although Kant does not explicitly, at least not in the *Critique*,²¹ venture an opinion on the possibilities P₃ and P₄, of which it is further open to question if they are anything more than merely formally distinguishable, these are surely logically inferable from the 'I think'-proposition. This is confirmed by some of Kant's assertions in the text of the Deduction. P_{3/4}-representations are representations, which, as Kant puts it, are 'nothing for me' (CPR B132), which is consistent with the rigid reference of the possessive determiner 'my' of P₁-representations.

There are many more intriguing sides to Kant's principle of transcendental apperception that call for further analysis and exegetical backup. I provide these elsewhere (Schulting forthcoming). Here I want to return to the particular problem that I set out to address, namely the modal fallacy that I claim issues from taking apperception as an analytic proposition in the terms proposed by Hossenfelder (PS'), which are implicitly endorsed by Strawson. This fallacy can be brought to light by further focusing on two interconnected features of apperception: first, the kind of *unity of consciousness* that is established by the self-ascription of representations and, secondly, the modality involved in making a claim regarding such unity. Only a particular modal claim is compatible with a more strictly defined principle of apperception (which I introduce below), which is in accordance with the breakdown of Kant's 'I think'-proposition provided above.

19 The predeterminer 'all' in the predicate 'all my representations' creates an ambiguity, for Kant's proposition could, superficially, be construed such that it posits that the 'I think' does not effectively accompany all, but only some of my representations, which could lead one to presume that P₂ is not strictly speaking false. This is indeed the route that most interpreters take. Elsewhere (Schulting, forthcoming), I argue that this view is thoroughly mistaken and runs into exegetical difficulties. At the systemic level, in any case, (1) it is logically nonsensical to assert, from a first-person perspective, that *whilst* I am thinking, I am only thinking some of my representations that are occurrently represented; (2) of representations that are *not* occurrently represented I cannot tell whether they could be mine unless they are effectively represented by me, that is, accompanied by the 'I think', and so, by implication, unaccompanied representations are not strictly speaking *my* representations. This excludes readings of the apperception principle, which hold that representations are at any rate *potentially* subject to transcendental apperception, as a great many commentators believe. In Schulting forthcoming, I also explain that the predicate 'all my representations' is a single complex representation, which *as such*, and *only* as such, is accompanied (effectively) by the 'I think'. I call such a representation *r_{all}* as opposed to reach.

20 Whilst it would seem that I can only think one thought at a time, the nature of discursive thought, according to Kant, is such that every singular thought, which is accompanied by an 'I think', consists of several representations taken together and thus thought simultaneously, under one common denominator (the 'I think'), as same, viz. as 'all my representations' in terms of a compound thought; unity always implies multiplicity, which in turn entails synthesis to the extent that one's various representations are identical or equal, namely related to the identical 'I think'. Kant makes this clear in the course of sections 15 and 16 of the B-Deduction (see Schulting, forthcoming).

21 In his *Anthropology*, Kant provides ample concrete examples of P₃/P₄-representations.

3. Unity

As regards unity, Strawson fails to notice that one can take Kant's argument for the unity of consciousness, which is established by the act of apperception, in two ways, only one of which is correct. One can take it either (1) as an argument for the psychological or existential unity of representative states $r_{\text{each}^{-1}} \dots r_{\text{each}^{-n}}$ in terms of mental states *had* by a representer and as such aggregated in any arbitrary sequence in conformity with the way they are prompted to occur, through psycho-physiological intermediaries or brain states as their proximate causes, by external objects (thus: a unity of representations in a *de re* sense); or (2) as an argument for the unity of representative states in terms of certain states *recognised* and *identified* by the representer herself as together constituting a unitary compound r_{all} ²² that belongs to the representer as her own (a unity in the *de dicto* sense), whereby it should be noted that the representer here is an epistemic agent (a thinker) and not *just* a representer. It is the latter kind of unity that, I contend, Kant is in fact arguing for. The difference between these two kinds of taking the unity of consciousness amounts to the difference between arguing for (whereby R stands for representation and UC for unity of consciousness)

UC1. For all R, R is united in UC, given certain conceptual constraints that have to do with the capacity for self-ascription and material constraints connected with the way the world is

and arguing for

UC2. For all R, *if* R is self-ascribed and recognised by a self-ascribing representer (i.e., a thinker 'I'), *then* R is united in UC (regardless of material or psychological²³ constraints)

Not heeding the distinction between UC1 and UC2 is tantamount to committing a modal fallacy. This needs to be made explicit (see section 4 below).

Let me first take a closer look at how distinguishing between UC1 and UC2 affects the Strawsonian construal of apperception in terms of PS (or PS'). The recognition alluded to in UC2 is of course not a case of actively reflecting on the part of a psychological subject on her mental states (by way of muttering to herself, as it were). Instead, it points to a function²⁴ performed by the occurrently representing self in that by being self-consciously aware of her identity as the performer of this function, that is, as an epistemic agent rather than in the modality of being primitively aware of one's conceptually indistinct environment, she knows the conditions under which her representations acquire objective reality, rather than have merely subjective validity, and thus become cognitively or epistemically relevant. Identity of self-consciousness is a rule or function of recognising that one's representations belong together in a unitary

²² See regarding the predeterminer 'all' in contrast to 'each' note 19 above.

²³ Strawson of course would equally deny that the argument for the unity of consciousness has anything to do with psychological constraints; the constraints of self-consciousness at issue are rather conceptual. However, I believe that, given his reading of the premise in terms of PS, the conceptual constraints that Strawson wants to argue for in effect are necessary, if not yet sufficient, conditions of empirical consciousness *tout court*, and therefore psychological.

²⁴ Cf. CPR A108 (AA IV: 82.12), where Kant speaks of the 'identity of [a] function'; a few lines further down (p. 82.21) Kant, similarly, speaks of the 'identity of [the mind's] action'. Notice that by 'function' Kant understands 'the unity of the action of ordering different representations under a common one' (CPR B93 [AA III: 85.18-19]), which is precisely what is meant by 'the synthesis of recognition in the concept', as the heading of the section, in which the phrase 'the identity of the function' occurs, reads.

form, namely in what Kant calls the transcendental unity of consciousness or the original-synthetic unity of apperception.²⁵ Kant calls this function *a priori* synthesis. But this conception of self-consciousness, as *including a priori* synthesis, does not comport well with the principle of self-ascription as defined above (PS), which stipulated that any representation whatsoever is subject to self-ascription. We must now redefine PS as:

{PS''}: $(\forall x)(\forall e)(x \text{ is a representation and } x \text{ is being represented at time } t_1) \wedge (e \text{ is an epistemic agent}) \rightarrow \text{necessarily } \{[(\exists y)(y \text{ is an identical thinker 'I'}) \wedge (y=e) \wedge (x \text{ is self-ascribed by } y)] \leftrightarrow [(\exists z)(z \text{ is an analytic unity of all representations recognised and retained by } e \text{ after } t_1) \wedge (e \text{ recognises } x \text{ to belong to } z)]\}$

PS'' is a better translation of B138 than PS, for it takes into consideration Kant's explicit stipulation (in particular in the lead-up to B138) that certain *a priori* conditions, namely, *a priori* rules for recognition that together amount to *a priori* synthesis (i.e., the categories), must be satisfied in order for self-ascription to an identical self first to be possible. This explains why, contrary to what Strawson believes, *a priori* synthesis must be seen as closely linked up with there being an analytic unity of representations at all.²⁶ PS'' is effectively tantamount to a biconditional, for not only is self-ascription conditional on the recognition of a unity of representations (z), but z is also only possible under the condition of self-ascription. In fact, self-ascription is nothing but the constitution of z through the act of recognition. There is thus an analytical unity of representations z and an 'I' thinking it, if and only if 'I' effectively self-ascribe all my representations in accordance with the *a priori* rules for recognition (i.e., synthesis). It is this biconditional relation between the self and her representations that she self-ascribes (by virtue of recognising their same- or oneness) that determines the analyticity of PS'', for which the condition of *a priori* recognition by means of a rule for unification must thus first be met, to wit *a priori* synthesis. (This latter requirement, which first makes the principle of self-ascription analytic, seems to indicate that self-ascription is *not* criterionless, as on PS.) This implies that one is not licensed to argue that for all representations that are had by a representer it necessarily holds that they are ascribable to the *same* representer, nor *a fortiori* that all (possible) representations *eo ipso* belong to an analytical unity of consciousness recognised by an epistemic agent in accordance with *a priori* rules.²⁷ For a representer is not always an epistemic agent.

So how can Strawson vouch for the metaphysically intemperate claim that *all* representations that one has (or potentially has) are self-ascribable, and that *a fortiori* they make up an analytic unity of consciousness (UC1)? Clearly, there is no analytic relation between all representations had (or potentially had) and self-ascribability; so PS is not really analytic, as Strawson and Hossenfelder would have us believe (hence Hossenfelder's substitution proposal PS' in an attempt to make the principle's analyticity more plainly visible). Strawson disregards the conditional necessity underlying

25 Kant identifies the transcendental unity of consciousness as an objective unity in contrast to a subjective unity at B139 (AA III: 113). Strawson's argument for the objective unity as that on which the unity of consciousness is transcendently dependent, comes close to Kant's talk of transcendental unity in intent, but not in execution, for contrary to Strawson, who confuses transcendental apperception with the capacity for subjective consciousness tout court and differentiates it from an objective connectedness, Kant's objective unity is none other than the principle of transcendental apperception itself.

26 How this works precisely is a topic on which I elaborate elsewhere (see Schulting, forthcoming).

27 In the A-Deduction, Kant might seem to endorse precisely this reading (see e.g. CPR A113, A116, A117n.). I confront this ostensible discrepancy in Schulting forthcoming.

the apperception argument (as observed by PS''), which first establishes sameness or identity of self-consciousness, from which, in a second step, objective connectivity is analytically derived. This fallacy regarding modality can be made more formally explicit in terms of a failure to distinguish between two kinds of claim regarding the unity of representations established by self-ascription. This is the topic of the next section.

4. Necessity

Let me summarise. I have argued that Kant's premise is such that it cannot be about the trivial truth that every representation requires a representer — a truth that would indeed be conceptual. If the analyticity of the principle of apperception were to concern a merely conceptual truth (as on PS and PS'), the considerable attention given it in the literature over the course of two centuries would be undeserved. Instead, the premise is about what is required for a representation to be part of a unitary representation, an analytic unity r_{all} , that has an objective validity, i.e., that is 'something to me' as the *identical* subject of all my representations (rather than something that just exists, just *is* a mental occurrence and therefore has mere subjective value to the occurrent representer). (Objective validity here denotes the satisfaction of the conditions for sameness or identity.) This concerns the fact that a representation has the quality of objective validity if and only if it shares the same mark as is shared by other representations that belong to the unity of self-consciousness. And it shares this mark when I *actually*, self-consciously self-ascribe and unify it with all other representations that I concurrently self-ascribe.

Strawson suggests that the transcendental argument starts off from a trivial truth, which consists in the necessary unifiability of one's representations tout court, leading to a conception of objectivity as their unity's condition of possibility. This is problematic. If it is true that all representations that one has (*sic*) are necessarily unified, or at least unifiable,²⁸ in and by a single subject of representation (oneself), then it is not clear why such a claim would logically require, as Strawson thinks in virtue of the inferential force of the putative transcendental argument, a concept of the objective as a means of distinguishing between one's representing and the object of one's representation. Furthermore, a concept of the objective is a representation no less than any other representation, so what difference does its being invoked as enabling condition for the unity of consciousness make regarding the differentiation requirement, i.e., the differentiation of the objective from the subjective?

So, why must objectivity figure as the ground of the unity of consciousness, if it is the case that all one's representations are united as a matter of course? It is far from clear on which grounds Strawson believes the unity argument to *rest on* the conceptualisability criterion, for conformably to construal PS of the premise of the transcendental argument Strawson in fact just posits UC1, suggesting no further condition under which the subclass of representations of the objective is capable of being differentiated from the broad class of representations. Strawson *stipulates* that the former are

28 It could be countered that a distinction should be observed between the claim (A) necessarily, all R are unified in and by S (whereby R stands for representation and S for subject, viz. the thinking 'I') and the claim (B) necessarily, all R are unifiable in and by S. Notice that Strawson also does not appear to respect the difference between A and B. However, for the purposes of indicating the metaphysically in-temperate sense in which Strawson construes the premise of self-ascription, this distinction is not relevant, for in both cases A and B a claim is made with regard to the modally absolute sense in which R is related to S: no R is not subject to the condition under which it is either united or unifiable in and by S.

necessary to satisfy the requirement of the self-ascription of representations; but stipulating that they are necessary falls short of specifying the condition under which concepts of the objective enable a differentiation between what is subjective and what is objective; and to all appearances this is what Strawson fails to do.

There is a modal issue here requiring our attention,²⁹ for UC1 is tantamount to the following claim:

{AN=Absolute Necessity}: $(\forall x)(x \text{ is a representation}) \rightarrow \text{necessarily } [(\exists y)(y \text{ is an identical thinker 'I'}) \wedge (\exists z)(z \text{ is an analytic unity of representations}) \wedge (y \text{ thinks } z \text{ by way of self-ascribing her representations}) \wedge (x \text{ is united or at least unifiable in } z \text{ with all other representations self-ascribed by } y)]$

On account of AN, no criterion for identification of singular representations is needed so as to differentiate them from representations that do not share the mark constituting their sameness. Singular representations share by implication a unitary mark that identifies them as belonging to a self who self-ascribes them (hence the widely held belief that self-ascription is criterionless). *All* representations are subject to AN. AN underlies PS; recall that, according to PS, *all* representations are necessarily ascribed or at least ascribable in and by an identical thinker 'I' and so, by implication, belong *eo ipso* to a single unitary consciousness (UC1). What is unclear, however, is in what way, assuming PS, Strawson thinks that the objective unity (premise T2 of the transcendental argument), as a means of differentiation of the subjective and the objective, constitutes the necessary ground of the subjective unity of representations (premise T1). That is to say, it is unclear, on account of AN and given T's analytic nature, how T2 can be shown logically to be inferable from T1. What is the nature of the grounding relation? More precisely: *What is it that makes*, logically, T2 derive analytically from T1 such that T2 is necessarily entailed by T1?³⁰ Surely, it cannot be T2 itself. It must be some analytically explicable criterion inherent to T1 for the inference to work. I see nothing in T1, if construed as amounting to UC1, that points to such a criterion.

However, in order to prevent metaphysically intemperate claims of the kind that Strawson seems committed to and the resultant argumentative lack of clarity as regards the inferential relation between the various premises of T, let us suppose that the claim regarding the premise of the unity of consciousness comes down to a mere conditional (in conformity with PS''). The conditional would comport with construal UC2 and can be formulated as follows:

{CN=Conditional Necessity}: $\text{necessarily } [(\exists x)(x \text{ is a representation}) \wedge (\exists y)(y \text{ is an identical thinker 'I'}) \wedge (y \text{ self-ascribes } x) \rightarrow (\exists z)(z \text{ is an analytic unity of all representations self-ascribed by } y) \wedge (y \text{ thinks } z) \wedge (x \text{ is united in } z \text{ by } y)]$

It is evident that AN and CN spell out two distinct modal claims that should not be confused. In the case of AN, as we have seen, no representations are excluded from being unified, or unifiable, in and by the single subject of representation (the identical

29 I want to cash out the puzzling dual modality in the verbal phrase 'must be able' in Kant's 'I think' proposition. I analyse this feature, which concerns the deduction of the categories of modality, in detail elsewhere (Schulking, forthcoming).

30 Notice that although the conclusion of any arbitrary syllogism is analytically (hence necessarily) entailed, its premises need not themselves be necessarily related. This is different with the type of inference that is a transcendental argument (T), whose premises are necessarily related. My question thus concerns the force of T in terms of how each of its premises are conceptually (analytically) linked. I thank Thaddeus Metz for pressing me for a clarification on this point.

thinker 'I'). In the case of CN, a condition is specified to the effect that representations are united so as to show an analytic unity of consciousness, if they are taken together by the subject by way of self-ascribing her own representations. Other than with AN, the assertion regarding unity requires an antecedent condition for its satisfaction, viz., an act or function of identification that the thinking self operates by way of self-ascription of representations; self-ascription is a condition for unity that I argued is not fulfilled by mere representing alone. This suggests that representations are not unified as a matter of course (in the strict sense), nor necessarily subject to a condition of unifiability for that matter. With CN, no claim is made with regard to the putative existential unity of representations (UC1), whereas with AN a claim is made to the effect that representations could not *be* otherwise than united or at least potentially united in the unity of consciousness, which boils down to an existential claim as to the unifiedness or unifiability of representations (unruly representations that fail to fit into the unity of consciousness are ruled out on AN). Also, on account of CN, the analytic relation between T1 and T2, between the subjective and the objective, can be made clearer, for on this reading both subjective and objective representations are grounded on the same condition of recognition for unitary representation.

5. Conclusion

We face a dilemma. If Strawson wants to argue that there is an analytical connection between the subjective and the objective, then he needs something more than the trivial truth to which the premise of self-ascription, as Strawson construes it (PS), amounts. For, as I have argued, there is nothing intrinsic to that trivial truth that leads us, inferentially, towards the conclusion of T, that is, that objective connectivity is a necessary condition of the self-ascription of representations. The most Strawson can get out of an argument relying on PS is a short argument from representation (or experience, which for Strawson is equivalent to representation) to *that of which* the representation or experience is, i.e., to objectivity. This is in effect what Strawson argues in terms of the conceptualisability criterion. In this way, however, Strawson fails to explain the self-reflexiveness regarding self and objectivity *from* the premise of self-ascription. But we have also seen that Strawson not just advances the conceptualisability argument, but in fact argues for a different claim that reveals an unambiguous commitment to AN. These two claims are clearly in tension.

However, if Strawson were to concede to the conditional construal CN of the premise of T, then the anti-sceptical force a transcendental argument is presumed to have is significantly diminished.³¹ For the argument would boil down to a hypothetical argument regarding the unity of representations that will not cajole a sceptic into conced-

31 At times Strawson does appear to understand the argument such that it concerns a mere *de dicto* claim with regard to the necessary requirements for representations to have objective reference (e.g. Strawson 1966: 89; notice the implicit conditional structure of Strawson's claim here that '[w]e could not employ any ordinary empirical concepts of objects *unless* our manifold perceptual experiences possessed the kind of coherence and interconnexion which is required for the application of such concepts' [emphasis added]). In like manner, he argues that '*if* any phase of experience is to count as a phase of experience of the objective, we must be able to integrate it with other phases as part of a single unified experience of a single objective world', thereby ruling out 'unruly perceptions' [Strawson 1966: 89; emphasis added]). However, as Strawson observes (1966: 92), the Deduction argument in terms of a proof is not 'simply a matter of [giving] the definition of "experience" that experience involves knowledge of objects'. Hence, the premise of the real proof in the Deduction cannot be the actuality of knowledge of objects, and so for Strawson the argument cannot in effect be a conditionally construed inference.

ing defeat.³² On that reading, a radical sceptic could still persist in the conviction that only representations exist that are not unified in the strict sense of being self-ascribed to an *identical* self (over time) and so do not belong to an analytic unity correspondent with the objective ways of the world, but instead are nothing but mere aggregates of representations $r_{\text{each-1}} \dots r_{\text{each-n}}$ with no common mark between them that would constitute their sameness. Of course, Strawson wants to avoid this result at all costs. Therefore, given how Strawson understands the purport of the transcendental argument, a conditional construal of its premise is not what he appears to have in mind.³³

I cannot of course in the space of a paper assess all aspects of Strawson's influential reconstructive strategy for reading Kant's argument. What I have tried to show is that Strawson's premise (T1) reveals an ambiguity regarding modality. The fallacy of reasoning resulting from it can, I have suggested, be avoided if one heeds the conditional purport of Kant's argument, which would include a commitment to *a priori* synthesis. But Strawson rejects outright the latter and appears to neglect the former. Consequently, he fails, to my mind, to provide the genuine analytical connection between the subjective and the objective that he rightly wishes to highlight.

References

- Ameriks, K. 2000a. *Kant's Theory of Mind*. New Edition. Oxford: Clarendon Press.
- Ameriks, K. 2000b. *Kant and the Fate of Autonomy. Problems in the Appropriation of the Critical Philosophy*. Cambridge: Cambridge University Press.
- Cassam, Q. 1987. 'Transcendental Arguments, Transcendental Synthesis and Transcendental Idealism', *The Philosophical Quarterly* 37, 355-378.
- Deppermann, A. 2001. 'Eine analytische Interpretation von Kants 'Ich denke'', *Kant-Studien* 92, 129-152.
- Guyer, P. 1980. 'Kant on Apperception and A Priori Synthesis', *American Philosophical Quarterly* 17, 205-212.
- Guyer, P. 1987. *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press.
- Hossenfelder, M. 1978. *Kants Konstitutionstheorie und die transzendente Deduktion*. Berlin, New York: Walter de Gruyter Verlag.
- Reich, K. 1992. *The Completeness of Kant's Table of Judgments*. Stanford: Stanford University Press.
- Schulting, D. forthcoming. *Kant's Deduction From Apperception. On Explaining the Categories*.
- Strawson, P. F. 1966. *The Bounds of Sense. An Essay on Kant's Critique of Pure Reason*. London: Methuen.

³² A sceptic could point out that, taken thus, the transcendental argument would appear to rest on a *petitio principii*, for what had to be proved, viz., objective experience as a condition of subjective experience, is already assumed to be a fact in the premise.

³³ However, even though objective experience, or the concept thereof, is presupposed in the argument's premise as a fact, a fact that the sceptic does not accept, on CN it would still be problematic for a sceptic to *make a claim* as to the denial or impossibility of a necessary unity of representations correspondent with an objective unity, for a sceptic in the business of making such negative claims must nonetheless *eo ipso* be in the business of forming identical thoughts of her own through self-ascription, and hence would be subject to self-refutation.