On Knowing Which Thing I Am*

JOEL SMITH

Are there conditions that every self-conscious subject must satisfy? If so, what are they? These are intriguing questions. An answer may well draw upon metaphysics, epistemology, and the philosophy of thought and language. If it can be shown, maybe via some form of transcendental argument, that there are substantial necessary conditions of self-consciousness then philosophers engaged in these disciplines should take note. This is particularly so, given the current popularity of a broadly neo-Kantian programme in the philosophy of mind, one that shows a keen interest in such transcendental arguments. In this vein, it is tempting to appeal to work in the philosophy of thought and language, in order to secure such necessary conditions. One such tactic, employed by Strawson[1], is to argue, first, that self-consciousness necessarily involves self-reference and, second, that there are non-trivial epistemic conditions of reference and hence of self-reference. This strategy has been made more explicit in the work of two of Strawson's students, Evans and Cassam[2]. In particular, they argue that in order to refer to something in thought one must know which thing it is. This means that self-reference, and hence self-consciousness, requires that one know which thing one is. This, they continue, requires a great deal more.

What is it to know which thing one is? In the work of Strawson, Evans and Cassam the epistemic condition in question is this: in order to be able to think about an object (refer to it in thought) one must be able to distinguish it from all other things. This principle is dubbed, by Evans, 'Russell's Principle'[3]. Applied to self-referential

* Thanks to Naomi Eilan, Keith Hossack, Lucy O'Brien and Ann Whittle for helpful comments.

[1] P. F. Strawson, *Individuals: An Essay in Descriptive Metaphysics* (London: Routledge, 1959), and *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason* (London: Methuen, 1966).

[2] G. Evans, *The Varieties of Reference*, edited by J. McDowell (Oxford: Clarendon Press, 1982), and Q. Cassam, *Self and World* (Oxford: Oxford University Press, 1997).

[3] Named, of course, after Russell's Principle of Acquaintance. See B. Russell, 'Knowledge by Acquaintance and Knowledge by Description',

doi:10.1017/S0031819104000464 ©2004 The Royal Institute of Philosophy

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

thoughts, Russell's Principle requires that I be able to distinguish myself from all other things. Thus, knowing which thing I am is a matter of being able to distinguish myself from all other things. The Strawsonian strategy claims that self-conscious subjects must be able to distinguish themselves from all other things.

However, as Evans himself saw, an adherence to Russell's Principle has the consequence that a subject can suffer from the illusion of self-referential thought. Given a plausible account of self-consciousness, to be elaborated later, it follows that self-consciousness need not involve (the capacity for) self-reference. Hence we find that the second stage of the Strawsonian strategy actually serves to undermine the first. That is, the non-trivial epistemic condition of reference calls into question the claim that self-consciousness necessarily involves self-reference. Thus, Russell's Principle cannot be used in an account of the necessary conditions of self-consciousness that proceeds via the notion of self-reference. Any philosopher who wishes to pursue the intriguing questions with which I began, should look elsewhere for the source of the conditions of self-consciousness.

In the first section I make good my claim that there is a distinctive strategy in the work of Strawson, Evans and Cassam, that employs Russell's Principle for the purpose of generating necessary conditions of self-consciousness via the notion of self-reference. In the second, I argue that an application of Russell's Principle to self-referential thoughts allows for the illusion of self-referential thought. There is, therefore, a dilemma for the Strawsonian strategy: either accept Russell's Principle and sever the tight connection between self-consciousness and self-reference, or protect that connection through a rejection of Russell's Principle as applied to self-referential thought. Either way, the strategy fails to yield any necessary conditions of self-consciousness.

I

There are several places in which Strawson employs Russell's Principle to argue that such and such a condition is required for self-consciousness; I will mention two. I am not concerned to evaluate these arguments in their full complexity, merely to show

reprinted in, *Mysticism and Logic*, (London: Unwin, 1917), and R. M. Sainsbury, *Russell* (London: Routledge, 1979). For an alternative principle of acquaintance, see C. Peacocke, *Sense and Content: Experience, Thought and their Relations* (Oxford: Clarendon Press, 1983), ch. 7.

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

that the strategy I have pointed to is already present in Strawson's own work. I begin with the argument against dualism in Ch. 3 of *Individuals*. Strawson sums this up very nicely in the following,

> One can ascribe states of consciousness to oneself only if one can ascribe them to others. One can ascribe them to others only if one can identify other subjects of experience. And one cannot identify others if one can identify them *only* as subjects of experience, possessors of states of consciousness. (p. 100)

Here we are presented with the rejection of Cartesian dualism as a necessary condition of the possibility of self-consciousness. Or rather, as a necessary condition of the ascription of states of consciousness to oneself. We are to understand self-consciousness as involving self-referential thought. We are then told that a necessary condition of this is that we can 'identify other subjects of experience'. What does Strawson mean by 'identify'? He describes the notion of identification as follows,

> It seems that the general requirements of hearer-identification could be regarded as fulfilled if the hearer knew that the particular being referred to was identical with some particular about which he knew some individuating fact, or facts, other than the fact that it was the particular being referred to. To know an individuating fact about a particular is to know that such-and-such a thing is true of that particular and of no other particular whatever … This, then, is the general condition for hearer-identification in the non-demonstrative case; and it is obvious that, if a genuine reference is being made, the speaker, too, must satisfy a similar condition. (p. 23)

This description of identification is framed explicitly in terms of language. But Strawson sees no problem in transferring the basics of his account to the domain of thought. He claims that, 'Each of us can *think* identifyingly about such particulars without talking about them' (p. 61). The condition for identification, or reference, is that the subject is able to distinguish the object of reference from all other things[4]. This, of course, is Russell's Principle. Strawson is employing Russell's Principle as a device with which to generate strong necessary conditions of self-consciousness. Admittedly,

[4] In Part Two of *Individuals*, Strawson writes that, 'in order for an identifying reference to be made, there must be some true empirical proposition known, in some not too exacting sense of this word, to the speaker, to the effect that there is just one particular which answers to a certain description.' (p. 183).

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

Russell's Principle is not here being applied to the self-referential thoughts themselves, but to thoughts about others; thoughts the possibility of which is held to be a necessary condition of self-referential thought. Are there any instances of Strawson bringing Russell's Principle to bear on self-referential thoughts themselves?

In Part 2, Ch. 2 of *The Bounds of Sense*, Strawson argues that self-consciousness entails the unity of consciousness (consciousness of the possibility of self-ascribing diverse experiences), that the unity of consciousness requires that different experiences could be ascribed to the self same subject, that this requires that the subject has some way of distinguishing him or herself from other things, and that this requires that the subject have a grasp of the seems-is distinction. The aspect of the argument that is relevant here is the following,

> It is a quite general truth that the ascription of different states or determinations to an identical subject turns on the existence of some means of distinguishing or identifying the subject of such ascriptions as one object among others. Applying this general truth to the case before us, we may say, in Kant's terminology, that the possibility of ascribing experiences to a subject of experiences and hence the possibility of self-ascription of experiences requires that there be some 'determinate intuition' corresponding to the concept of a subject of experiences; or, substituting another terminology for Kant's, we may say that this possibility requires that there be empirically applicable criteria of identity for subjects of experience. (p. 102)

Here we see Strawson applying an epistemic condition of reference to self-referential thoughts. Whilst he is not explicit about what sense he is giving to the phrase 'distinguishing or identifying', his earlier adoption of Russell's Principle lends plausibility to the thought that this is what he has in mind here also. That is, self-consciousness requires self-reference and the conditions on self-reference require that further conditions hold. This is what I have called the 'Strawsonian strategy'.

Whilst Strawson did employ Russell's Principle to argue that such and such a thing is a necessary condition of self-consciousness, it was Evans who first gave an explicit account of the principle in a general theory of reference. Unsurprisingly, the principle derives from Russell. Russell articulated his principle of acquaintance in several places, one of which runs as follows,

> *Every proposition which we can understand must be composed wholly of constituents with which we are acquainted* … The chief

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

> reason for holding this principle true is that it seems scarcely possible to believe that we can make a judgement or entertain a supposition without knowing what it is that we are judging or supposing about…It seems to me that the truth of this principle is evident as soon as the principle is understood.[5]

Acquaintance is an epistemic notion; being acquainted with an object involves having a certain kind of knowledge of it. This knowledge is 'knowing which object it is that one is thinking about'. In line with his epistemological views, Russell himself thought that the only items with which we are acquainted in the requisite way were sense data, universals, our own mental states and, possibly, ourselves. Understood as involving these epistemological limitations, the principle of acquaintance has little to recommend it[6]. It follows from Russell's official position that either the proposition that London is the capital of England can be analysed into a proposition making reference only to sense data etc. or that it cannot be understood.

But it seems reasonable to think that the heart of the principle of acquaintance can be divorced from Russell's own epistemological strictures. This is the strategy adopted by Evans. Evans claims that to be in a position to entertain a singular thought about *a* one must know which thing *a* is, but he thinks that one can have this kind of knowledge about, amongst other things, everyday physical objects. The kind of knowledge that Evans believes is required for one to know which object one is thinking about is, what he calls, 'discriminating knowledge'.

> a subject cannot make a judgement about something unless he knows which object his judgement is about…the knowledge which it [Russell's Principle] requires is what might be called *discriminating knowledge*: the subject must have the capacity to distinguish the object of his judgement from all other things.[7]

This means that unless I have such a capacity (to distinguish the object of my thought from all other things), I cannot think about it. Following both Strawson and Evans, I shall sometimes refer to the

[5] B. Russell, 'Knowledge by Acquaintance and Knowledge by Description', reprinted in, *Mysticism and Logic* (London: Unwin, 1917), pp. 159–60.

[6] See, for example, R. M. Sainsbury, *Russell* (London: Routledge), pp. 26–41, who argues against Russell's position.

[7] G. Evans, *The Varieties of Reference*, edited by J. McDowell (Oxford: Clarendon Press, 1982), p. 89.

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

act of distinguishing an object from all other things as 'identifying' it. Now there are, according to Evans, three ways in which I can satisfy Russell's Principle with respect to any given object: (i) I can identify it by perceiving it, (ii) I can identify it by having the ability to recognize it were I to perceive it, and (iii) I can identify it through my knowing distinguishing facts about it.

As we shall see, Evans' account of how it is that we identify ourselves is a version of (i). That is Evans thinks that the way in which we distinguish ourselves from all other things, and thereby satisfy Russell's Principle with respect to ourselves (and are thereby able to entertain singular thoughts about ourselves) is, roughly speaking, by perceiving ourselves. In the next section, I will return to Evans' application of Russell's Principle to self-referential thoughts. For the time being it suffices to say that Evans' account clearly draws heavily upon that of Strawson, and has been highly influential, particularly in the work of Cassam.

In his *Self and World* Cassam puts forward what he calls 'the intuition version of the identity argument' as an attempt to show that a necessary condition of self-consciousness is that a subject must experience him or herself as a bodily object. Cassam begins by pointing out that, 'a self-conscious subject must be capable of thinking of her experiences as *her* experiences, that is, of self-ascribing them' (pp. 118–9). This is, he claims, 'at least one important element of what might be described as our intuitive notion of self-consciousness' (p. 119). Of course, this self-ascription cannot be merely accidental, but must be 'qua subject'. Cassam then claims that it, 'is a quite general truth that the ascription of experiences to an identical subject turns on the existence of some means of distinguishing the subject of such ascriptions as one object among others, because it is an even more general truth that a thinker will not count as having latched on to a particular item in the world and predicated something of it unless she has a capacity to distinguish the object of her judgement from all other things' (pp. 122–3). One way of distinguishing oneself from all other things involves knowing 'what *sort* of thing one has referred to' (p. 123). But this kind of knowledge does not appear to be a necessary condition of self-consciousness, since we can consider, 'the Cartesian dualist who regards the persisting subject of her thoughts as an immaterial substance. This belief may well be philosophically indefensible … but this surely has no bearing on her ability to think first-personally' (p. 127). That is, both dualists and materialists are self-conscious, even though at least one of them does not appear to know what sort of thing he or she is. So there must be a way of distinguishing oneself from all

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

other things that does not presuppose that one know what sort of thing one is. The answer is that, 'awareness of the object [oneself] as a shaped, located, and solid 'articulated unity' is what puts S in a position to 'isolate' it and predicate something of it' (pp. 136–7). So, 'The dualist satisfies a substantive 'knowing which' requirement on self-reference because and only because she is intuitively aware of that to which she ascribes her experiences as an articulated physical unity' (p. 140). Thus, 'Awareness of the subject of one's experiences as something with a determinate shape, as well as solidity and location, is a transcendental condition of consciousness of self-identity because it is in being aware of one's spatial properties that one satisfies the discrimination requirement on self-reference' (p. 142).

I have sketched Cassam's argument at some length as it is the clearest example yet of the Strawsonian strategy. Cassam takes self-consciousness to involve self-reference, applies Russell's Principle to self-referential thoughts and then derives a substantial necessary condition. We can see, then, that this strategy is a common thread running from Strawson, through Evans, to Cassam. The strategy relies on two claims, first, that self-consciousness requires self-reference and, second, that self-reference must adhere to Russell's Principle. Self-consciousness requires me to know which thing I am, in the sense of being able to distinguish myself from all other things.

II

What, then, is self-consciousness? Not a question to be answered lightly. There are, broadly speaking, two styles of thought to be considered. The first takes self-consciousness to be a special form of awareness of the self. Self-consciousness is, on this view, something like an inner perception of the self. Traditionally this has been thought to involve an awareness of the self's mental properties only, however recently philosophers in the Strawsonian tradition have suggested that bodily-awareness is a form of self-consciousness. That is, there is an inner perception of the self as the possessor of non-mental, bodily, properties. These views agree that self-consciousness is a form of *awareness*[8].

A second way of thinking about self-consciousness is as a property of certain thoughts. The question now becomes, what does

[8] This sort of self-consciousness is what Hume famously denied that he had. See D. Hume, *A Treatise of Human Nature*, edited by L. A. Selby-Bigge (Oxford: Oxford University Press, 1978).

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

it mean to say that a thought is self-conscious? Well, to begin with, self-conscious thoughts have a certain functional role. Self-conscious thoughts immediately dispose one to act in various ways[9]. Thoughts that are not self-conscious do not. Thus, thinking 'I am about to be attacked by a tiger' will dispose me to run for my life, whilst thinking 'Smith is about to be attacked by a tiger' may not, and this is true even though I am Smith. Thus, the first-personal thought, the 'I'-thought, is self-conscious whilst the third-personal thought, the 'Smith'-thought, is not. In addition to output, the functional role of self-conscious thoughts is distinctive on the input side as well. Thus, a subject's self-conscious thoughts are liable to be affected in an immediate way by the deliverances of certain specific ways of gaining knowledge of one's mental and bodily situation[10]. Thus, each of us has a peculiar way of gaining authoritative and immediate knowledge of our own mental life. Let us call this the faculty of introspection. Self-conscious thoughts will immediately accommodate themselves to the deliverances of introspection. If introspection informs me of the occurrence of a desire for ice-cream, I will be disposed to think that I want some ice-cream. However, unless I also believe that I am Smith, I will not be similarly disposed to think that Smith wants some ice-cream, even if as a matter of fact I am Smith. It is a distinctive feature of Evans' account of the functional role of self-conscious thoughts that ways of gaining knowledge of one's bodily situation also feed into one's self-conscious thinking in this way. For example, the deliverances of bodily-awareness will immediately dispose me to think various self-conscious thoughts about my bodily orientation[11].

As should be clear from the above, this second conception of self-consciousness is closely associated with the first-person pronoun. Self-conscious thoughts are most naturally expressed using the word 'I'. As such, we can distinguish the two conceptions of self-

[9] This aspect of the functional role of self-conscious thoughts is well articulated in J. Perry, 'The Problem of the Essential Indexical', reprinted in *The Problem of the Essential Indexical and Other Essay*, expanded edition. (Stanford: CLSI Publications, 2000).

[10] This aspect is made clear in G. Evans, *The Varieties of Reference*, edited by J. McDowell (Oxford: Clarendon Press, 1982).

[11] I do not wish to give the impression that, according to this second conception, self-consciousness can be defined exclusively in terms of the functional role of self-conscious thought. It is an open question whether there is an intrinsic essence to self-conscious thought that cannot be defined functionally. Rather than provide a definition, my comments on the functional role of self-conscious thoughts serve only to locate the phenomenon we are interested in.

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

consciousness by calling the first 'self-awareness' and the second 'first-person thought'. Now, whilst I have distinguished between these two conceptions of self-consciousness, it is clear that they are intimately related. For, on the one hand it would seem natural for the proponent of self-awareness to assert that the capacity for first-person thought is dependent on the internal awareness of the self; that self-awareness is the source of self-conscious thought. This view is well expressed in the argument from Cassam outlined in the previous section. His argument proceeds from self-consciousness, construed as first-person thought, to self-consciousness, construed as bodily self-awareness. The latter, he believes, is a necessary condition of the former[12]. Cassam's approach here is typical of the Strawsonian strategy in general. The conception of self-consciousness that is employed is that of first-person thought. It is first-person thought that the Strawsonian strategy claims necessarily involves self-reference, and it is these self-referring first-personal thoughts to which Russell's Principle is applied. It is, therefore, with first-person thought that we shall be concerned.

What, then, is the relationship between first-person thought and self-reference? The proponent of the Strawsonian strategy is committed to there being a conceptual connection between first-person thought and self-reference. This could take one of two forms. First, it might be held that each and every first-person thought is self-referential. One would, for example, think this if one accepted the following account of how the referent of first-person thoughts is fixed: first-person thoughts refer to their thinker[13]. This rule would clearly justify the first step in the Strawsonian strategy; that the capacity for self-consciousness requires the capacity for self-reference. Second, it might be held that whilst the most central cases of first-personal thought are self-referential, there are atypical examples of first-personal thoughts that are not self-referential[14].

[12] This is also the general strategy of the rather different position defended in J. L. Bermúdez, *The Paradox of Self-Consciousness* (Cambridge, Mass.: MIT Press, 1998).

[13] This self-reference rule is the analogue in thought of what D. Kaplan, 'Demonstratives' in J. Almog, J. Perry and H. Wettstein (eds.), *Themes From Kaplan*. (Oxford: Oxford University Press, 1989), takes to be the character of the word 'I'. There may be ways, other than acceptance of this rule, to support the view that each and every first-person thought is self-referential, but the self-reference rule is the most obvious.

[14] For different motivations for holding such a view see D. Velleman, 'Self to Self', *The Philosophical Review* 105 (1996), and E. Corazza, W. Fish and J. Gorvett, 'Who is I?', *Philosophical Studies* 107 (2002).

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

What do I mean here by 'central cases'? Well, I take it that the self-ascription of a mental predicate will constitute a central case. Thinking that one wants an ice-cream is a central case of first-person thought. Let us suppose that, of necessity, self-consciousness involves such central cases. It will follow that self-consciousness necessarily involves self-reference. Again the first step of the Strawsonian strategy will be justified[15].

The second step of the Strawsonian strategy involves the application of Russell's Principle to these self-referential first-person thoughts. If referring to an object requires that one know which thing it is, then referring to oneself requires knowing which thing one is. Further, if knowing which thing the object of one's thought is requires that one be able to discriminate it from all other things, then knowing which thing one is requires that one be able to distinguish oneself from all other things. Our question, then, is this: how is it that first-person thinkers are able to distinguish themselves from all other things? In the rest of this section I shall be arguing for the following claim: an adequate answer to this question has the consequence of undermining the tight conceptual connection between self-consciousness and self-reference. That is, if one applies Russell's Principle to first-person thought, one opens up the possibility of a self-conscious subject who is not capable of self-referring. The second step of the Strawsonian strategy undermines the first.

In general, how might one identify (in the sense of 'distinguish from all other things') an object of thought? According to the picture presented to us by Evans, there are three ways in which a subject can identify an object: (i) by perceiving it, (ii) by having the ability to recognize it, and (iii) by knowing distinguishing facts about it. Let me begin by discarding the recognitional model as an account of the self-identification required for first-person thought. The recognitional mode rides on the coattails of the perceptual mode. Our way of identifying a certain class of thing cannot *primarily* be through an ability to recognize it. Being able to distinguish oneself from all other things is not then, or at least not quite generally, a matter of being able to recognize oneself. The difficulty comes when deciding between the other two candidates: the perceptual model and the description-based model. On the first,

[15] I have skated over deep waters here, for it is not at all clear how the functional role of first-person thought relates to the way in which the reference of such thoughts is accomplished. For a sense of the difficulty involved in this, see J. Campbell, *Past, Space, and Self* (Cambridge, Mass.: MIT Press, 1994), Ch. 4.

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

one's capacity to distinguish oneself from all other things is a matter of one's being in perceptual, or quasi-perceptual, contact with oneself. This is what accounts for one's ability to think first-personally. On the second, that capacity is not dependent on any perceptual contact, but simply requires one to know a uniquely individuating description of oneself.

As a matter of fact, Strawson, Evans, and Cassam all opt for the first of these. Strawson, for example, speaks of a 'determinate intuition' of the subject. The reasons for rejecting the description-based account of self-identification are familiar. The first is the claim that for any description 'the *F*' it is always possible to doubt the identity 'I am the *F*'. Hence, learning that the *F* is such and such, is not equivalent to learning that I am such and such[16]. In Cassam's terminology, introduced earlier, even if one is in possession of a uniquely individuating description of oneself, this will not account for the special 'qua subject' character of first-person thought. As a response to this, it might be suggested that it is not possible to doubt the identity, 'I am the thinker of this thought', or, 'I am the subject of these conscious states'[17]. If so, it might be argued, one could satisfy Russell's Principle with regards to oneself by knowing that one is the thinker of one's current thoughts. But there is a well known answer to this line of thought. To begin with, it is not absolutely clear that we can demonstratively refer to our thoughts *at all*. But, even if we can, there is an influential line of thought, due to Strawson, to the effect that since the identity of a mental state is determined by the identity of its owner, the demonstrative identification of a thought is parasitic on the demonstrative identification of the person whose thought it is[18]. If this is correct, the proposed account would be circular.

[16] See J. Perry, 'The Problem of the Essential Indexical', reprinted in *The Problem of the Essential Indexical and Other Essay*, expanded Edition. (Stanford: CLSI Publications, 2000), and H-N Castañeda, 'He: A Study in the Logic of Self-Consciousness', reprinted in J. G. Hart and T. Kapitan *The Phenomeno-Logic of the I: Essays on Self-Consciousness* (Bloomington and Indianapolis: Indiana University Press, 1999).

[17] See, for example, C. Peacocke, *Sense and Content: Experience, Thought and their Relations* (Oxford: Clarendon Press, 1983), Ch. 6. I should point out that Peacocke's account is not an attempt to satisfy Russell's Principle. Also see, B. Russell, 'On The Nature of Acquaintance', reprinted in *Logic and Knowledge* (London: Unwin, 1956), p. 165.

[18] '*particular* states of consciousness … cannot be thus identifyingly referred to except as the states or experiences *of* some identified *person*. States, or experiences, one might say, *owe* their identity as particulars to the

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

I shall simply assume that the description-based account of the satisfaction of Russell's Principle with regards to first-person thought is unsatisfactory. In any case, the fact that the main proponents of the Strawsonian strategy reject the description-based account, gives enough of a justification for pursuing the alternative line. I shall be concerned, therefore, with the perceptual model. The idea is this: self-consciousness necessarily involves self-reference, self-reference requires that one be able to distinguish oneself from all other things, and being able to do this is a matter of being in perceptual, or quasi-perceptual contact, with oneself[19]. Since this perceptual mode of identification characteristically gives rise to demonstrative thoughts, this amounts to treating first-personal thoughts on the demonstrative model. The result is that, as is the case with other demonstrative thoughts, first-personal thoughts are opened up to the possibility of reference failure.

On Evans' view, demonstrative thoughts require two components: information links and modes of identification. That is, not only must one be able to distinguish the object of thought from all other things, one must also be receiving information from the object. There are, therefore, two ways in which a demonstrative thought can be 'ill-grounded'. A demonstrative thought can be ill-grounded if either one's purported mode of identification fails to distinguish the object of thought from all other things, or if one is not receiving information from the relevant object. As an example of the first sort of ill-groundedness, consider a case in which the subject is receiving perceptual information from a table in front of them but, due to considerable distortion in the channel through which this (mis)information is flowing, neither the table, nor any other object, is picked out by the subject's mode of identification. As an example of the second kind of ill-groundedness, consider a case in which the subject is suffering from a hallucination of a table in a place where, as a matter of fact, there is just such a table. Thus, the subject's (mis)information derives from nowhere, even though the mode of identification does serve to uniquely identify an object[20].

[19] For a good discussion of the perceptual-demonstrative model, see J. Campbell, *Past, Space, and Self* (Cambridge, Mass.: MIT Press, 1994), Chs. 3 and 4.

[20] For the full range of cases of ill-groundedness, see G. Evans, *The Varieties of Reference*, edited by J. McDowell (Oxford: Clarendon Press, 1982), p. 133.

identity of the person whose states or experiences they are.' P. F. Strawson, *Individuals: An Essay in Descriptive Metaphysics* (London: Routledge, 1959), p. 97. Also see G. Evans, *The Varieties of Reference*, edited by J. McDowell (Oxford: Clarendon Press, 1982), p. 253.

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

Depending on one's theoretical commitments, there are broadly speaking two stances that one might take with regards to such cases of ill-groundedness. One might say that in these cases the subject thinks a thought that fails to refer, or one might say that in these cases one fails to think a thought at all, and so suffers from an 'illusion of thought'. Evans' view is, of course, the latter. According to this picture, when one attempts to think a demonstrative thought, but that thought is ill-grounded, one fails to think anything. But, since it is the case that, from the subject's perspective, the situation is indiscriminable from that in which a real thought is being thought, the subject is under the illusion that he or she is thinking a thought. The alternative view has it that whilst there is no illusion of thought, the genuine thought lacks an object.

Treating first-personal thoughts on the perceptual-demonstrative model opens the way for a similar situation with regards to them. First we need to consider the kind of perceptual, or quasi-perceptual, contact that one generally has with regards to oneself. To begin with, through bodily-awareness one is in receipt of information about one's location and orientation in space. We can also count that aspect of external perception that serves to locate the subject in relation other perceived objects. Third, through introspection, one is in receipt of information concerning one's current mental situation. Finally, memory provides one with information concerning one's history. All these ways of gaining information with regards to one's mental and bodily situation are to be regarded as the information links in the case of first-person thought[21]. It is when this contact goes awry that reference failure can ensue. Consider the following cases: (1) Suppose that it were possible to remove a person's brain whilst keeping it alive and retaining all the relevant links for perception and the control of action, maybe by radio transmitter. We could call the separate brain the body's 'control centre'. Now, imagine two qualitatively indistinguishable 'control centres' outside of but linked to a single body, such that each appears to perceive the world 'through' that body[22]. Each control centre is a centre of consciousness, and each receives information from the single body, so each has the requisite information link. However, each control centre's mode of identification is faulty, since neither one can

[21] Of course, it is highly controversial whether all or any of these are genuinely perceptual.

[22] See, G. Evans, *The Varieties of Reference*, edited by J. McDowell (Oxford: Clarendon Press, 1982), pp. 254–5. For a dramatic account of such a fantasy, see D. Dennett, 'Where Am I?', in his *Brainstorms: Philosophical Essays on Mind and Psychology* (London: Penguin, 1978).

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

distinguish between the two centres. That is, each subject is incapable of distinguishing itself from the other, and so from all other things. In this case, the first-personal thoughts of each will be ill-grounded. (2) A single 'control centre' receiving illusory information that derives from no body. Even if the subject's mode of self-identification accurately picks out exactly one body, located somewhere in the world, the fact that there is no information link to that body will result in the ill-groundedness of first-person thought. This is for the reason that if the reference of first-person thoughts is determined (partly) by information links, when those links are deviant, reference will fail[23].

These cases are analogous to possible cases of ill-groundedness with regards to ordinary demonstrative thoughts[24]. Again, with regards to first-person ill-groundedness there are two possible ways of describing the situation. On the one hand it might be claimed that these subjects do manage to think first-personally, but those first-personal thoughts that they think fail to self-refer. On the other, it might be claimed that these subjects fail to think first-personally, they merely suffer from the illusion of first-personal thought. Suppose, for the moment, that one accepts the first description of such cases. It now seems that the connection between first-person thought and self-reference has been severed. For, in giving such cases, we have described situations in which a subject has the capacity for first-personal thought (i.e. is self-conscious), yet lacks the capacity for self-reference. That is, in applying

[23] Of course, this second sort of ill-groundedness results from treating first-person thought as requiring an information link, rather than applying Russell's Principle to it. One could, in principle, drop the information-link requirement whilst retaining Russell's Principle. But the very fact that Russell's Principle leads to the possibility of reference failure is all that is required for the point I am making.

[24] Another candidate for an ill-grounded first-person thought is Anscombe's example of a subject in a state of complete sensory deprivation thinking, 'I won't let this happen again'. See, E. Anscombe, 'The First Person', in S. Guttenplan, *Mind and Language: Wolfson College Lectures 1974* (Oxford: Clarendon Press, 1975). My reason for not relying on this example is that Evans takes himself to be able to rely on a dispositional account of the relevant information links, and maintain that self-reference can still occur in such a situation. See G. Evans, *The Varieties of Reference*, edited by J. McDowell (Oxford: Clarendon Press, 1982), pp. 215–6. O'Brien takes Evans to task for this claim. See L. O'Brien, 'Evans on Self-Identification', *Noûs* 29 (1995). For another pertinent discussion of Anscombe's example, see Q. Cassam, *Self and World* (Oxford: Oxford University Press, 1997), pp. 147–50.

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

Russell's Principle (on its perceptual-demonstrative model) to first-person thought, we have opened up the possibility of a subject who is self-conscious but who cannot self-refer. But this spells the end of the Strawsonian strategy, which requires both that self-consciousness entails the capacity for self-reference, and that Russell's Principle be applicable to self-referential thought.

Then maybe the proponent of the Strawsonian strategy should, following Evans, accept the second description of these cases, and endorse the possibility of the illusion of first-personal thought. In this way, in might be thought, we can avoid the problem, as it will turn out that any case in which a subject fails to satisfy the conditions on self-reference is a case in which that subject fails to think first-personally and, given that we defined self-consciousness as first-person thought, this means that such a subject is not self-conscious. Taking this option, allows us to retain both the view that self-consciousness (conceived as first-person thought) necessarily involves self-reference, and the view that first-person thought is subject to the perceptual-demonstrative model of Russell's Principle.

But this move is unconvincing. On this picture, self-consciousness is a matter of first-person thought and first-person thought is potentially illusory. This means that self-consciousness is potentially illusory. Two subjects may be phenomenologically indiscriminable and yet one be self-conscious and the other not. This has the consequence that I would be able to question whether or not I am self-conscious, since it is possible that I am not (if, for example, I am a brain-in-a-vat). This, I propose, is deeply implausible. The idea of a subject truly *thinking* to itself *I am not self-conscious*, is barely coherent[25]. An account of self-consciousness that leaves it as an epistemic possibility that I am not self-conscious is defective.

Furthermore, given the dialectic of the Strawsonian strategy, this position is troubling. The idea of spelling out the necessary conditions of self-consciousness is an idea that shares a great deal with traditional transcendental arguments. But, as a rule, transcendental arguments take as their starting points aspects of experience that are beyond rational doubt. One of the reasons why it is philosophically interesting to give the necessary conditions of self-consciousness is that self-consciousness is a bedrock phenomenon, something that we know with certainty that we possess. The view currently under consideration undermines this, and for that reason we should be

[25] I say *thinking* since by hypothesis the subject cannot think genuine first-personal thoughts.

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

suspicious of it. Defining self-consciousness in such a way that it is not certain that we have it deprives the project of giving the necessary conditions of self-consciousness of much of its interest.

I suggest that, if we are to endorse the possibility of the illusion of first-person thought, we should define self-consciousness in such a way as to accommodate this fact. We should define self-consciousness phenomenologically. On this picture self-conscious subjects are those for whom it seems that they can think first-personally. Self-consciousness is certain, beyond sceptical doubt. This allows that brains-in-vats are self-conscious even if we want to say of them that their first-personal thinking is illusory.

The upshot of this is, again, that the second step of the Strawsonian strategy undermines the first. The application of Russell's Principle to first-personal thought, when held together with the rejection of the description-based account of self-identification, is in direct conflict with the view that self-consciousness requires the capacity for self-reference. This is damaging both to the view that first-person thoughts are, without exception, self-referential and also to the view that the central cases of first-person thought (the self-ascription of mental predicates) are self-referential. The situations that we have described involve self-conscious subjects that are incapable of self-reference. It is the falsity of this that the first step of the Strawsonian strategy requries.

To illustrate, let us return to Cassam's 'intuition version of the identity argument'. Cassam's argument explicitly relies on the following three claims: (A) A self-conscious subject must be able to self-ascribe his or her experiences. (B) The self-ascription of experiences requires that one know which thing one is, in the sense that one can distinguish oneself from all other things. (C) The way in which one satisfies Russell's Principle with regards to oneself is by being aware of oneself as an 'articulated physical unity'[26]. We can read the first claim as the assertion that self-consciousness requires the capacity for first-person thought. The second claim is the application of Russell's Principle to first-person thought. The third is the acceptance of the perceptual-demonstrative account of knowing which thing one is. I have been arguing that one cannot hold these three theses in conjunction.

How might the proponent of the Strawsonian strategy respond to these charges? One way might be to reject the perceptual-demonstrative model of first-person thought, and revert to the description-based account. That is, it could be maintained that the way in

[26] Q. Cassam, *Self and World* (Oxford: Oxford University Press, 1997), pp. 118–42.

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

which we satisfy Russell's Principle with regards to our first-personal thinking is by knowing a proposition such as, 'I am the thinker of these conscious states'. As I have already pointed out, defenders of the Strawsonian strategy have, possibly for good reason, not been keen to take this route. Another option would be to attempt to make the perceptual-demonstrative model immune from reference failure. But it is difficult to see how this might be achieved. If the way in which a subject's first-person thoughts satisfy Russell's Principle is by their being in perceptual, or quasi-perceptual, contact with themselves, being in such contact is a necessary condition of first-person thought. But, as we have seen, it is relatively simple to describe cases in which this condition does not obtain; cases in which the contact that one has with oneself is sufficiently distorted to make one's first-person thoughts ill-grounded.

As a third option, it might be suggested that whilst Russell's Principle should be dropped as a principle governing reference, it can still be employed in an account of what it is to think of oneself as an object. So, whilst we can reject the line of thought that led from Russell's Principle to the possibility of reference failure for first-person thought, we can still endorse the thought, which occurs in Cassam's argument, that self-consciousness requires that one be able to distinguish oneself from all other things. To clarify, it might be maintained that although the reference of first-person thought is not effected via Russell's Principle, but by the self-reference rule, still Russell's Principle must apply to these, self-referential, first-person thoughts. The problem with this suggestion is that it is hard to see how such a combination of views is consistent. If first-person thought guarantees self-reference in the absence of any epistemic relation to oneself, how can it be at the same time maintained that self-consciousness is thinking of myself as an object and that *that* requires the satisfaction of Russell's Principle? If it is insisted that first-person thought requires knowing which thing one is, and that this knowledge is to be thought of on the perceptual-demonstrative model, whether or not this epistemic relation is thought to be determinative of the referent of first-person thoughts or merely a condition on them, then we cannot guarantee that self-consciousness requires self-reference.

Either self-consciousness requires me to know which thing I am or it does not. If it does then self-consciousness does not require self-reference, but if it does not then Russell's Principle cannot be used in an argument for the necessary conditions of self-consciousness. As Evans himself puts it, 'our ordinary thoughts about ourselves are liable to many different kinds of failings, and … the

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press

Cartesian assumption that such thoughts are always guaranteed to have an object cannot be sustained.'[27] But this means that Russell's Principle is not going to yield necessary conditions of self-consciousness.

III

What I have been calling the Strawsonian strategy has enjoyed a great deal of popularity. The thought that it is possible to derive interesting necessary conditions of self-consciousness via the notion of self-reference is appealing to those who are sympathetic to a broadly neo-Kantian programme in the philosophy of mind. But there is a serious obstacle to be overcome. The proponent of the Strawsonian strategy must make a choice: either accept Russell's Principle, but in so doing undermine the strategy, or reject the principle, but in doing so deprive the strategy of its greatest weapon. Either way the Strawsonian strategy is in trouble. In my view, it is the second of these two routes that should be taken. By severing the connection between self-consciousness and self-reference, the acceptance of Russell's Principle does irrevocable damage to the Strawsonian strategy. The rejection of Russell's Principle, on the other hand, leaves open a window of opportunity[28]. For the rejection of Russell's Principle is only the rejection of a particular understanding of the requirement that one know which thing one is. It is open for the proponent of the Strawsonian strategy to offer up a weaker interpretation of this knowledge. The task, for those who find that neo-Kantian programme attractive, is to find a non-trivial epistemic condition on self-reference that does not sever the link between self-consciousness and self-reference. Formulating such a principle should be the task of those who wish to see the Strawsonian strategy progress.

University of Essex

[27] G. Evans, *The Varieties of Reference*, edited by J. McDowell (Oxford: Clarendon Press, 1982), p. 249.

[28] There is a strong case to be made that Russell's Principle is quite generally false, see R. M. Sainsbury, 'Critical Notice of *The Varieties of Reference*', *Mind* 94 (1985), and M. Rozemond, 'Evans on *De Re* Thought', *Philosophia* 22 (1992–1993).

https://doi.org/10.1017/S0031819104000464 Published online by Cambridge University Press