

such supposition and *therefore* are unprepared with an answer. We need not have determinately meant the word “see” one way or the other.

We may make a similar point about ‘phantom limb’. I take the part of the body where pain is felt to be the object of a transitive verb-like expression “to feel pain in —”. Then when there is, e.g., no foot, but X, not knowing this, says he feels pain in his foot, he may say he was wrong (“I did not see a lion there, for there was no lion”) or he may alter his understanding of the phrase “my foot” so that it becomes a purely intentional object of the verb-like expression. But it need not be determined in advance, in the normal case of feeling pain, whether one so intends the expression “I feel pain in —” as to withdraw it, or merely alters one’s intentions for the description of the place of the pain, if one should learn that the place was missing.

## 2 The First Person

Descartes and St Augustine share not only the argument *Cogito ergo sum* – in Augustine *Si fallor, sum* (*De Civitate Dei*, XI, 26) – but also the corollary argument claiming to prove that *the mind* (Augustine) or, as Descartes puts it, *this I*, is not any kind of body. “I could suppose I had no body,” wrote Descartes, “but not that I was not”, and inferred that “this I” is not a body. Augustine says “The mind knows itself to think”, and “it knows its own substance”: hence “it is certain of being that alone, which alone it is certain of being” (*De Trinitate*, Book X). Augustine is not here explicitly offering an argument in the first person, as Descartes is. The first-person character of Descartes’ argument means that each person must administer it to himself in the first person; and the assent to St Augustine’s various propositions will equally be made, if at all, by appropriating them in the first person. In these writers there is the assumption that when one says “I” or “the mind”, one is naming something such that the knowledge of its existence, which is a knowledge of itself as thinking in all the various modes, determines what it is that is known to exist.

Saul Kripke has tried to reinstate Descartes’ argument for his dualism. But he neglects its essentially first-person character, making it an argument about the non-identity of *Descartes* with his own body. Whatever else is said, it seems clear that the argument in Descartes depends on results of applying the method of doubt.<sup>1</sup> But by that method Descartes must have doubted the existence of the man Descartes: at any rate of that figure in the world of his time, that Frenchman, born of such-and-such a stock and christened René;

<sup>1</sup> *Principles of Philosophy*, I, LX, contains Descartes’ best statement, which is I think immune to the usual accusation of substitutional fallacy: “Each of us conceives of himself as a conscious being, and can in thought exclude from himself any other substance, whether conscious or extended; so from this mere fact it is certain that each of us, so regarded, is really distinct from every other conscious substance and from every corporeal substance. And even if we supposed that God had conjoined some corporeal substance to such a conscious substance so closely that they could not be more closely joined, and had thus compounded a unity out of the two, yet even so they remain really distinct” (*Philosophical Writings*, trans. G. E. M. Anscombe and P. T. Geach). Rendering Descartes’ premise here as “I can conceive myself not to include or be my body”, we come close to Kripke’s version (but in the first person) “Possibly I am not A”, where “A” means my body. But why can I so conceive myself if not because I can doubt the existence of my body?

But “doubting” here does not mean merely reflecting that I am ignorant of the existence of my body though not of myself. So understood, the argument would indeed involve the substitutional fallacy. “Doubting” means clearly understanding that the existence of my body is not guaranteed by something which is thoroughly understood, and is all I am sure of: the existence of myself. We see the importance of the premise supplied by St Augustine “The mind knows its own substance”.

but also, even of the man – unless a man isn't a sort of animal. If, then, the non-identity of himself with his own body follows from his starting-points, so equally does the non-identity of himself with the man Descartes. "I am not Descartes" was just as sound a conclusion for him to draw as "I am not a body". To cast the argument in the third person, replacing "I" by "Descartes", is to miss this. Descartes would have accepted the conclusion. That mundane, practical, everyday sense in which it would have been correct for him to say "I am Descartes" was of no relevance to him in these arguments. That which is named by "I" – that, in his book, was not Descartes.

It may seem strange to say: "The non-identity of himself with Descartes was as valid a conclusion as the other" and not treat this as already a *reductio ad absurdum*. For is that phrase not equivalent to "the non-identity of Descartes with Descartes"?

No. It is not. For what is in question is not the ordinary reflexive pronoun, but a peculiar reflexive, which has to be explained in terms of "I". It is the reflexive called by grammarians the 'indirect reflexive' and there are languages (Greek, for example) in which there is a special form for it.<sup>2</sup>

"When John Smith spoke of James Robinson he was speaking of his brother, but he did not know this." That's a possible situation. So similarly is "When John Smith spoke of John Horatio Auberon Smith (named in a will perhaps) he was speaking of himself, but he did not know this." If so, then 'speaking of' or 'referring to' oneself is compatible with not knowing that the object one speaks of is oneself.

Yet we are inclined to think that "It's the word each one uses in speaking of himself" explains what "I" names, or explains "I" as a 'referring expression'. It cannot do so if "He speaks of himself" is compatible with ignorance and we are using the reflexive pronoun, in both cases, in the ordinary way.

Nor can we explain the matter, as we might suppose, by saying "I" is the word each one uses when he knowingly and intentionally speaks of himself". For did not Smith knowingly and intentionally speak of Smith? Was not the person he intended to speak of – Smith? and so *was* not the person he intended to speak of – himself?

It may be said: "Not in the relevant sense. We all know you can't substitute every designation of the object he intended to speak of and keep the statement about his intention true." But that is not the answer unless the reflexive pronoun itself is a sufficient indication of the way the object is specified. And that is something the ordinary reflexive pronoun cannot be. Consider: "Smith realizes (fails to realize) the identity of an object he calls 'Smith' with himself." If the reflexive pronoun there is the ordinary one, then it specifies

<sup>2</sup> ε, οὐ, οί. See Thucydides II. 13. The form is rare. Credit for discerning the indirect reflexive in English, which does not have a distinct form for it, belongs in the present day to H.-N. Castaneda in "The Logic of Self-Knowledge", *Noûs*, I (1967), 9–22. But his presentation is excessively complicated and I believe it has not attracted enough attention to the substantive point.

for us who frame or hear the sentence an object, whose identity with the object he calls "Smith" Smith does or doesn't realize: namely the object designated by our subject word "Smith". But that does not tell us what identity Smith himself realizes (or fails to realize). For, as Frege held, there is no path back from reference to sense; any object has many ways of being specified, and in this case, through the peculiarity of the construction, we have succeeded in specifying an object (by means of the subject of our sentence) without specifying any conception under which *Smith's* mind is supposed to latch on to it. For we don't want to say "Smith does not realize the identity of Smith with Smith".

We only have to admit a failure of specification of the intended identity, if we persist in treating the reflexive in "He doesn't realize the identity with himself" as the ordinary reflexive. In practice we have no difficulty at all. We know what we mean Smith doesn't realize. It is: "I am Smith." But if that is how we understand that reflexive, it is not the ordinary one. It is a special one which can be explained only in terms of the first person.

If that is right, the explanation of the word "I" as 'the word which each of us uses to speak of himself' is hardly an explanation! – At least, it is no explanation if that reflexive has in turn to be explained in terms of "I"; and if it is the ordinary reflexive, we are back at square one. We seem to need a sense to be specified for this quasi-name "I". To repeat the Frege point: we haven't got this sense just by being told which object a man will be speaking of, whether he knows it or not, when he says "I". Of course that phrase "whether he knows it or not" seems highly absurd. His use of "I" surely guarantees that he does know it! But we have a right to ask *what* he knows; if "I" expresses a way its object is reached by him, what Frege called an "Art des Gegebenseins", we want to know what that way is and how it comes about that the only object reached in that way by anyone is identical with himself.

To say all this is to treat "I" as a sort of proper name. That's what gets us into this jam. Certainly "I" functions syntactically like a name. However, it has been observed not to be a proper name. Now this observation may strike us as obvious enough in a trivial sense. After all, we don't call it a proper noun but a personal *pronoun*. It is at any rate not an ordinary proper name. It could not have a lot of the characteristic use of a proper name. For if it is such, it is one that everyone has, and, worse still, one that each person uses only to refer to that person that he himself is. So it's no use for introducing people to one another, or for calling to someone, or for summoning him. And while it might be used as a signature (like the signature of an aged and doddering parson that I heard of, on someone's marriage lines: Me, Vicar), one would be quite dependent on other clues to the identity of the signatory. If this were the only name anyone had, the situation would be worse than it is for a bank in a Welsh village. These inconveniences are avoided, of course, because there are other more various proper names which people have as well. So the observation that "I" is not a proper name seems to reduce to the

triviality that we perhaps would not call a word a proper name if everyone had it and used it only to speak of himself. – But is even that true? After all, all Sikhs seem to be called “Singh”. So the real difference lies in that one point that each one uses the name “I” only to speak of himself. Is that a ground not to call it a proper name? Certainly to the eyes of our logicians it is a proper name. Are their eyes dim? Or is it really logically a proper name?

Let us ask: is it really true that “I” is only not called a proper name because everyone uses it only to refer to himself? Let us construct a clear case of just such a name. Imagine a society in which everyone is labelled with two names. One appears on their backs and at the top of their chests, and these names, which their bearers cannot see, are various: “B” to “Z” let us say. The other, “A”, is stamped on the inside of their wrists, and is the same for everyone. In making reports on people’s actions everyone uses the names on their chests or backs if he can see these names or is used to seeing them. Everyone also learns to respond to utterance of the name on his own chest and back in the sort of way and circumstances in which we tend to respond to utterance of our names. Reports on one’s own actions, which one gives straight off from observation, are made using the name on the wrist. Such reports are made, not on the basis of observation alone, but also on that of inference and testimony or other information. *B*, for example, derives conclusions expressed by sentences with “A” as subject, from other people’s statements using “B” as subject.

It may be asked: what is meant by “reports on one’s own actions”? Let us lay it down that this means, for example, reports issuing from the mouth of *B* on the actions of *B*. That is to say: reports from the mouth of *B* saying that *A* did such-and-such are *prima facie* verified by ascertaining that *B* did it and are decisively falsified by finding that he did not.

Thus for each person there is one person of whom he has characteristically limited and also characteristically privileged views: except in mirrors he never sees the whole person, and can only get rather special views of what he does see. Some of these are specially good, others specially bad. Of course, a man *B* may sometimes make a mistake through seeing the name “A” on the wrist of another, and not realizing it is the wrist of a man whose other name is after all not inaccessible to *B* in the special way in which his own name (“B”) is.

(It may help some people’s imagination if we change the example: instead of these rather inhuman people, we suppose machines that are equipped with scanning devices are marked with signs in the same way as the people in my story were marked with their names, and are programmed to translate what appears on the screens of their scanners into reports.)

In my story we have a specification of a sign as a name, the same for everyone, but used by each only to speak of himself. How does it compare with “I”? – The first thing to note is that our description does not include self-consciousness on the part of the people who use the name “A” as I have described it. They perhaps have no self-consciousness, though each one knows a lot about the object that he (in fact) is; and has a name, the same as

everyone else has, which he uses in reports about the object that he (in fact) is.

This – that they have not self-consciousness – may, just for that reason, seem not to be true. *B* is conscious of, that is to say he observes, some of *B*’s activities, that is to say his own. He uses the name “A”, as does everyone else, to refer to himself. So he is conscious of himself. So he has self-consciousness.

But when we speak of self-consciousness we don’t mean that. We mean something manifested by the use of “I” as opposed to “A”.

Hence we must get to understand self-consciousness. Unsurprisingly, the term dates only from the seventeenth century and derives from philosophy. Getting into ordinary language, it alters, and by the nineteenth century acquires a sense which is pretty irrelevant to the philosophical notion: it comes to mean awkwardness from being troubled by the feeling of being an object of observation by other people. Such a change often happens to philosophical terms. But this one also gets into psychology and psychiatry, and here its sense is not so far removed from the philosophical one.

The first explanation of self-consciousness that may occur to someone, and what the form of the expression suggests, is this: it is consciousness of a self. A self will be something that some things either have or are. If a thing has it it is something connected with the thing, in virtue of which the thing that has it is able to say, and mean, “I”. It is what he calls “I”. Being able to mean “I” is thus explained as having the right sort of thing to call “I”. The fanciful use of the word, if someone should put a placard “I am only a wax-work” on a wax policeman, or in the label on the bottle in *Alice in Wonderland* “Drink me”, is a pretence that the objects in question have (or are) selves. The *self* is not a Cartesian idea, but it may be tacked on to Cartesian Ego theory and is a more consequent development of it than Descartes’ identification of ‘this I’ with his soul. If things are, rather than having, selves, then a self is something, for example a human being, in a special aspect, an aspect which he has as soon as he becomes a ‘person’. “I” will then be the name used by each one only for himself (this is a direct reflexive) and precisely in that aspect.

On these views one would explain “self” in “self-consciousness” either by explaining what sort of object that accompanying self was, or by explaining what the aspect was. Given such explanation, one might have that special ‘way of being given’ of an object which is associated with the name one uses in speaking of it.

Now all this is strictly nonsensical. It is blown up out of a misconstrue of the reflexive pronoun. That it is nonsense comes out also in the following fact: it would be a question what guaranteed that one got hold of the right self, that is, that the self a man called “I” was always connected with him or was always the man himself. Alternatively, if one said that “the self connected with a man” meant just the one he meant by “I” at any time, whatever self that was, it would be by a mere favour of fate that it had anything else to do with him.

But “self-consciousness” is not any such nonsense. It is something real, though as yet unexplained, which “I”-users have and which would be lacking to “A”-users, if their use of “A” was an adequate tool for their consciousness of themselves.

The expression “self-consciousness” can be respectably explained as ‘consciousness that such-and-such holds of oneself’. Nor should we allow an argument running: since the occurrence of “oneself” is just like the occurrence of “himself” which left us perfectly well understanding what Smith failed to realize, the word “self” must itself connote the desired ‘way of being given’ that is associated with “I” as (logically speaking) a proper name. We must reject this argument because “oneself” is here nothing but the indirect reflexive: that is to say, the reflexive of indirect speech. Understanding indirect speech we know what the related direct speech is. That is all.

These considerations will lack appeal. The question was, what does “I” stand for? If that question is asked, and “I” is supposed to stand for its object as a proper name does, we need an account of a certain kind. The use of a name for an object is connected with a conception of that object. And so we are driven to look for something that, for each “I”-user, will be the conception related to the supposed name “I”, as the conception of a city is to the names “London” and “Chicago”, that of a river to “Thames” and “Nile”, that of a man to “John” and “Pat”. Such a conception is requisite if “I” is a name, and there is no conception that can claim to do the job except one suggested by ‘self-consciousness’. That is why some philosophers have elaborated the notion of ‘selves’ (or ‘persons’ defined in terms of self-consciousness) and conducted investigations to see what such things may be. And just as we must be continuing our reference to the same city if we continue to use “London” with the same reference, so we must each of us be continuing our reference to the same self (or ‘person’) if we continue to use “I” with the same reference.

This led to an imaginative *tour de force* on the part of Locke: might not the thinking substance which thought the thought “I did it” – the genuine thought of agent-memory – nevertheless be a different thinking substance from the one that could have had the thought: “I am doing it” when the act was done? Thus he detached the identity of the self or ‘person’ from the identity even of the thinking being which does the actual thinking of the I-thoughts.

Considerations about reflexive pronouns are certainly not going to dam up the flood of inquiries about ‘the self’ or ‘selves’, so long as “I” is treated as a name and a correlative term is needed for its type of object. Nevertheless, these are embarrassing credentials for such inquiries. And a self *can* be thought of as what “I” stands for, or indicates, without taking “I” as a proper name. The reasons for considering it as a proper name were two: first, that to the logician’s eye it is one, and second, that it seemed to be just like our “A” (which was clearly a proper name) except that it expressed ‘self-consciousness’. So we tried to explain it as a proper name of a self. Now a lot

of people who will have no objection to the talk of ‘selves’ will yet feel uneasy about calling “I” a proper name of a self or anything else. I assume it was made clear that the different reference in each mouth was not an objection (there is no objection to calling “A” a proper name), and so there is some other reason. The reason, I think, is that, so understood, a repeated use of “I” in connection with the same self would have to involve a reidentification of that self. For it is presumably always a use in the presence of its object! There is no objection to the topic of reidentification of selves – it is one of the main interests of the philosophers who write about selves – but this is not any part of the role of “I”. The corresponding reidentification *was* involved in the use of “A”, and that makes an additional difference between them.

So perhaps “I” is not a name but rather another kind of expression indicating ‘singular reference’. The logician’s conception of the proper name after all only required *this* feature. There are expressions which logically and syntactically function as proper names without being names. Possibly definite descriptions do, and certainly some pronouns. “I” is called a pronoun, so we will consider this first. Unluckily the category ‘pronoun’ tells us nothing, since a singular pronoun may even be a variable (as in “If anyone says that, *he* is a fool”) – and hence not any kind of singular designation of an object. The suggestion of the word “pronoun” itself is not generally borne out by pronouns. Namely, that you get the same sense in a sentence if you replace the pronoun in it by a name, common or proper: what name in particular, it would be difficult to give a general rule for. Perhaps “pronoun” seemed an apt name just for the personal pronouns and especially for “I”. But the sense of the lie “I am not E.A.” is hardly retained in “E.A. is not E.A.”. So that suggestion is of little value.

Those singular pronouns called demonstratives (“this” and “that”) are a clear example of non-names which function logically as names. For in true propositions containing them they provide reference to a distinctly identifiable subject-term (an object) of which something is predicated. Perhaps, then, “I” is a kind of demonstrative.

Assimilation to a demonstrative will not – as would at one time have been thought – do away with the demand for a conception of the object indicated. For, even though someone may say just “this” or “that”, we need to know the answer to the question “this *what?*” if we are to understand him; and he needs to know the answer if he is to be meaning anything.<sup>3</sup>

Thus a singular demonstrative, used correctly, does provide us with a proper logical subject so long as it does not lack a ‘bearer’ or ‘referent’, and

<sup>3</sup> This point was not grasped in the days when people believed in pure ostensive definition without the ground’s being prepared for it. Thus also in those days it was possible not to be so much impressed as we ought to be, by the fact that we can find no well-accounted-for term corresponding to “I” as “city” does to “London”. It was possible to see that there was no ‘sense’ (in Frege’s sense) for “I” as a proper name, but still to think that for each one of us “I” was the proper name of an ‘object of acquaintance’, a *this*. What *this* was could then be called “a self”, and the word “self” would be felt to need no further justification. Thus, for example, McTaggart. See *The Nature of Existence* (Cambridge, 1921–7), Vol. II, 493–4, 586–7, 390–1, 394.

so it conforms to the logician's requirement for a name. And the answer to the question "this what?" might be taken to be "this self", if it can be shown that there are selves and that they are apparently what is spoken of by all these people saying "I". Thus would these philosophical inquiries about selves have a certain excuse.

It used to be thought that a singular demonstrative, "this" or "that", if used correctly, could not lack a referent. But this is not so, as comes out if we consider the requirement for an answer to "this what?" Someone comes with a box and says "This is all that is left of poor Jones." The answer to "this what?" is "this parcel of ashes"; but unknown to the speaker the box is empty. What "this" has to have, if used correctly, is something that it *latches on to* (as I will put it): in this example it is the box. In another example it might be an optical presentation. Thus I may ask "What's that figure standing in front of the rock, a man or a post?" and there may be no such object at all; but there is an appearance, a stain perhaps, or other marking of the rock face, which my "that" latches on to. The referent and what "this" latches on to may coincide, as when I say "this buzzing in my ears is dreadful", or, after listening to a speech, "That was splendid!" But they do not have to coincide, and the referent is the object of which the predicate is predicated where "this" or "that" is a subject.

There is no other pronoun but a demonstrative to which "I" could plausibly be assimilated as a singular term that provides a reference. Of course someone may say: "Why assimilate it at all? Each thing is what it is and not another thing! So 'I' is a pronoun all right, but it is merely the pronoun that it is." But that is no good, because 'pronoun' is just a rag-bag category; one might as well say: "It is the word that it is." The problem is to describe its meaning. And, if its meaning involves the idea of reference, to see what 'reference' is here, and how accomplished. We are now supposing that it is not accomplished as it is for a regular proper name; then, if "I" is not an abbreviation of a definite description, it must catch hold of its object in some other way – and what way is there but the demonstrative?

But there is a contrast between "I" and the ordinary demonstrative. We saw that there may be reference-failure for "this", in that one may mean "this parcel of ashes" when there are no ashes. But "I" – if it makes a reference, if, that is, its mode of meaning is that it is supposed to make a reference – is secure against reference-failure. Just thinking "I . . ." guarantees not only the existence but the presence of its referent. It guarantees the existence because it guarantees the presence, which is presence to consciousness. But note that here "presence to consciousness" means physical or real presence, not just that one is thinking of the thing. For if the thinking did not guarantee the presence, the existence of the referent could be doubted. For the same reason, if "I" is a name it cannot be an empty name. I's existence is existence in the thinking of the thought expressed by "I . . ." This of course is the point of the *cogito* – and, I will show, of the corollary argument too.

Whether "I" is a name or a demonstrative, there is the same need of a

'conception' through which it attaches to its object. Now what conception can be suggested, other than that of *thinking*, the thinking of the I-thought, which secures this guarantee against reference-failure? It may be very well to describe what selves are; but if I do not know that I am a self, then I cannot mean a self by "I".

To point this up, let me imagine a logician, for whom the syntactical character of "I" as a proper name is quite sufficient to guarantee it as such, and for whom the truth of propositions with it as subject is therefore enough to guarantee the existence of the object it names. He, of course, grants all that I have pointed out about the indirect reflexive. It cannot perturb him, so long as the 'way of being given' is of no concern to him. To him it is clear that "I", in my mouth, is just another name for E.A., "I" may have some curious characteristics; but they don't interest him. The reason is that "I" is a name governed by the following rule: 'If X makes assertions with "I" as subject, then those assertions will be true if and only if the predicates used thus assertively are true of X.'<sup>4</sup> This will be why Kripke – and others discussing Descartes – make the transition from Descartes' "I" to "*Descartes*".

Now first, this offers too swift a refutation of Descartes. In order to infer straight away that Descartes was wrong, we only need the information that Descartes asserted "I am not a body", together with the knowledge that he was a man: that is, an animal of a certain species; that is, a body living with a certain sort of life.

But there would and should come from Descartes' lips or pen a denial that, strictly speaking, *the man Descartes* made the assertion. The rule was sound enough. But the asserting subject must be the thinking subject. If you are a speaker who says "I", you do not find out what is saying "I". You do not for example look to see what apparatus the noise comes out of and assume that that is the sayer; or frame the hypothesis of something connected with it that is the sayer. If that were in question, you could doubt whether anything *was* saying "I". As, indeed, you can doubt whether anything is saying it out loud. (And sometimes *that* doubt is correct.)

Thus we need to press our logician about the 'guaranteed reference' of "I". In granting this, there are three degrees of it that he may assert.

(1) He may say that of course the user of "I" must exist, otherwise he would not be using "I". As he is the referent, that is what 'guaranteed reference' amounts to. In respect of such guaranteed reference, he may add, there will be no difference between "I" and "A". But the question is, why "I" was said to *refer* to the "I"-user? Our logician held that "I" was logically a proper name – a singular term whose role is to make a reference – for two reasons: one, that "I" has the same syntactical place as such expressions, and the other, that it can be replaced *salva veritate* by a (more ordinary) name of X

<sup>4</sup> My colleague Dr J. Altham has pointed out to me a difficulty about this rule about "I". How is one to extract the *predicate* for purposes of this rule in "I think John loves me"? The rule needs supplementation: where "I" or "me" occurs within an oblique context, the predicate is to be specified by replacing "I" or "me" by the indirect reflexive pronoun.

when it occurs in subject position in assertions made by *X*. In saying this, he no doubt thought himself committed to no views on the sense of "I" or what the "I"-user means by "I". But his second reason amounts to this: one who hears or reads a statement with "I" as subject needs to know whose statement it is if he wants to know of whom the predicate holds if the statement is true. Now, this requirement could be signalled by flashing a green light, say, in connection with the predicate, or perhaps adding a terminal '-O' to it. (I apologize to anyone who finds this suggestion altogether too fanciful, and beg him to suspend disbelief.) What would make such a signal or suffix into a referring expression? The essential argument cannot be an argument back from syntax to reference, for such an argument would depend only on the form of sentence and would be absurd. (e.g. no one thinks that "it is raining" contains a referring expression, "it".) And so it seems that our logician cannot disclaim concern with the sense of "I", or at any rate with what the "I"-user must mean.

(2) So the "I"-user must intend to refer to something, if "I" is a referring expression. And now there are two different things for "guaranteed reference" to mean here. It may mean (2a) guaranteed existence of the object meant by the user. That is to say, that object must exist, which he is taking something to be when he uses the expression in connection with it. Thus, if I suppose I know someone called "X" and I call something "X" with the intention of referring to that person, a guarantee of reference in this sense would be a guarantee that there is such a thing as *X*. The name "A" which I invented would have this sort of guaranteed reference. The "A"-user means to speak of a certain human being, one who falls under his observation in a rather special way. That person is himself, and so, given that he has grasped the use of "A", he cannot but be speaking of a real person.

If our logician takes this as an adequate account of the guaranteed reference of "I", then he will have to grant that there is a third sort of 'guaranteed reference', which "I" does *not* have. Guaranteed reference for that name "X" in this further sense (2b) would entail a guarantee, not just that there is such a thing as *X*, but also that what I take to be *X* is *X*. We saw that the "A"-user would not be immune to mistaken identification of someone else as 'A'. Will it also be so with "I"?

The suggestion seems absurd. It seems clear that if "I" is a 'referring expression' at all, it has both kinds of guaranteed reference. The object an "I"-user means by it must exist so long as he is using "I", nor can he take the wrong object to be the object he means by "I". (The bishop may take the lady's knee for his, but could he take the lady herself to be himself?)

Let us waive the question about the sense of "I" and ask *only* how reference to the right object could be guaranteed. (This is appropriate, because people surely have here the idea of a sort of pure direct reference in which one simply first means and then refers to an object before one.) It seems, then, that this reference could only be sure-fire if the referent of "I" were both freshly defined with each use of "I", and also remained in view so long as

something was being taken to be *I*. Even so there is an assumption that something else does not surreptitiously take its place. Perhaps we should say: such an assumption is extremely safe for "I", and it would be altogether an excess of scepticism to doubt it! So we accept the assumption, and it seems to follow that what "I" stands for must be a Cartesian Ego.

For, let us suppose that it is some other object. A plausible one would be *this body*. And now imagine that I get into a state of 'sensory deprivation'. Sight is cut off, and I am locally anaesthetized everywhere, perhaps floated in a tank of tepid water; I am unable to speak, or to touch any part of my body with any other. Now I tell myself "I won't let this happen again!" If the object meant by "I" is this body, this human being, then in these circumstances it won't be present to my senses; and how else can it be 'present to' me? But have I lost what I mean by "I"? Is that not present to me? Am I reduced to, as it were, 'referring in absence'? I have not lost my 'self-consciousness'; nor can what I mean by "I" be an object no longer present to me. This both seems right in itself, and will be required by the 'guaranteed reference' that we are considering.

Like considerations will operate for other suggestions. Nothing but a Cartesian Ego will serve. Or, rather, a stretch of one. People have sometimes queried how Descartes could conclude to his *RES cogitans*.<sup>5</sup> But this is to forget that Descartes declares its essence to be nothing but thinking. The thinking that thinks this thought – that is what is guaranteed by "cogito".

Thus we discover that if "I" is a referring expression, then Descartes was right about what the referent was. His position has, however, the intolerable difficulty of requiring an identification of the same referent in different "I"-thoughts. (This led Russell at one point to speak of 'short-term selves'.)

Our questions were a combined *reductio ad absurdum* of the idea of "I" as a word whose role is to 'make a singular reference'. I mean the questions how one is guaranteed to get the object right, whether one may safely assume no unnoticed substitution, whether one could refer to oneself 'in absence', and so on. The suggestion of getting the object right collapses into absurdity when we work it out and try to describe how getting hold of the wrong object may be excluded.

How, even, could one justify the assumption, if it is an assumption, that there is just one thinking which is this thinking of this thought that I am thinking, just one thinker? How do I know that 'I' is not ten thinkers thinking in unison? Or perhaps not quite succeeding. That might account for the confusion of thought which I sometimes feel. – Consider the reply "Legion, for we are many", given by the possessed man in the gospel. Perhaps we should take that solemnly, not as a grammatical joke.<sup>6</sup> These

<sup>5</sup> For example A. J. Ayer. See *Language, Truth and Logic* (2nd edn, London, 1946), p. 142.

<sup>6</sup> Ambrose Bierce has a pleasant entry under "I" in the *Devil's Dictionary*: "I is the first letter of the alphabet, the first word of the language, the first thought of the mind, the first object of the affections. In grammar it is a pronoun of the first person and singular number. Its plural is said to be *We*, but how there can be more than one myself is doubtless clearer to the grammarians

considerations refute the 'definite description' account of "I". For the only serious candidate for such an account is "The sayer of this", where "sayer" implies "thinker".

Getting hold of the wrong object is excluded, and that makes us think that getting hold of the right object is guaranteed. But the reason is that there is no getting hold of an object at all. With names, or denoting expressions (in Russell's sense) there are two things to grasp: the kind of use, and what to apply them to from time to time. With "I" there is only the use.

If this is too hard to believe, if "I" is a 'referring expression', then Descartes was right. But now the troubles start. At first, it seems as if what "I" stands for ought to be the clearest and certainest thing – what anyone thinking of his own thinking and his own awareness of anything is most evidently aware of. It is most certain because, as Augustine said, it is involved in the knowledge of all mental acts or states by the one who has them. They could not be doubted. But the *I*, the 'mind', the 'self', was their subject, not their object, and looking for it as an object resulted, some people thought, in total failure. It was not to be found. It was rather as it were an area of darkness out of which light shone on everything else. So some racked their brains over what this invisible subject and the 'thinking of it' could be; others thought there was no such thing, there were just all the objects, and hence that "I", rather, was the name of the whole collection of perceptions. But that hardly fitted its grammar, and anyway – a problem which utterly stumped Hume – by what was *I* made into a unity? Others in effect treat selves as postulated objects for "I" to be names of in different people's mouths. Yet others denied that the self was invisible, and claimed that there is a unique feeling of oneself which is indescribable but very, very important, especially in psychology, in clinical psychology, and psychiatry.

With that thought: "The *I* was subject, not object, and hence invisible", we have an example of language itself being as it were possessed of an imagination, forcing its image upon us.

The dispute is self-perpetuating, endless, irresoluble, so long as we adhere to the initial assumption, made so far by all the parties to it: that "I" is a referring expression. So long as that is the assumption you will get the deep division between those whose considerations show that they have not perceived the difficulty – for them "I" is in principle no different from my "A"; and those who do – or would – perceive the difference and are led to rave in consequence.

And this is the solution: "I" is neither a name nor another kind of expression whose logical role is to make a reference, *at all*.

Of course we must accept the rule "If *X* asserts something with 'I' as subject, his assertion will be true if and only if what he asserts is true of *X*." But if someone thinks that is a sufficient account of "I", we must say "No, it

than it is to the author of this incomparable dictionary. Conception of two myselfs is difficult, but fine. The frank yet graceful use of "I" distinguishes a good author from a bad; the latter carries it with the manner of a thief trying to cloak his loot."

is not", for it does not make any difference between "I" and "A". The truth-condition of the whole sentence does not determine the meaning of the items within the sentence. Thus the rule does not justify the idea that "I", coming out of *X*'s mouth, is another name for *X*. Or for anything else, such as an asserting subject who is speaking through *X*.

But the rule does mean that the question "Whose assertion?" is all-important. And, for example, an interpreter might repeat the "I" of his principal in his translations. Herein resides the conceivability of the following: someone stands before me and says, "Try to believe this: when I say "I", that does not mean this human being who is making the noise. I am someone else who has borrowed this human being to speak through him." When I say "conceivability" I don't mean that such a communication might be the truth, but only that our imagination makes something of the idea. (Mediums, possession.)

If I am right in my general thesis, there is an important consequence – namely, that "I am E.A." is after all not an identity proposition. It is connected with an identity proposition, namely, "This thing here is E.A.". But there is also the proposition "I am this thing here".

When a man does not know his identity, has, as we say, 'lost his memory', what he doesn't know is usually that *that* person he'd point to in pointing to himself (this is the direct reflexive) is, say, Smith, a man of such-and-such a background. He has neither lost the use of "I", nor would he feel at a loss what to point to as his body, or as the person he is; nor would he point to an unexpected body, to a stone, a horse, or another man, say. The last two of these three points may seem to be part of the first of them; but, as we have seen, it is possible at least for the imagination to make a division. Note that when I use the word "person" here, I use it in the sense in which it occurs in "offences against the person". At this point people will betray how deeply they are infected by dualism, they will say: "You are using 'person' in the sense of 'body'" – and what *they* mean by "body" is something that is still there when someone is dead. But that is to misunderstand "offences against the person". None such can be committed against a corpse. 'The person' is a living human body.

There is a real question: with what object is my consciousness of action, posture and movement, and are my intentions connected in such fashion that *that* object must be standing up if I have the thought that I am standing up and my thought is true? And there is an answer to that: it is this object here.

"I am this thing here" is, then, a real proposition, but not a proposition of identity. It means: this thing here is the thing, the person (in the 'offences against the person' sense) of whose action *this* idea of action is an idea, of whose movements *these* ideas of movement are ideas, of whose posture *this* idea of posture is the idea. And also, of which *these* intended actions, if carried out, will be the actions.

I have from time to time such thoughts as "I am sitting", "I am writing",

"I am going to stay still", "I twitched". There is the question: in happenings, events, etc. concerning what object are these verified or falsified? The answer is ordinarily easy to give because I can observe, and can point to, my body; I can also feel one part of it with another. "This body is my body" then means "My idea that I am standing up is verified by this body, if it is standing up". And so on. But observation does not show me which body is the one. Nothing shows me that.<sup>7</sup>

If I were in that condition of 'sensory deprivation', I could not have the thought "this object", "this body" – there would be nothing for "this" to latch on to. But that is not to say I could not still have the ideas of actions, motion, etc. For these ideas are not extracts from sensory observation. If I do have them under sensory deprivation, I shall perhaps *believe* that there is such a body. But the possibility will perhaps strike me that there is none. That is, the possibility that there is then nothing that I am.

If "I" were a name, it would have to be a name for something with this sort of connection with this body, not an extra-ordinary name for this body. Not a name for this body because sensory deprivation and even loss of consciousness of posture, etc., is not loss of *I*. (That, at least, is how one would have to put it, treating "I" as a name.)

But "I" is not a name: these I-thoughts are examples of reflective consciousness of states, actions, motions, etc., not of an object I mean by "I", but of this body. These I-thoughts (allow me to pause and think some!) . . . are unmediated conceptions (knowledge or belief, true or false) of states, motions, etc., of this object here, about which I can find out (if I don't know it) that it is E.A. About which I did learn that it is a human being.

The I-thoughts *now* that have *this* connection with E.A. are I-thoughts on the part of the same human being as the I-thoughts that had that connection twenty years ago. No problem of the continuity or reidentification of 'the I' can arise. There is no such thing. There is E.A., who, like other humans, has such thoughts as these. And who probably learned to have them through learning to say what she had done, was doing, etc. – an amazing feat of imitation.

Discontinuity of 'self-feeling', dissociation from the self-feeling or self-image one had before, although one still has memories – such a thing is of course possible. And so perhaps is a loss of self-feeling altogether. What this 'self-feeling' is is no doubt of psychological interest. The more normal state is the absence of such discontinuity, dissociation and loss. That absence can therefore be called the possession of 'self-feeling': I record my suspicion that this is identifiable rather by consideration of the abnormal than the normal case.

Self-knowledge is knowledge of the object that one is, of the human animal that one is. 'Introspection' is but one contributory method. It is a rather doubtful one, as it may consist rather in the elaboration of a self-image than in noting facts about oneself.

<sup>7</sup> Professor Føllesdal and Mr Guttenplan tell me that there is some likeness between what I say and what Spinoza says. I am grateful for the observation; but cannot say I understand Spinoza.

If the principle of human rational life in E.A. is a soul (which perhaps can survive E.A., perhaps again animate E.A.) *that* is not the reference of "I". Nor is it what I am. I am E.A. and shall exist only as long as E.A. exists. But, to repeat, "I am E.A." is not an identity proposition.

It will have been noticeable that the I-thoughts I've been considering have been only those relating to actions, postures, movements and intentions. Not, for example, such thoughts as "I have a headache", "I am thinking about thinking", "I see a variety of colours", "I hope, fear, love, envy, desire", and so on. My way is the opposite of Descartes'. These are the very propositions he would have considered, and the others were a difficulty for him. But what were most difficult for him are most easy for me.

Let me repeat what I said before. I have thoughts like "I am standing", "I jumped". It is, I said, a significant question: "In happenings, events, etc., concerning what object are these verified or falsified?" – and the answer was: "this one". The reason why I take only thoughts of actions, postures, movements and intended actions is that only those thoughts both are unmediated, non-observational, and also are descriptions (e.g. "standing") which are directly verifiable or falsifiable about the person of E.A. Anyone, including myself, can look and see whether that person is standing.

That question "In happenings, events, etc., concerning what object are these verified or falsified?" could indeed be raised about the other, the Cartesianly preferred, thoughts. I should contend that the true answer would be "if in any happenings, events, etc., then in ones concerning this object" – namely the person of E.A. But the description of the happenings, etc., would not be just the same as the description of the thought. I mean the thought "I am standing" is verified by the fact that this person here is *standing*, falsified if she is not. This identity of description is entirely missing for, say, the thought "I see a variety of colours". Of course you may say, if you like, that this is verified if this person here sees a variety of colours, but the question is, what is it for it to be so verified? The Cartesianly preferred thoughts all have this same character, of being far removed in their descriptions from the descriptions of the proceedings, etc., of a person in which they might be verified. And also, there might not be any. And also, even when there are any, the thoughts are not thoughts of such proceedings, as the thought of standing is the thought of a posture. I cannot offer an investigation of these questions here. I only want to indicate why I go after the particular "I"-thoughts that I do, in explaining the meaning of "I am E.A." This may suffice to show why I think the Cartesianly-preferred thoughts are not the ones to investigate if one wants to understand "I" philosophically.

Suppose – as is possible – that there were no distinct first-person expression, no pronoun "I", not even any first-person inflection of verbs. Everyone uses his own name as we use "I". (Children sometimes do this.) Thus a man's own name takes the place of "I" in this supposed language. What then? Won't his own name still be a name? Surely it will! He will be using what is syntactically *and* semantically a name. That is, it is semantically

a name in other people's mouths. But it will not be so in his mouth, it will not signify like a name in his utterances.

If I used "E.A." like that, and had no first-person inflections of verbs and no such words as "I", I should be in a difficulty to frame the proposition corresponding to my present proposition: "I am E.A." The nearest I could get would be, for example, "E.A. is the object E.A." That is, "E.A. is the object referred to by people who identify something as E.A."

There is a mistake that it is very easy to make here. It is that of supposing that the difference of self-consciousness, the difference I have tried to bring before your minds as that between "I"-users and "A"-users, is a private experience. That there is this asymmetry about "I": for the hearer or reader it is in principle no different from "A"; for the speaker or thinker, the "I"-saying subject, it is different. Now this is not so: the difference between "I"-users and "A"-users would be perceptible to observers. To bring this out, consider the following story from William James. James, who insisted (rightly, if I am right) that consciousness is quite distinct from self-consciousness, reproduces an instructive letter from a friend: "We were driving . . . in a wagonette; the door flew open and X, alias 'Baldy', fell out on the road. We pulled up at once, and then he said 'Did anyone fall out?' or 'Who fell out?' – I don't exactly remember the words. When told that Baldy fell out he said 'Did Baldy fall out? Poor Baldy!'"<sup>8</sup>

If we met people who were A-users and had no other way of speaking of themselves, we would notice it quite quickly, just as his companions noticed what was wrong with Baldy. It was not that he used his own name. That came afterwards. What instigated someone to give information to him in the form "Baldy fell out" was, I suppose, that his behaviour already showed the lapse of self-consciousness, as James called it. He had just fallen out of the carriage, he was conscious, and he had the idea that someone had fallen out of the carriage – or he knew that someone had, but wondered who! That was the indication of how things were with him.

Even if they had spoken a language without the word "I", even if they had had one without any first-person inflexion,<sup>9</sup> but everybody used his own name in his expressions of self-consciousness, even so, Baldy's conduct would have had just the same significance. It wasn't that he used "Baldy" and not "I" in what he said. It was that his thought of the happening, falling out of the carriage, was one for which he looked for a subject, his grasp of it one which required a subject. And that could be explained even if we didn't have "I" or distinct first-person inflexions. He did not have what I call 'unmediated agent-or-patient conceptions of actions, happenings and states'. These conceptions are subjectless. That is, they do not involve the connection of what is understood by a predicate with a distinctly conceived subject. The (deeply rooted) grammatical illusion of a subject is what generates all the errors which we have been considering.

<sup>8</sup> *Principles of Psychology*, II (London 1901), p. 273 n.

<sup>9</sup> In Latin we have "ambulo" = "I walk". There is no subject-term. There is no need of one.

### 3 Substance

The raising of certain difficulties about the notion of substance belongs especially to the British Empiricist – that is to say our – tradition. We can see a starting-point for them in Descartes' considerations about the wax in the Second Meditation. Descartes concluded there that it was by an act of purely intellectual perception that we judge the existence of such a thing as *this wax* – a doctrine the meaning of which is obscure.

Let me sketch at least some of the troubles that have been felt on this subject. First, there is the idea of the individual object. What sort of idea is that, and how got? This individual object is the same – 'persists' as we say – through many changes in its sensible properties, or sensible appearances; what is the individual itself all this time? Second, supposing that question should be answered in the case in hand by 'It is wax', is it not one objection to this answer that it gives a *general* term "wax" as an answer to the question "What is the individual?" Surely we wanted to know: what is the individual thing *qua* individual, in its individuality? And this cannot be answered by giving a predicate which not merely logically can be true of many individuals, but does actually fail to mark out this one from others. Next, even accepting this answer: "It is wax", what can being wax be except: being white and solid at such and such temperatures, melting at such and such temperatures . . . etc., etc.? Are not the ideas of kinds of substances given by more or less arbitrary lists chosen from the properties found by experience to go together? In that case, the general idea 'wax' will be equivalent to the chosen list; and the particular, individual, parcel of wax is at any time the sum of its sensible appearances. Any other notion of substance surely commits us on the one hand to unknowable real essences, and on the other to an unintelligible 'bare particular' which underlies the appearances and is the subject of predication but just for that reason can't *in itself* be characterized by any predicates. This picture of the appearances or the properties as a sort of clothes reminds one of Butler's lines about Prime Matter:

He had first matter seen undrest;  
He took her naked, all alone,  
Before one rag of form was on.

The picture of substance is too unacceptable, so following Russell we must speak of 'bundles of qualities' or following Ayer of 'totalities of appearances' which are unified not by their relation to some further entity but by their own interrelations. We would rather not admit anything so doubtful as that