Self-Consciousness and the Double Immunity

ANDREA CHRISTOFIDOU

> Not long ago after a trying railway journey by night, and much fatigued, I got into an omnibus, just as another gentleman appeared at the other end. 'What shabby pedagogue is that, that has just entered?' thought I. It was myself; opposite me hung a large mirror.
>
> Ernst Mach—*Analysis of Sensations*

Concern with the metaphysics of self-consciousness is, in particular, concern with the relation between two distinct features which constitute its essence: self-identification and self-ascription—on the one hand, 'I'-thoughts about oneself *as oneself*, as subject, and on the other, 'I'-thoughts about oneself *as being* in a certain way, that is, self-ascriptive thoughts not only of one's experiences and the content of one's experiences, but also of one's bodily states and properties. The latter is crucial, for the possibility then arises that in self-conscious thoughts more is involved in the total cognitive or self-ascriptive awareness than the mere content of one's experiences.

The central thesis of this paper pivots on identifying and clearly distinguishing between two *kinds* of immunity to error in first-person thoughts, and argues that, contrary to the dominant view, the distinction drawn between absolute and circumstantial, or non-absolute, immunity to error through misidentification relative to 'I' is not correct: *all* first-person statements have absolute immune to error through misidentification relative to 'I'. Such immunity does not allow degrees. One of the main sources of the way the incorrect distinction is forced upon our considerations in this area can be traced to a conflation in our discussions of the two essential features of self-consciousness. Discussions of self-ascription (and by implication of self-knowledge and self-awareness), are subsumed under discussions of self-identification, or conversely, discussions of self-identification are conflated with discussions of self-ascription. The point here is not that self-*consciousness* does not require self-knowledge, nor that self-*identification* does not require self-knowledge. Rather, the point is that self-consciousness involves more than self-knowledge, and that self-knowledge presupposes self-identification. Thus, the point is not to say that self-identification and self-ascrip-

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

tion are not intimately related, but to deny that we can understand their distinctive contribution to our conception of the self and self-consciousness if we do not first clearly distinguish between them. If there is a problem with first-person statements it is one that concerns self-ascription and not self-identification—self-ascription of both mental and physical predicates[1]. It is one that concerns a second kind of immunity which I shall call: immunity to error through *misascription*. To anticipate, I shall argue that the question of how each of the two essential features is indispensable to our understanding of the self and self-consciousness cannot be answered until we distinguish between the two kinds of immunity. Implicitly, the aim is to provide a framework within which the reductionist challenge as to the nature of the self, and as a corollary the no-reference view of 'I', on the one hand, and the idealist conception of the self, on the other, can be addressed.

Discussions of self-identification abound in the literature and are already familiar enough not to need an extensive survey here; nevertheless, it will be helpful to offer a brief discussion of such a notion by drawing on the ways in which the works of various philosophers have stimulated the debate. This I undertake in §2, where I also raise the question of whether self-identification is sufficient for our understanding of self-consciousness. A negative reply leads me to the central thesis of this paper in §3. In §4 I sharpen the discussion by focusing on Sydney Shoemaker's work, since it was his eloquent expression of the problem which set the debate in motion. I argue that the force of his initial characterization of what I call the *first* immunity (immunity to error through misidentification relative to 'I') is diminished by his distinction between absolute and circumstantial immunity. Such a modification to the initial characterization diminishes, in turn, our understanding of the self and self-consciousness. In §5 I sketch some of the implications of the *second* immunity, and before concluding, in §6 I pull together the various strands central to my thesis. But first, in §1, I outline the problem.

1. The Problem

In considering the rôle played by 'I' in first-person thoughts, the following seems clear:

(a) 'I' functions as a singular *referring* expression;
(b) 'I' *identifies* the subject (of the thought, or judgment).

[1] For convenience I shall use 'bodily' and 'physical' interchangeably.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

Shoemaker argues that although 'I' exhibits this distinctive dual rôle, and therefore needs special attention, it is governed by a simple rule:

(R) The reference of 'I' is to the speaker, or thinker.

Such a rule leaves no latitude for the speaker's intentions in the determination of its reference.[2] Reflecting on the simple self-reference rule, Shoemaker introduced a notion that has helped to a large extent to shape our understanding of the puzzles that surround the 'I', what he calls *immunity to error through misidentification relative to the first-person pronoun*. Gareth Evans, following Shoemaker, argues that in first-person *identification* the self-reference principle guarantees that the object of an 'I'-thought is its subject. Even if this is not an adequate answer to the question 'what am I?', in self-conscious thoughts we have direct, or unmediated self-reference—what Evans calls *identification-free* self-reference.[3] And the primary evidence, Evans argues, for the claim that first-person reference is identification-free is the phenomenon of immunity to error through misidentification. Identification-free self-reference serves to explain why 'I' is immune to error through misidentification.

Self-identification concerns first-person thoughts expressed by using the pronoun 'I'; it is not concerned with thoughts that a subject may have about himself which are expressed using a proper name, or a definite description, for it might be that he does not realise that they refer to him. For any name or definite description *D* that refers to me, I might not know that I am *D*, yet I can still refer to myself as *I*. Even an amnesiac, or someone deluded, if he understands the first-person pronoun, can still refer to and identify himself despite not *knowing who* he is, or having any other information about himself. The irreducibility of 'I' in self-identification is of a piece with the irreducibility of 'I' in self-reference; but the issue is not so much the importance of this feature in self-identification, nor is it simply a matter of reference. Rather, it is a matter of 'I'-thoughts that are immune to error through misidentification[4]. In the ordinary case of perceptual awareness of objects, identification and re-identification bring with them the possibility of

[2] Sydney Shoemaker, 'Self-reference and Self-awareness', *The Journal of Philosophy*, vol. 65, no. 19, 1968. References to reprint in his *Identity, Cause and Mind: Philosophical Papers* (Cambridge University Press, 1984), pp. 6–18.

[3] Gareth Evans, *The Varieties of Reference* (Oxford University Press, 1982).

[4] See my, *The Metaphysics of the Self: Self-Identification and Self-Ascription* (Ph.D. thesis; University of London, 1993).

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

misidentification. But this perceptual model is inapplicable to self-*identification*—there is no possibility of misidentification in the first person case, for there is no such thing as picking out oneself and then trying to identify or distinguish it from other selves or objects. The '*know who*' requirement is inappropriate in self-identification. The inappropriateness of the requirement has been seen as posing a serious objection to the thesis that 'I' is a referring and an identifying expression. This seems to me to be a mistake. The inappropriateness of the requirement in self-identification neither establishes the implausibility of imposing it, or a '*know which*' (knowing *which* entity one is), or a '*know what*' (knowing *what* kind of entity one is) requirement on *self-consciousness*, nor does it entail that if 'I' is a referring expression then the referent must be a disembodied being. Such conclusions are far too quick, they are in fact fallacious since no weighty conclusions concerning the self and self-consciousness can be drawn simply on the basis of self-identification—this is precisely the central point of my main thesis: self-identification is insufficient for an understanding of self-consciousness. The appropriateness of a '*know who*', or '*know which*', or '*know what*' requirement can only be considered in the context of a properly developed discussion and consideration of self-ascription (see §4 below). In self-identification one is presented (assuming that 'presented' in this context is not understood on the model of perception of ordinary objects) to oneself in a 'primitive way', that is, in a direct or unmediated way. The point about self-identification is a peculiarly strong form of identification-free first-person reference, but this strong peculiarity does not entail that the referent is a peculiar sort of entity. No weighty metaphysical conclusions can be derived from considerations about self-identification alone.

Evans begins his discussion by distinguishing between self-reference (which may be thought of as a *formal* feature of the guarantee against reference-failure of 'I') and self-identification (which, as I should put it, may be thought of as a *substantive* feature of the guarantee against identification-failure), and goes on to argue (and I agree with him) that self-reference and self-identification are inexorably linked, indeed, according to Evans they must be in our discussions of self-consciousness. Implicit in Evans's argument seems to be the thought that although the immunity to error through misidentification appears, *prima facie*, to be an epistemic phenomenon (it seems to be an immunity to a certain sort of error in judgment which would impose a '*know who*' or '*know which*' requirement), it is in fact caught up, as I should put it, in the metaphysics of self-consciousness, namely, in the part that concerns the guaran-

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

tee against failure of *identification* of the referent. Despite this, Evans seems to slip when he argues that 'at the same time the subject must *know which* object it is of which he thus has [...] knowledge'[5]. Such a demand has been seen to have startling consequences, namely, that 'I' may fail to refer, for if one were to suffer from amnesia one's self-identification would miss its referent. I shall come back to all this late. The upshot thus far is that, for both Shoemaker and Evans the self-reference principle is inexorably bound up with the immunity to error through misidentification. Put another way, although self-reference and self-identification are *distinct*, in the sense that the former might be thought of as belonging to semantics and the philosophy of language whereas the latter might be thought of as belonging to metaphysics, such a distinction is a distinction of reason and not a real distinction[6].

Given the importance of the inexorable relation between the self-reference principle and the immunity to error through misidentification, it is clear that rule (R) as it stands neither captures such a relation, nor brings out the dual rôle of 'I' with which I began. (R) could be refined thus: in self-reference 'I' never fails to refer to the thinker or speaker—it is immune to reference-failure; and in self-identification 'I' never fails to identify the thinker or speaker—it is immune to error through misidentification. Following Evans, we might say that in first-person thoughts the subject is identical with the referent (the 'object') of his thought or utterance. (R) can now be reformulated:

[5] Evans, op. cit., [1982], p. 212.

[6] Two things need to be made clear here. First, the claim I am making concerning the relation between self-reference and self-identification should not be thought of as either trying to derive metaphysical conclusions from linguistic premises concerning the nature of reference and the semantic value of 'I', or, conversely, trying to derive conclusions about the semantics and the nature of reference of indexicals such as 'I', from metaphysical premises (or assumptions) concerning the nature of things. The issues are *distinct* though they can be related given that the 'I' plays a crucial rôle in arguments and paradoxes concerning the nature of the subject or the self. Secondly, given the limitations of space of a single paper, I can only discuss very few of the issues that would otherwise deserve attention; in particular, I shall not consider questions concerning Perry's *rôles*, Kaplan's *character*, Castañeda's notion of *'He'*. My focus is primarily on Shoemaker's work on the immunity to error through mis*identification* relative to 'I'. (See: John Perry 'The Problem of the Essential Indexicals', *Nous*, 1979, vol. 13, pp. 3–21; David Kaplan 'Demonstratives', *Themes from Kaplan*, [ed.] J. Almong *et al*., 1977, Oxford: Oxford University Press, pp. 481–563; H-N. Castañeda '"He": A Study in the Logic of Self-consciousness' *Ratio*, 1966, vol. 8, pp. 130–157).

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

(R*) Necessarily 'I' refers to *and* identifies the thinker, or speaker.

The reference (and identification) of 'I' is its referent—the object, or rather the subject referred to and identified. My use therefore of the immunity to error through misidentification is to be understood as Shoemaker's (and Evans's) initial characterization and understanding of it—namely, as involving an inexorable relation between self-reference and self-identification.

Evans's idea of identification-free self-reference of 'I' seems, however, to present us with a dilemma: either first-person identification-free self-reference is direct or unmediated, thus constituting an exception to the Fregean conception of *sense* and reference—in which case it needs some special epistemological support; or if the (canonical) epistemology of first-person thoughts is to be relevant to their semantics, something more substantial is needed than what is given by the identification-free self-reference principle alone: it needs a Fregean *sense*. The first-person *mode of presentation* (the Fregean *sense*) is important, not only in the context of indexical expressions, but also in unasserted first-person thoughts. Since I have dealt with the dilemma in another paper arguing that direct or identification-free first-person reference can meet the Fregean requirement, I do not intend a further discussion here except to emphasise that a Fregean *sense* is needed, the grasping of which requires us to be sensitive to the truth-conditions of statements, and not merely as a rule that fixes the reference of the expression—indeed, as Evans argues, a Fregean theory of *sense* cannot come apart from a theory of reference.[7] But the point about self-reference and self-identification is that the truth-conditions of first-person statements are immune to error.

The importance about (R*) is thus: in order for 'I' to be immune to error both in the formal sense of self-reference, and in the substantive sense of self-*identification*, a condition upon its use is that that use be understood. (This was the reason for the qualification to my earlier claim concerning the amnesiac (or the deluded), that he should understand the first-person pronoun.) And that is how I

[7] See my, 'First Person: The Demand for Identification-free Self-reference', *The Journal of Philosophy*, vol. XCII, No. 4, 1995, pp. 223–234. See also Evans [1982], and John McDowell, 'Reductionism and the First Person', *Reading Parfit*, Jonathan Dancy (ed.) (Cambridge: Blackwell, 1997, pp. 230–50). The question of whether Frege thought that sense fixes reference, or whether this is the Neo-Fregeans' (or at least some of them) interpretation is something that requires textual evidence and hence not something that can be dealt with in this paper. Suffice it to say that Evans' work on this issue is not committed to such an interpretation and for present purposes I shall follow Evans.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

should understand by (R*) henceforth. Understanding the use of 'I' is inextricably linked to the way it both refers and identifies the subject of thought—the primary evidence of which is the immune to error through misidentification. (R*) enables me now to state, and argue for it later: (a) the first immunity does not allow degrees—*all* first-person statements have absolute immunity to error with regard to the *first* immunity; and (b) *all* first-person statements cannot be false in virtue of an error in the *identification component* of these statements—they cannot be false in virtue of an error in the identification of the referent of 'I'. These claims ground the central thesis of this paper.

2. Immunity to Error through Misidentification

Wittgenstein was perhaps the first to point out that certain first-person statements enjoy an immunity to error through misrecognition of a person—or, as Shoemaker puts it, are immune to error through misidentification relative to the first-person pronoun. Philosophers who find the referring rôle of 'I' problematic claim that, at least in certain first-person statements, 'I' is not a referring expression at all. Others try to deny that first-person statements are immune to errors of identification. One way of attempting this is to suggest that there are different uses of 'I'. The *locus classicus* of this approach is a passage from the *Blue Book*, that has been quoted to (near) death:

> There are two different cases in the use of the 'I' (or 'my') which I might call 'the use as object' and 'the use as subject'. Examples of the first kind of use are these 'My arm is broken', 'I have grown six inches', 'I have a bump on my forehead', 'The wind blows my hair about'. Examples of the second kind are: '*I* see so and so', '*I* hear so and so', '*I* try to lift my arm', '*I* think it will rain', '*I* have a toothache'. One can point to the difference between these two categories by saying: The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error [...]. On the other hand there is no question of recognizing a person when I say I have toothache. To ask 'are you sure it's *you* who have pains? would be nonsensical [...].[8]

Wittgenstein's distinction between the 'object-use' and the 'subject-use' of 'I' has led philosophers in one of two directions: either to the

[8] Ludwig Wittgenstein, *Blue and Brown Books* (Oxford: Basil Blackwell, 1978), pp. 66–7.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

idea that the whole tenor of his account is that 'I', at least in its 'as-subject' use, does not *identify* the subject, or to the idea that 'I' is a grammatical illusion, and hence there is nothing to which it refers. I think Wittgenstein's position is more elusive (see n. 17).

Philosophers (such as Strawson, Shoemaker, Evans, and others) who reject the no-reference view of 'I' argue against the sort of move that, for example, Elizabeth Anscombe makes from the premise that it is impossible to misidentify the entity to which 'I' refers, to the conclusion that there is no identification of such an entity—indeed that there is no such thing as the self.[9] For example, Strawson explains why 'I' 'can be used without criteria of subject-identity and yet refer to a subject', saying that, even in such a use, the links between criterionless *self-ascription* and the third-person criteria 'are not in practice severed'[10] (a number of questions arise out of Strawson's claim; see §§3 and 5 below). Shoemaker also argues:

> the identification of a presented object as oneself would have to go together with the possibility of misidentification, and it is precisely the absence of this possibility that characterizes the use of 'I' [...] this is one of the main sources of the mistaken opinion that one cannot be an object to oneself, which in turn is a source of the view that 'I' does not refer.[11]

He goes on to say that the source of this mistake is the tendency of the 'no-reference' theorist to think of self-awareness on the model of sense perception (for example, that one's sense perception of John explains one's knowledge that John has a beard).[12] But, he continues, since this model is of dubious coherence when applied to self-awareness, it follows that the 'no-reference' theorist's conclusion that there seems to be no self-awareness and that 'the word 'I' is a pseudo-subject' must also be of dubious coherence.

[9] See G. E. M. Anscombe, 'First Person', *Mind and Language*, S. Guttenplan (ed.) (Oxford: Clarendon Press, 1975), pp. 45–65.

[10] P. F. Strawson, *The Bounds of Sense* (London: Methuen, 1966), p. 165. See also his 'P. F. Strawson Replies' in *Philosophical Studies: Essays Presented to P.F. Strawson*, S. van Straaten (ed.) (Oxford: Clarendon Press, 1980), pp. 260–96.

[11] Shoemaker, op. cit., [1984], pp. 12–13.

[12] Descartes also, contrary to the Cartesians, rejected not only the perceptual model of self-awareness, but also the notion of introspection (or what he called inner sense) saying that it was subject to error and deception, just like the external senses, and therefore could not be trusted (see *The Philosophical Writings of Descartes*, vol. II, John Cottingham, Robert Stoothoff and Dugald Murdoch (eds), (Cambridge University Press, 1984), AT VII 77).

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

Evans, in similar vein, traces the mistake to the equivocation between thoughts that involve demonstrative identification and are identification-free (of which immunity to error through misidentification is a straightforward consequence—a paradigm case is that of 'I'-thoughts), and thoughts that involve an identification component (that is, thoughts which are identification-dependent):

> There is a danger of moving from the fact that there is no identification in the latter sense (that no criteria of recognition are brought to bear, and so forth) to the conclusion that there is no identification in the former sense.[13]

These examples serve to emphasise the centrality of the immunity to error through misidentification relative to 'I'. But although the immunity is undeniably central to any discussion of self-consciousness, a theory of self-identification is insufficient for an understanding of self-consciousness. For otherwise, first-person statements such as 'I am bleeding' or 'I have grown three inches' should also describe subjective facts (as they are usually called; but see §5 below). As Kant argued, self-consciousness involves a distinction between subjectivity and objectivity; such a distinction, however, does not seem necessary for self-identification. So, if facts about self-consciousness are thought to involve such a distinction, there must be something more to self-consciousness than self-identification—and that is *self-ascription* of both mental properties and bodily properties.

Hence, despite the truth of rule (R*), in considering the problems of the self we need to distinguish two important questions: how is it possible for 'I' to be immune to error in identifying the referent? And, how is it possible that there should be properties the self-ascription of which is immune to error through mis*ascription*? These two questions have at their hearts the two distinct but related essential features of self-consciousness with which I began: self-identification and self-ascription. However, it is only after we distinguish clearly between the two kinds of immunity in first-person thoughts that we can go on to ask how each of these essential features is significant in our understanding of the self and self-consciousness. To clarify and explain the two immunities would be: (a) to provide an answer to the two questions, and (b) to begin to understand something about what we are.

The claim I am making here is not that the first-person pronoun does not raise problems about the self. The point is that, if 'I' does raise such problems, these are not restricted to first-person judg-

[13] Evans, op. cit., [1982], p. 218.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

ments that involve the self-ascription of mental properties, as it is generally thought, but extend to the self-ascription of bodily properties. To put it another way: such problems are not restricted to certain sort of 'canonical' evidence that can plausibly be taken to have some fairly direct bearing on the semantics of first-person judgments (for example, my feeling pain, or my having a toothache) as it is usually thought, but rather, if 'I' does raise such problems they also extend to first-person judgments that are considered in the literature unlikely to be similarly relevant to the semantics of first-person judgments (for example, my seeing in a mirror a leg bleeding). As I have argued elsewhere, 'I' is essential and irreducible[14]; however, it remains true that we won't have a proper understanding of its distinctive rôle in, and its implications to self-consciousness unless we show how its immunity to error through misidentification is absolute. Despite the *prima facie* attractiveness of current work on the idea that one must be aware, or 'know from the inside', that one is a flesh-and-blood object in the spatio-temporal world, such a premise will not get us very far; it would certainly not get us away from the traditional empiricist position: no fixed and abiding subject, as Kant argues, can be derived from what is known from the inside. A conception of such a subject demands a metaphysical underpinning. Nor will it get us very far by insisting as Strawson does, that the immunity against misidentification is explained and guaranteed by the 'ordinary practice, well established among human beings, of reference to themselves and each other'[15], beings who are flesh-and-blood objects. We need an argument for, rather than an assumption of the immunity of 'I' in *all* first-person statements whatever the self-ascriptive component. If I succeed in my aim, I shall have provided a framework for a conception of a fully-fledged and abiding subject tracing a route through the spatio-temporal world.

3. Drawing the Distinction between Two Kinds of Immunity

The discussion so far has generally been concerned with the one kind of immunity to error in first-person statements. I shall argue that there are in fact *two* kinds of immunity. This is with a view to arguing that, contrary to Wittgenstein's claim, there are not two dif-

[14] See my, op. cit., [1995].

[15] P. F. Strawson, 'The First Person—And Other', *Self-Knowledge*, Quassim Cassam (ed.) (Oxford: Oxford Readings in Philosophy, 1994), pp. 210–15, p. 211.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

ferent uses of 'I', but only two different kinds of self-*ascriptive* statements, and therefore the first immunity, immunity to error through misidentification, covers all referring uses of 'I'. Like Shoemaker, I consider self-identification to be an essential feature of self-consciousness, but unlike Shoemaker I contend that the immunity to error relative to 'I' is absolute. To offer the self-ascription of mental predicates such as 'having toothache', or 'being in pain', as examples of the immunity to error through misidentification, is not only unilluminating but misleading. It is misleading because it conflates self-identification with self-ascription; as we shall see these examples are also subject to a second kind of immunity. And it is unilluminating because the examples cited are not paradigm cases of the first kind immunity to error relative of 'I'; they are only *one part* of a whole range of first-person statements which are immune to error relative to 'I'—a range that includes mental self-ascription, bodily self-ascription, and a combination of both. Self-identification and self-ascription are not only syntactically distinct, they are also semantically distinct; the former concerns the identification component of a first-person statement, the indexical 'I' and its cognates, the latter concerns the predication component of such a statement, which includes property-names or predicates. Self-ascription presupposes self-identification and as such it can be neither identified nor conflated with self-identification (see §4 below).

The second kind of immunity, immunity to error through *misascription*, concerns the predication component of first-person statements. This immunity occurs with first-person statements that involve the self-ascription of certain mental predicates. I say *certain,* since this second immunity might not occur in every case of mental self-ascription (such as 'seeing a bird in the garden', or 'seeing a leg bleeding'). However, in cases such as 'I am in pain', there is no question, not only that it is *I* who am in pain (a result of the first kind of immunity), but also that it is *pain that* I feel (a result of the second kind of immunity).[16] There are two points to be made at the outset. First, the second immunity should not be confused with old-style incorrigibility, or with first person authority (see §5 below). Secondly, and more importantly, the distinction between the two kinds of immunity does not correspond in any direct way

[16] There might be instances of uncertainty and hesitation concerning some of our mental states, but these cases need not be understood as expressions of error or of ignorance of our determinate mental states, but as reflections of the indeterminacy of those states at those times. But in the self-ascription of sensations such as pain, truth and truthfulness cannot come apart.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

to the traditional distinction between mental properties and bodily properties. As we shall see, given that self-ascription presupposes self-identification, in first-person thoughts, judgments, or statements there is no such thing as being subject to error through *misidentification* relative to 'I' whatever the self-ascriptive property, but only being subject to error through *misascription*. The first immunity is absolute whatever the self-ascriptive component might be. Only a conflation of the two components can impose on us the incorrect distinction of degrees in terms of the first immunity. There is no cogent reason to suppose that 'I' is sometimes immune to error absolutely (in its *as subject* use), and at other times is not (in its so-called *as object* use).[17] The point I am pressing can be made vivid by considering some of the dominant accounts found in this area.

Shoemaker[18] argues that certain self-ascriptive statements involving physical states have only 'circumstantial immunity to error' relative to 'I', not absolute. For example, if a speaker asserted 'I am facing a table' this statement would be immune to error through misidentification *relative to 'I'* only circumstantially and not absolutely—indeed, he argues, any immunity it has it is derivable from a statement involving mental self-ascription—in the sense that it would be known or believed by the speaker 'as a consequence of the proposition "I see a table in the centre of my field of vision"'.[19] But does this mean that the use of 'I' is circumstantially immune to error through misidentification, or that the identification of the *table* is so immune? Indeed, why should statements like 'I am facing a table' have only circumstantial immunity to error relative to 'I'? Why should the predication or self-ascrip-

[17] Wittgenstein did not consider the distinction between the two immunities, and thus drew his 'subject-use'–'object-use' distinction, mistakenly thinking that 'I' is immune to error through misidentification only in its use 'as subject'. Indeed, Peter Hacker remarks that Wittgenstein (at least in the 'Notes for Lectures') says that a statement such as 'I have a toothache' 'is a remark about the use of the word 'I', whoever uses it' (p. 309; Ludwig Wittgenstein, 'Wittgenstein's Notes for Lectures on "Private Experience" and "Sense Data"', R. Rhees (ed.), *Philosophical Review*, vol. LXXXVII, no. 3, 1968, pp. 271–320) and traces the whole problem that surrounds 'first-person present tense psychological sentences [...] to our not understanding the function of 'I' (P. M. S. Hacker, *Insight and Illusion* (revised 2nd ed.) Oxford: Oxford University Press, 1986), pp. 261–2.

[18] Shoemaker, op. cit., [1984]. I shall discuss Shoemaker's work in greater detail in §4 and shall therefore limit myself here to a few introductory remarks.

[19] Shoemaker, op. cit., [1984], p. 8.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

tion component of such statements be suddenly allowed to compromise the immunity of self-identification? In Shoemaker's initial characterisation of the immunity, there was no reference to a self-ascriptive or predicative component. As we shall see later, the force of Shoemaker's initial insight is diminished by the distinction he draws between absolute immunity and circumstantial immunity.

Strawson argues that immunity to error through misidentification applies to cases of both corporeal and mental self-ascription—as he puts it, to '*both* states of consciousness *and* corporeal characteristics'.[20] This is true and insightful. However, Strawson's explanation, as we saw in §2, for why 'I' 'can be used without criteria of subject-identity and yet refer to a subject', is that, even in such a use, the links between criterionless *self-ascription* and the third-person criteria 'are not in practice severed'. There are two points to be made here. First, a striking feature of Strawson's argument is that it conflates the impossibility of misidentifying the entity to which 'I' refers, what might be called criterionless self-*identification* of 'I', with criterionless self-*ascription* of states of consciousness. The reason that 'I' can be used without criteria of subject-identity is *not* because of criterionless *self-ascription*, but because of what Evans has called, identification-free self-reference. Secondly, the claim that the immunity to error through misidentification applies to cases of both corporeal self-ascription and mental self-ascription—to '*both* states of consciousness *and* corporeal characteristics'—must surely be argued for, especially in this area of philosophy. Such an account of the immunity does not seem to address the central problem. One cannot simply

> repeat the point that the immunity of 'I' from reference-failure (of either kind) in the thought, or speech, of any human user of it, whatever his condition, is guaranteed by the role of the expression in the ordinary practice, well established among human beings, of reference to themselves and each other[21] (see also §6 below).

Evans quite rightly expresses dissatisfaction with Wittgenstein's ambivalent treatment of the issue, saying that it was this treatment that gave rise to the widespread wrong belief "that the phenomenon of immunity to error through misidentification, which is so central to the notion of self-consciousness, does not extend to self-ascrip-

[20] P. F. Strawson, *Individuals* (London: Methuen, 1959), p. 104; see also Strawson, op. cit., [1966], pp. 164–5.

[21] Strawson, op. cit., [1994], p. 211.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

tions of physical properties"[22] (see also the discussion in §4 below). However, Evans, like Strawson, when he discusses immunity to error he subsumes *self-ascription* under the immunity to error through mis*identification*—such subsumption is evidenced in a number of ways in his work.

First, Evans argues that there is a class of first-person judgments, such as 'my legs are crossed', which are immune to error through misidentification on the grounds that 'when the first component expresses knowledge gained in the appropriate way' it does not make sense to say 'someone's legs are crossed, but is it my legs that are crossed?'. It does not make sense because:

> There just does not appear to be a gap between the subject's having information [...] in the appropriate way, that the property of being *F* is instantiated, and his having information [...] that *he* is F.[23]

This might be true. But if the reference of 'I' (or 'my') is identification-free as Evans had initially argued, then 'I' is immune to error through misidentification whatever the self-ascriptive property might be—whether it is mental self-ascription or bodily self-ascription. If there is a problem with such statements, the discussion must be directed to the self-ascriptive component and not to the self-identification component. The proviso 'knowledge gained in the appropriate way' is appropriately applicable to the self-ascriptive component (in this case to bodily self-ascription), but it is irrelevant to the immunity to error through misidentification relative to 'I' and its token reflexives. First-person judgments made on the basis of current proprioceptive and other bodily properties, are immune to error through misidentification—it could not be someone else's legs that are being referred to and identified by the token reflexive 'my', it could not be anyone else than me that 'my' identifies (though as Evans observes there can be causally deviant cases linking one's brain with someone else's legs in such a way that one experiences the legs as being crossed; however, such a scenario, he says, merely shows the possibility of error and not that there is no immunity in the normal way—but see

[22] Evans, op. cit., [1982], p. 218. It is worth noting that McDowell says that Evans's plan seems to have been to work up from the idea of 'how wrong it is to associate the immunity to error of misidentification of 'I'-thoughts exclusively with *mental* self-ascription [...] to a picture of fully self-conscious thought, perhaps partly by a consideration of the 'vertiginous' thought that the object of this thought is its subject' (see Evans, [1982], pp. 261–2, n. 6). It is evident that there is no indication in these remarks of Evans's recognition of a second kind of immunity.

[23] Evans, op. cit., [1982], p. 221.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

also §4 below). There is a disparity of course between knowing that one's legs are crossed and knowing that one's legs are bleeding, but that is, at least *prima facie*, because in the former case one's knowledge is gained on the basis of sensory awareness—the feeling or sensation of one's crossed legs; that's the reason for there not appearing to be a gap between the subject's knowing that the property of being *F* is instantiated, and his knowing that *he* is F. Statements that involve feelings or sensations, as I shall argue, fall in the class of statements that have *double* immunity—see §4 below.

Secondly, Evans, following Shoemaker, is committed to denying that 'I' is immune to error through misidentification absolutely; he argues that it is not just with physical self-ascription that one cannot claim absolute immunity to error relative to 'I',

> one cannot make this sort of absolute claim of immunity to error through misidentification for *mental* self-ascriptions either—at least not self-ascription [...] which includes 'I see so-and-so' and 'I hear so-and-so'.[24]

Evans identifies a class of mental self-ascriptions in which the immunity to error through misidentification relative to 'I' is non-absolute. Again, it seems to me that if there is a problem here it is not one that concerns non-absolute claims relative to the first immunity, as Evans thinks, nor is it one that concerns circumstantial immunity, as Shoemaker thinks. The self-identification component in first-person statements remains untouched and admits no distinction between absolute and circumstantial immunity, or between logical and *de facto* immunity to error through misidentification.[25] If there is a problem with first-person statements it is one that concerns the self-ascriptive component of such statements of both mental and physical predicates, one that is relative to the second and not the first immunity.

Colin McGinn, in line with the dominant accounts, discusses these problems in the following passage:

> judgment *types* expressed by 'I see a canary' and 'I am waving my arm' are not immune to error through misidentification, since there are possible tokens of these types in which just such an error is committed.[26]

[24] Evans, op. cit., [1982], p. 219.

[25] This is again a distinction that Shoemaker draws in another paper which is concerned in particular with bodily self-ascription, a topic I shall discuss below. See Sydney Shoemaker, 'Persons and their past', in his, op. cit., [1984], pp. 19–48.

[26] Colin McGinn, *The Subjective View* (Oxford University Press, 1993), p. 51.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

On the other hand, McGinn says judgments such as 'I *seem* to see a canary' or 'I am *trying* to wave my arm' are immune to error through misidentification, since such judgments are grounded in purely inner and subjective awareness of certain mental states. It is certainly true that these judgments *are* immune to error through misidentification, but surely that is not the point here. What *is* relevant is that these judgments are also immune to error through misascription. It is a mistake to associate the immunity to error through misidentification only with 'purely' mental predicates or only with *seeming* experiences, as I shall argue in §4 below.

The upshot of all this is that our concern with the nature of the self remains incomplete in the absence of a recognition of the two kinds of immunity.[27] In order to sharpen the discussion I shall focus on Shoemaker's approach to the problem.

4. The Shoemaker Approach

Shoemaker holds, as we have seen, that certain uses of 'I'—those in which 'I' occurs as subject—are absolutely immune to error through misidentification. This claim is important to his account for he also holds that whether 'I' is used to identify depends on whether misidentification is possible. The question is: what are the conditions for such a possibility? Following Wittgenstein, Shoemaker claims that in statements such as 'I am bleeding':

[27] None of above philosophers supports the 'no-reference' view of 'I', though Wittgenstein's position is the most elusive. It seems to me that even in the *Blue and Brown Books* he is not at all sure whether he wants to claim that 'I' in its 'as subject' use, that is, in its use in conjunction with the self-ascription of mental predicates, has no referent, or, that there is no need to identify such a referent. I am inclined to the latter interpretation. A similar ambivalence is found in *Philosophical Remarks,* Rush Rhees (ed.) (Oxford: Basil Blackwell, 1975) where he observes that the use of 'I' 'is one of the most misleading representational techniques in our language [...] particularly when it is used in representing immediate experience' (p. 88, §57). It is possible to argue that, in his later work, Wittgenstein came to see that the no-reference view of 'I' was inadequate for resolving the problem of immediate experience; he therefore turned his efforts to the thesis of avowals as non-assertoric. This seems to me to be an evasion of the complexities and difficulties involved in this area (indeed, it is implicit in what I am arguing that the problem of immediate experience has to do with the second immunity). Hacker notes that the 'distinction between the use of 'I' as subject and as object does not reappear in [Wittgenstein's] later writings' (Hacker, op. cit., [1986], p. 233, n. 13).

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

> it can happen that what I say is false even though I am giving expression to the knowledge that a certain person is bleeding.[28]

It might happen that in the mirror I see a leg bleeding, and mistakenly think that the person whose leg is bleeding is myself. It is cases such as this that lead Shoemaker into thinking that the immunity to error is not absolute, or more strongly, that there is a possibility of misidentification *relative to 'I'*. On his view, at least a necessary condition for the possibility of misidentification would be that such statements:

> are subject to error through misidentification relative to the first-person pronouns, where to say that a statement 'a is ' is subject to error through misidentification relative to the term 'a' means that the following is possible: the speaker knows some particular thing to be , but makes the mistake of asserting 'a is ' because, and only because, he mistakenly thinks that the thing he knows to be is what 'a' refers to. The statement 'I feel pain' is not subject to error through misidentification relative to 'I': it cannot happen that I am mistaken in saying 'I feel pain' because, although I do know of someone that feels pain, I am mistaken in thinking that person to be myself.[29]

Two questions arise. First, is it correct to say that 'I' is immune to error through misidentification only when used as subject? Secondly, given the truth of (R*), is it correct to say that whether 'I' involves an identification depends on whether misidentification is possible?

Shoemaker differs from Wittgenstein in drawing, not on two uses of 'I', but on two classes of 'I'-judgment (which, however, correspond to Wittgenstein's two uses), but this does not affect my worries about his position. In effect, Shoemaker holds that the proposition 'I am ' would be subject to error through misidentification relative to the term 'I' only if:

> There could be someone who knows that some person has a certain property, and who mistakenly uses 'I' to refer to that person.

There are two ways of analysing this statement. The first is Shoemaker's; on his analysis, consistent and irreproachable though it may seem, in order to account for the error involved in statements which are expressed by self-ascribing physical predicates, we have to accept the distinction between the subject-use and object-use of 'I' (see also §6). The second analysis is mine; on this analysis,

[28] Shoemaker, op. cit., [1984], p. 7.
[29] Shoemaker, op. cit., [1984], pp. 7–8.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

Wittgenstein's distinction can be avoided, and with it the danger of being led to the view that 'I' in its subject-use does not refer to anything, or that '"the self" is a piece of philosopher's nonsense consisting in a misunderstanding of the reflexive pronoun'.[30] According to my analysis, which draws instead on the distinction between the two kinds of immunity, statements such as 'I am bleeding' may be subject to error through *misascription*, relative to a physical predicate, but they are not subject to error through *misidentification* relative to 'I', contrary to what Wittgenstein and Shoemaker argue. Shoemaker endorses the idea that although

> the identification of a presented object as oneself would have to go together with the possibility of misidentification, [...] it is precisely the absence of this possibility that characterizes the use of 'I'.[31]

He also quite correctly concedes that we ascribe certain predicates to others and they ascribe them to us on the basis of observation of behaviour, but that there are certain other predicates, mental predicates such as 'being in pain', that we ascribe to ourselves without the need for any behavioural observation. He presents the self-ascription of the latter predicates as examples of immune to error through misidentification relative to 'I'. This is true at one end of the spectrum, that which involves the identification component of first-person statements, but at the other end of the spectrum, the ascriptive or predicative component of first-person statements, it amounts to a conflation of the two kinds of immunity. On closer examination we find that Shoemaker does not seem to recognize that the self-ascription of (some) mental predicates is also immune to error through *misascription* and hence statements such as 'I am in pain' have a *double* immunity.

It might be objected here that there are clear counter-examples to my central thesis that 'I' is immune to error in *all* first-person statements. Consider the following two examples. First, suppose that on getting into an omnibus I look at the mirror hanging at the opposite end of it, and on the basis of that visual experience I judge: I look so shabby today. That certainly is a referential use of 'I', but it could be mistaken through misidentification relative to 'I', or so the objection goes, for although I think I am seeing myself when I see the shabby-looking person reflected in the mirror, it is really someone else on the omnibus who happens to look like me. Secondly,

[30] Anthony Kenny, *The Self* (Marquette: Marquette University Press, 1988), p. 4.

[31] Shoemaker, op. cit., [1984], p. 12.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

suppose that in a mirror I see a leg bleeding and think that it is mine, but in fact it is another person's leg. According to the objection examples such as these, which abound in the literature, are clear cases of failure of the immunity to error through misidentification relative to 'I'.

Let us consider the cogency of two counter-examples. If in the mirror at the end of the omnibus I see a shabby-looking person and judge that I look shabby, but it turns out that such reasoning is wrong, the flaw can be traced to the enthymematic premise: 'I am the person who is looking shabby'. To argue that this is a case of misidentification is to argue that the enthymematic premise is false; those who want to argue that misidentification relative to 'I' is possible, presumably think that it makes perfect sense for me to ask: 'someone is looking shabby, but is it me?'. In such cases, it is reasonable for me to think that someone is looking shabby, but not be entirely certain that the shabby-looking person reflected in the mirror is me, or someone else. But how does it follow from this that this is a case of misidentification relative to 'I'—that this is a case, as the objection claims, of misidentifying myself using 'I'? There is an important difference between the following two statements: (a) ' I am thinking "someone is looking shabby, but is it me?"', and (b) 'I am thinking "I am looking shabby"'. The first is a case of misidentifying *the person* who is looking shabby; the second is a case of *misascribing* the predicate 'looking shabby' to myself. But neither is a case of misidentification relative to 'I'. The objection seems to be trading on an equivocation between the two different statements. It moves fallaciously from the first statement where the question 'someone is looking shabby, but is it me?' makes sense (only because misidentification of *the person* who is looking shabby is possible), to the second statement 'I am looking shabby' where it makes no sense to wonder whether the reference and identification is to me, and misidentification is impossible. The statement 'I am looking shabby' is false precisely because the reference to, and identification of 'I' is unshaken: it is immune to error through misidentification; the same cannot not be said about the definite description. 'I' cannot be equivalent to, or be substituted *salvo sensu* by any definite description (or, a proper name) and preserve the truth-value of a first-person statement.[32]

Similarly, if in the mirror I see a leg bleeding and think that it is mine but in fact it is another person's leg, then I am mistaken in thinking that my leg is bleeding—but again, this is not a case of erroneously misidentifying myself or *my* leg. It is a case of misiden-

[32] See my, op. cit., [1995], pp. 223–34.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

tifying *the leg* that is bleeding, or the person whose leg is bleeding. My realization that my leg is bleeding is never equivalent to my realization that some leg is bleeding, or that someone's leg is bleeding and wondering whether it is mine or not. On discovering that it is not my leg that is bleeding I should say: 'so my leg is not bleeding', and then I should try to explain how I could have made this mistake. My mistake is established by determining that neither of my legs was bleeding—by determining the ascriptive component and not the identification component of the statement, since there can be no doubt as to whether it is *my* leg that is being referred to and identified by the token reflexive 'my'. I have *misascribed* to myself a predicate, but the reference *and* identification of 'my' or 'I' is unshaken: 'I' (or 'my') is immune to error through misidentification absolutely, whatever the self-ascriptive predicate might be.

There is an asymmetry or a disparity between 'I' and definite descriptions, or any other term including other indexicals and demonstratives like 'this' and 'that' (the pronominal use of demonstratives can be ambiguous and need to be accompanied by an implicit or explicit gesture; the indexical 'I' or 'my' is never ambiguous), and proper names. Hence a mistake in the identification component can occur only when I am using any of those other expressions, but not when I am using 'I'. My realization that I myself *am* the person bleeding is never equivalent to my realization that someone satisfying any definite description is the person bleeding. We can clarify this by giving a Russellian analysis[33]:

x[((this leg is x's) & (anyone whose leg this is, is x)) & (x=me)]

On this analysis, the definite description 'the person whose leg this is' is not a singular term, so there need only be one reference to any person (no doubt there is also a reference to a leg).

In drawing the distinction between the two kinds of immunity in §3 I said that self-ascription presupposes self-identification, and as such can be neither identified nor conflated with self-identification. We can now see the centrality of this claim. Any self-ascriptive statement, whatever the self-ascriptive predicate might be, entails that it is I who is the subject of that statement, and hence identification-failure is impossible. The reason for this is that self-identification is *always* presupposed by any possible self-ascription, or by any self-ascriptive experience, whether it be self-ascription of inner experience, of outer experience, of bodily experience, or of bodily property. It is not possible to consider some *self*-ascriptive experi-

[33] I owe the suggestion for providing this analysis to a discussion with David Wiggins.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

ence as separable or separate from self-identification (of the self or subject) so that such a self-ascribed experience can be considered and decided upon independently of self-identification. What remains open for consideration is whether I satisfy or not the predicate self-ascribed, and so the question of whether or not a first-person statement is subject to error depends entirely on self-knowledge (on the method of finding out truths about oneself as *being* in a certain way, that is, self-ascriptive thoughts), and therefore on what predicates one ascribes to oneself. But now we are in the realm of the second immunity, the immunity to error through mis*ascription*; if it turns out that I do not after all satisfy the predicate then it would be a case of misascription and not of misidentification. The first immunity remains absolute and direct whatever the self-ascriptive predicate might be.

The direct identificational use of "I" or "my", as in for example, 'my leg', does not involve 'being presented to [my]self as an object'[34], in the way, say, that my car is presented to me as an object. Whatever claims about the world are presupposed in the fact of using 'I' (or 'my') are true, provided I use it correctly, that is, that I understand its use. The same cannot be said about the claims made by my use of 'that leg', or 'that person' concerning which I may be in error. This is not a purely linguistic matter; I am not simply trying to explain what we mean when we use 'I', but also what the metaphysical underpinning of such talk is—what makes the immunity of 'I' possible[35]. It is impossible for me to be using 'I' and 'my' correctly if I am using it to speak of, refer to, or identify another person, or another person's body[36]. The impossibility is not grounded in the speaker's or thinker's intentions, or in his psychological or epistemic condition, but in the fact that self-identification is direct—that rule (R*) that governs the use of 'I' is true.

In judging that 'I am F' I invariably know without any doubt that the 'I' refers to and identifies me and no one else (it should be obvious that I am not concerned with indirect speech here). If I also know without any doubt that I am *F*, this is because *'F'* stands for a certain kind of mental property—say, pain. Moreover, a judgment of the form 'I am F' preserves its first immunity through past- and

[34] Shoemaker, op. cit., [1984], p. 14.

[35] Any metaphysical underpinning to 'I' ought not to be controversial, even if defending such an underpinning would require a framework that might be considered not to be entirely devoid of metaphysical neutrality.

[36] Another way of expressing this thought is to say that 'I' cannot be competently used to make the sorts of mistakes with which Keith Donnellan illustrates his famous distinction between referential and attributive uses of definite descriptions.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

future-tense judgments. If I say 'I was F', but in fact I was *not-F*, then I am wrong in attributing *F* to myself—but not because I am misidentifying myself using 'I'. I cannot be wrong that it was *I* who was F, or not-F (or that it is *I* who will be F, or not-F)[37]. 'I'-thoughts do not form *dynamic* Fregean thoughts like other demonstrative thoughts—e.g. thinking of today as yesterday, as the day before yesterday, and so on. There is no shifting from 'I'-thoughts to 'you'-thoughts, or from 'my-body thoughts' to 'your-body thoughts' (as one thinks of oneself over time) because they "cannot be connected by expressing a single dynamic thought".[38] Given that 'I' preserves its immunity, whatever the ascriptive property (mental or bodily) through time, for there is no need for any *skill* or *care* not to lose track of the object of one's 'I'-thoughts as they recede into the past, or extend into the future, this suggests that there is a persisting subject of whom 'I'-judgments involve no possibility of misidentification through space and time.

If my arguments are correct, then two points follow. First, they demonstrate that the thinking *qua* thinking 'I' cannot be the only entity that secures guaranteed self-identification, but that a fully-fledged and abiding self also secures such immunity to error through misidentification. It is now evident, as I said in §2 above, that the appropriateness of a '*know who*', or '*know which*', or '*know what*' requirement, in short, self-knowledge, can only be considered after a properly developed discussion and consideration of self-ascription. This seems to be where Evans has slipped because of his conflation of self-identification with self-ascription (see §1 above). Thus, it is only after we have provided such a basis that we can go on to argue that when I self-ascribe different properties I have 'discriminating knowledge' of myself. Secondly, they provide a framework for knowing *what* I am—a ground for knowing to what kind of entity these properties are ascribed. This in turn involves an ability to individuate myself from other objects in the world—that is, to *know which* elements I include within my 'boundary'. In the case of

[37] McDowell thinks that Evans (op. cit., [1982]) 'seems to slip when he suggests [p. 237] that the point can be put in terms of "immunity to error through misidentification": i.e. the circumstance that a judgment's predication is not attached to its subject by way of a judgment of identity' (McDowell, op. cit., [1997], p. 248, n. 3). But it seems to me that the point is correctly put, and that if there is a problem it has to do with the 'judgment's predication'. Only if we subsume self-ascription under self-identification can we think that the point cannot be put in terms of the first immunity.

[38] John McDowell, 'De Re Senses', *Philosophical Quarterly*, vol. 34, no. 136, 1984, pp. 283–294; p. 285, n. 8. See also my, op. cit., [1995].

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

seeing someone else's leg in the mirror and thinking it is my leg, I have mistakenly included in the conception I have of myself an element that does not belong to it. Insofar as I might incorrectly draw a boundary around (that is, know the 'fundamental ground of difference'[39] specific to) myself, it is possible that I should misascribe a property to myself. But it is certainly not possible that I should misidentify myself, if I am using 'I' and its cognates.

5. Immunity to Error through Misascription

The primary concern in this paper was the drawing and defending of the *distinction* between the two kinds of immunity—it was not with the *implications* of the second kind of immunity[40]; in this sense the principal task can be seen as accomplished. However, in the remaining space, I should like to sketch out some of the implications of the second immunity. The introduction of the second immunity raises a number of interrelated but distinct problems, such as first-person authority, epistemic asymmetry, and immediacy, all of which are central to our understanding of the unity and substantiality of self-consciousness, and of the autonomy of 'I'. In §2 I said that if facts about self-consciousness involve, as Kant thought, a distinction between subjectivity and objectivity, then there must be something more to self-consciousness than self-identification. It has now transpired that the introduction of the second immunity has brought out clearly the centrality and significance of self-ascription to self-consciousness. My attempt to offer some elucidation of the implications of the second immunity splits into two parts. The first involves the distinction between subjectivity and objectivity; the second considers the relation between the second immunity and a set of problematic notions that its introduction raises.

Subjective and objective: Self-ascriptive judgments about states of consciousness seem different in various ways from ordinary ascriptive judgments about the rest of the world—the latter are usually distinguished by their commitment to objectivity. Some philosophers have argued that this disparity is captured in the difference in the nature and function of subjective predicates. Yet, subjective predicates seem neither in their grammatical, nor in their logical behaviour to exhibit any signs of being anything other than ordinary predicates, much less to betray any sign of not being subject to

[39] Evans, op. cit., [1982], p. 107.

[40] These are explored in full detail in my doctoral thesis (op. cit., n. 4), and in another paper under progress.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

truth or falsity. If self-ascriptive judgments about states of consciousness can be true or false in the ordinary sense, where might the difference between such judgments and non-perspectival judgments about the rest of the world lie?

We can begin by saying that all states of consciousness are subjective in the sense that they do not exist independently of the subject of experience—they are subjective in the sense that they are essentially connected with a particular point of view (an instance of a *type*). If this is a fact about conscious and self-conscious beings, then there is a sense in which our self-ascriptive judgments about states of consciousness may be perfectly objective—not in the sense that such states exist independently of the subject, which is absurd, but in the sense that such judgments meet the objectivity requirement of truth and falsity. That is, although self-ascriptive judgments are essentially perspectival, they provide an opportunity not only for assessment of right or wrong, but they also aspire to truth, since 'true' as used in self-ascriptive judgments has its familiar and ordinary sense that it has when it is used in judgments about the rest of the world. From this elucidation of 'subjective', contrary to what might be supposed, it does not follow that questions concerning self-ascriptive judgments about states of consciousness are not answerable to a standard that is founded both in our objective conceptual categories and in the states, properties, and responses of self-conscious subjects of experience; such questions would be 'answerable to what registers upon the critically-regulated [property-involving] responses of [subjects of experience]'.[41]

However, and herein lies the difference, unlike non-perspectival judgments about the rest of the world, self-ascriptive judgments do not aspire to objectivity absolutely, or *sub specie æternitatis*. Subjective predicates presuppose a special centrality in the world, a special station in the world, because of their essential connection with the first-person point of view. It goes without saying that an absolute, centreless conception of reality cannot accord the self or the first person any special status or centrality, since it is a conception from no point of view. It does not follow, however, that the absolute conception does not or must not encompass the self and its point of view, as it is some times claimed; if the absolute conception is to be consistent with itself, then, as I have argued elsewhere, it must encompass the totality of facts.[42] But, as Adrian

[41] David Wiggins, *Needs, Values, Truth* (revised 2nd ed., Oxford: Basil Blackwell, 1991), p. 346.

[42] See my, 'Subjectivity and the First Person: Some Reflections', *Philosophical Inquiry*, vol. XXI, (Summer-Fall 1999), pp. 1–27.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

Moore argues, if "this thought amounts to anything, it certainly does not amount to the absurd idea that being from no point of view is the same as, or even equivalent to, being from every point of view".[43]

Despite the fact that self-ascriptive judgments do not aspire to objectivity absolutely, or *sub specie æternitatis*, objectivity (what I called elsewhere *restricted objectivity* as opposed to *absolute objectivity*[44]) about self-ascriptive judgments of states of consciousness, I am suggesting, entails truth in the whole sense of truth[45], and admits vindicatory explanations of such judgments: it represents a strongly (but not crudely) *cognitivist* conception of the subject matter of subjectivity. If this is correct, the distinction between truth and falsity, or rather, truth and ineffability does not neatly correlate with the distinction between objectivity and subjectivity; both the subjective and the non-perspectival aspire to truth. And despite the metaphysical difference between the two domains, such a difference does not entail that there is an ambiguity, or a difference in the meaning of the notion of *fact*, nor does it entail that there are distinct kinds of fact. What it does entail is that there are facts about metaphysically distinct domains of the world. 'Subjective' and 'non-subjective' do not apply to facts; they apply at the level of the nature of things. There are facts about the subject and its states of consciousness, and there are facts about the non-subjective domain of the world[46] (and presumably some would argue that there are facts about the moral domain, the mathematical domain, and so on). This leads me to suggest that although the two metaphysically distinct domains can and do interact, they are neither in competition with one another, nor can one be classified as appropriate and the

[43] A. W. Moore, *Points of View* (Oxford, 1997), p. 66.

[44] See my, op. cit., [1999].

[45] As Wiggins says we should have regard 'for the way in which the marks of truth are common marks, present whenever truth is present'—hence his formula 'there is nothing else to think'. (David Wiggins, 'Replies', *Essays for David Wiggins: Identity, Truth and Value*, Sabina Lovibond and S. G. Williams (eds), (Oxford: Blackwell, 1996), p. 277). We could say that it is sufficient for some judgment that *p* (that one is in pain, for instance) being true that one could come to believe that *p* precisely because *p*; but we must be careful here not to think that 'the explanation postulates the existence of the [subjective] property'; rather, the belief is 'an intelligible and correct response' to the property (Wiggins, [1996], p. 261). Wiggins calls this explanation of belief *vindicatory*. As he argues: 'The question of vindication is a question not of sense but of reference or truth-value' (Wiggins, ibid., [1996], p. 255).

[46] For a fuller discussion and defence of all this see my, op. cit., [1999].

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

other inappropriate. If both domains are equally respectable, legitimate, and essential to our understanding of what there is, then it is futile to look for, or insist on a single homogeneous set of facts that uniquely classifies and exhausts the totality of what there is—the totality of facts[47].

The second part of my elucidation is concerned with the relation between the second immunity and some of the problematic notions that arise by its introduction. What the second immunity clearly brings out is that the self-ascription of states of *inner experience*, such as pains, or sensations, precisely because they make no claims about anything independent of the subject, guarantees, as we have seen, that such ascription is immune to error through *misascription*.[48] However, the self-ascription of states of *outer experience*, such as seeing, or believing (and, as I argued earlier, the self-ascription of bodily states, such as bleeding, or of physical states, perhaps with the exception of indexicals such as 'here'), are guaranteed only *one* immunity—that relative to the self-identification component—their predication or self-ascriptive component allows for the possibility of error.

First-person authority: States of outer experience have intentionality—the *aboutness* or *object-directedness* of such states—which is not only about how things are with me, but also about something independent of myself; they are about 'what is there *anyway*' and might therefore be subject to error. But it does not follow from this that such states are not subjective (as explained above), nor does it follow that, just because they are about something other than myself —their content is partly, or wholly constituted by the way the world is, or they are object-dependent—that there is a threat to *first-person authority* over my own mind. All my conscious states of mind are known to be had me, but not to you, *immediately*—that is, directly,

[47] See my, op. cit., [1999].

[48] This should not be thought of as committing me to any kind of *privacy*. The meaning of the term 'pain', like that of any other term, is public and objective. All our concepts are objective, if the Fregean theory of sense and reference is correct; it is just that some are physical concepts and some are mental. 'Objective' should not be confused with 'physical' (for a defence of this, see my, op. cit., [1999]. Wittgenstein suggests that our mental concepts replace the primitive, the natural expression of mental states (*Philosophical Investigations*, [trans.] G. E. M. Anscombe, (Oxford: Basil Blackwell, 1958), §244). Mental states are in need of outward manifestations (*PI* §580), and therefore mental ascriptions must be tied up with those manifestations, but mental concepts neither refer nor are reduced to those manifestations.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

without the need for criteria.[49] It is a condition of learning and understanding the first-person case that I do not consult any extraneous evidence, or criteria for knowing my own mind. It ought to be clear that it does not follow from the fact that I *need* no extraneous criteria for knowing my own states of consciousness that there *are* none. Central to understanding a predicate, a term, or an expression is grasping the conditions under which such a predicate (or term, etc.) can be correctly ascribed or self-ascribed. In Kantian terms, I self-ascribe not just in accordance with rules and criteria, but in accordance with *the idea* of a rule or a criterion, what might be referred to as the normativity of meaning—the recognition or grasping of normative relations and the space of reasons. The same holds for criterionless self-identification, or identification-free self-reference, vis-à-vis the concept 'I'—the recognition of the irreducibility of 'I' or the self. Just because there is no room for mediation (in the sense that there is no such thing as knowing from *evidence* in one's own case, or in self-identification), it does not follow that there is no room for truth either (in the sense that there is no such thing as knowing *without* evidence).[50]

Epistemic asymmetry: All this is in turn interrelated with *epistemic asymmetry*. Clearly other people can wonder whether, for example, I believe that *p*, and can in fact be unable to answer the question 'does she believe that *p*?' They need to establish not simply whether *p* is true, but also whether I believe that *p*. But from my first-person point of view I do not establish this at all; I only need to establish that *p* is true. It is absurd to think that in my own case I need to establish that it is true of *me* that *I* believe that *p*[51]. Thus, normally I am in a position to answer the same question asked of me straight off, directly or immediately. Any wondering or uncertainty on my part concerns the question whether *p* and not whether I believe that *p*; or it might be that I haven't made up my

[49] We should be careful not to conflate first- and third-person senses of directly knowing—directly knowing one's states of mind and directly knowing that someone is cheerful. In the latter sense, direct knowledge means non-inferential third-person knowledge, but it does not imply the absence for a need of criteria. As Evans and McDowell put it: 'Seeing cheerfulness in a face is not inferring that its owner is cheerful from the way his face looks. (Of course it is only because his face looks that way that we can see the cheerfulness in it)' (Gareth Evans and John McDowell (eds), *Truth and Meaning: Essays in Semantics*, (Oxford: Clarendon Press, 1976), p. xxii). No seeing, however, is needed in the former sense of directly knowing one's own mind.

[50] See my, op. cit., [1995].

[51] I discuss this in greater detail in my doctoral thesis, but also in a paper currently in progress on 'Wittgenstein on Moore's Paradox'.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

mind on the subject—perhaps I am wavering between belief and disbelief.

What can be brought out of these elucidatory remarks is that neither epistemic asymmetry nor first-person authority are the same as, or even equivalent to, the second kind of immunity. Neither epistemic asymmetry nor first-person authority *eo ipso* require that the self-ascription of states of consciousness is immune to error; I can believe that *p*, whether *p* is true or false. And in so far as I can be mistaken about the self-ascription of states of outer experience this is to do with the intentionality of those states, and not with what is available to first-person authority, since *all* states of consciousness are in principle available to such an authority, and the fact that some are intentional and some are not is irrelevant. What is relevant is the distinction between self-ascription which is subject to the second immunity and self-ascription which might not be. Furthermore, as I said in §3, the second immunity should not be confused with old-style incorrigibility, since the latter has been traditionally seen as involving an epistemological distinction between incorrigibility and corrigibility which, it has been argued, correlates with the ontological distinction between subjective and objective facts. Indeed, 'it may be regarded as a condition of the truth of the claim that they are subjective and they are also incorrigible'[52]. If my arguments are cogent, however, they demonstrate that there is no strict correlation between the two distinctions.

Shoemaker argues that it is important to see that the distinction Wittgenstein draws between the 'subject-use' and 'object-use' of 'I' 'is not the controversial distinction between 'corrigible' and 'incorrigible' first-person statements'[53], since the examples that Wittgenstein uses which involve the subject-use of 'I' are incorrigible. Unlike Wittgenstein, Shoemaker goes on to argue that with first-person statements which are clearly not incorrigible:

> I can be mistaken in saying 'I see a canary' since I can be mistaken in thinking that what I see is a canary [...] but it cannot happen that I am mistaken in saying this because I have misidentified as myself the person I know to see a canary.[54]

It is this insight that I have tried to bring out and defend, an insight that remains buried in Shoemaker's discussion since the explanation he offers is simply the reiteration of the claim that such statements

[52] McGinn, op. cit., [1993], p. 45.
[53] Shoemaker, op. cit., [1984], p. 7.
[54] Shoemaker, op. cit., [1984], p. 8.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

do not have 'absolute immunity' to error through misidentification, they have only 'circumstantial immunity' relative to 'I'.

6. The Double Immunity

I have argued that 'I' is governed by rule (R*) and leaves no latitude for the speaker's intentions in the determination of both its reference and its identification. Although Shoemaker can be seen as endorsing rule (R*), he still holds that if the self-reference of 'I' *as subject* were not possible, a lot of other reference would also not be possible. This argument is important to Shoemaker in establishing his thesis that the use of 'I' as subject has priority over its uses as object. But why does Shoemaker still insist on such priority? What would its paradigm case be? Does this not take us back to the worries with which we began? There are, I believe, two possible ways out of this. First, it might be argued that my argument does little to show the limitation of Shoemaker's position, even if he inadvertently lapses back into the two uses of 'I' and insists on the priority of 'I' in its *as subject* use, still, given the availability to him of the Strawsonian considerations that the use of 'I' can refer to '*both* states of consciousness *and* corporeal characteristics'. That would be true only if Strawson did not conflate discussion of self-ascription with discussion of self-identification, and more importantly, only if the Strawsonian considerations were supported by argument; however, when one scrutinises Strawson's argument one finds that those considerations are simply asserted (at best they are backed up by an implicit verificationism). One cannot assume that the immunity of 'I' to identification-failure "is guaranteed by the role of the expression in the ordinary practice, well established among human beings, of reference to themselves and each other"[55]. The claim that 'I' refers to an embodied subject must surely be the conclusion of one's enquiry in this area of philosophy and not an assumption, if one's arguments are to have the force they are intended to have against both the reductionist or 'no-reference' view of 'I', and the idealist conception of the self.

This leaves us with the second possible way out, the one I am proposing, for if my arguments are correct, then what is distinctive about them is that the conception of the self as an embodied self-conscious being is the *outcome* of my central argument. I have argued that first-person judgments that involve the self-ascription of states of inner experience, such as pain, have a *double* immunity:

[55] Strawson, op. cit., [1994], p. 211.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

they are immune to error both with regard to their identification and to their predication or ascriptive component. My argument has been that the unique nature of 'I' guarantees that there is a speaker or thinker that 'I' refers to and identifies, and that such identification is, in all cases, immune to error through misidentification *simpliciter*; the upshot is that those who claim that there are counter-examples to my thesis and argue that some first-person thoughts are subject to error through misidentification relative to 'I', are trading on an equivocation between two different statements. I have also argued that considerations of the '*know who*', '*know which*' or '*know what*' requirement are appropriately relevant only to a properly developed discussion of self-ascription, for although self-ascription presupposes self-identification, and therefore can neither be subsumed under, nor conflated with self-identification, the latter is identification-free. Moreover, my argument neither undermines the idea that 'I' is a referring and an identifying expression, nor entails that its referent can only be a disembodied being. Such conclusive and weighty metaphysical commitments cannot be derived, if at all, from considerations of self-reference and self-identification alone. Rather, the subject making or expressing a first-person statement cannot use 'I' so as wrongly to identify who is speaking or thinking, *who*ever he might be, or *what*ever he might be. All this brings out the importance of (R*), for it clearly makes the immunity to error through misidentification a feature of the *sense* of 'I': if the subject did make a mistake, this would show that he has ceased to understand the term 'I' and its cognates, or that he has always understood it wrongly—a more disturbing possibility. It would show that the first-person case had collapsed.[56] Thus, if my challenge is cogent, the immunity to error through misidentification is absolute: it is not limited to first-person statements that involve the self-ascription of states of inner experience, but extends to all first-person statements, it extends to those that involve the self-ascription of states of outer experience as well as the self-ascription of bodily states. We could, therefore, conclude that:

(a) there is only *one* use of 'I';
(b) in *all* first-person statements, the reference and identity effected by the use of 'I' or its cognates is immune to error through misidentification—it is governed by (R*);
(c) first-person statements that involve the self-ascription of states of inner experience have a *double* immunity.

[56] For a defence of this see my, op. cit., [1995].

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press

7. Self-consciousness

The picture that emerges from these considerations is that it ought to be undeniable that where there is consciousness there is a point of view from which the world is perceived and experienced, and where there is self-consciousness there is a first-person point of view, an 'I' or a self that has reason and reflection not only on its own states of consciousness but also on the way the world is independently of itself. What gives self-consciousness its unity and substantiality is precisely its being held together by an 'I' or a self that is capable both of self-identification and of self-ascription of states of consciousness (inner and outer) and of bodily states.

Evans has argued that the essence of self-consciousness is self-identification[57], but if my arguments are correct, then the essence of self-consciousness is *both* self-ascription *and* self-identification. Arguments in this area that have been stimulated by the works of the philosophers discussed here, have tried to explain the immunity to error in first-person judgments; at best they explain, to a certain extent, the first immunity but give us no grip on the second. These arguments share the mistaken premise that judgments involving the self-ascription of bodily states cannot be absolutely immune to error in their self-identification component, but are, at best, circumstantially or non-absolutely immune to error through misidentification relative to 'I', and at worst, subject to error. My arguments, I believe, have exposed the mistaken premise of conceiving the body simply as external to the thinking subject, a view that fails to appreciate the fact that there is an intimate and substantial union between them, as Descartes rightly argued, and which constitutes what he called man or person. This is important, for it helps to bring home to us the fact that there is nothing in the use of 'I' that forces upon us either the two uses of 'I' (the subject-use of 'I' and the so-called object-use of 'I'), or a distinction between absolute and circumstantial immunity to error; more seriously, it does not force upon us an actual separation between the two components of our nature.[58] Indeed, there is nothing in the use of 'I' that forces upon us an actual disjunction: either a physicalist conception, or an idealist conception of what we are. We come to realize that we are embodied self-conscious beings and that 'I' is immune to error through misidentification *simpliciter*.[59]

Worcester College, Oxford

[57] Evans, op. cit., [1982], p. 213.

[58] As Descartes argued, *real* distinction between mind and body does not mean *actual* separation or segregation, something that his critics both at the time and at present have often missed.

[59] I should like to thank Stephen Blamey, Peter J. King, David Wiggins, and Stephen Williams for their helpful comments on an earlier draft.

https://doi.org/10.1017/S0031819100000668 Published online by Cambridge University Press