

1

Historical Background

§ 1. When we ask about *I* and the thinking it expresses, we are commonly told that the topic is straightforward: a simple rule gives all that needs to be said, '*I* refers to the speaker or writer.'¹ If we complain that this ignores signing and typing, that it wrongly attributes reference to the expression itself rather than its uses (which one thing is such that *it* is '*the* referent of *I*?'), and that it is inapplicable to sessions of sweet silent thought, we may be offered something more discriminating: for example,

The term *I* is governed by a simple rule: any use of *I* refers to whoever produced it. To characterise the meaning of the term, this is all that is needed.²

Like a clock-face, this plain front hides a mass of intricate workings whose construction and design represent substantial achievements. For there are very few other referring expressions of which it might plausibly be said that their meaning is given by a simple rule. Indeed, *Now* and *Here* may exhaust the group. So the prevailing view presupposes an apparatus, a background system made up out of logical distinctions between different kinds of referring expression, together with the principles justifying them.

Such systematicity as we now have begun to be introduced in the nineteenth and twentieth centuries. Previously, it seemed possible to effect considerable progress in philosophy without the mechanism to which the current standard conception appeals. When Descartes, for example, made the proposition 'I am thinking, therefore I exist' the first principle of his philosophy, describing it as 'so firm and secure that all the most extravagant suppositions of the sceptics were not capable of overthrowing it', he did not trouble to say what *I* means, how it may be used to express what kinds of thoughts, what kind of referring expression it is, or indeed whether it should be considered a referring expression at all.³

This insouciance would have been acceptable had Descartes's principle been cogent and foundational no matter how these questions are to be answered. But this is evidently not the case. For he assumes that *I* refers to a self, and his second step is precisely to infer something about that self. If 'I am thinking' is true, then that to which *I* refers must be a thinking thing. But suppose that, as some profess

¹ e.g. Kaplan (1989: 505).

² Campbell (1994: 73).

³ *Discourse*, iv, XXX.

to believe, *I* merely *appears* to be a referring expression, like the expletive use of *It* in 'It is raining' which (unlike the referring use in 'It is a rain-drop') does not purport to pick anything out. Then 'I am thinking' only really means something like 'there's thinking going on', and there is no self to draw conclusions about. The crucial step from Descartes's first principle to every other is missing.

The absence of apparatus for specifying the meaning of *I* caused confusion to ramify exponentially in the course of the next century. For Descartes's casualness permits several interpretations of his first principle, and each such reading triggers different objections to the Cartesian programme, each of which in turn admits of different interpretations. The sets of 'Objections' by Antoine Arnauld (the Fourth) and Pierre Gassendi (the Fifth), when coupled with Descartes's 'Replies', are examples of this deadly ramification. But perhaps the best instance springs from Lichtenberg's addition to the debate. For at least three (mutually inconsistent) readings of his well-known anti-Cartesian remark are now prominent in philosophy:

'It thinks'; that is what one should say, like 'it is raining'. To say *Cogito* is already saying too much—if one translates it as 'I think'.⁴

On the first interpretation, Lichtenberg is claiming that *I* is a referring expression, one that picks out a substantial self, a 'somewhat'. He is protesting that *Cogito* adds one thought too many to those available to the Cartesian meditator. Given what this person has just methodically doubted away in the First Meditation, he is not entitled to assert anything about states of affairs involving a substantial self in the Second. A very different reading makes Lichtenberg assert that *I* is an expletive, or at least that the sentence 'I am thinking' is parallel in grammar and synonymous with the expletive sentences 'It is thought' or 'There's thinking going on'. He is then protesting against translating *Cogito* as 'I am thinking' because until *I* is recognized for what it is, no more a referring expression than the expletive use of *There* or *It*, philosophers (like Descartes himself) will continue to assume that it refers and look for its referent. According to the third interpretation, Lichtenberg accepts that *I* is a grammatical subject and that it does not function as an expletive. He is protesting that this feature of *I* is not sufficient of itself to justify the claim that the term stands for or identifies anything at all, let alone a substantial self. The justification for postulating a substantial self to which *I* refers may be forthcoming. But, *pace* Descartes, not from the mere fact that *I* is used as a grammatical subject.⁵

The point is that, without contextual apparatus for making the relevant concepts distinct and sharp, Lichtenberg's remark is simply not robust enough

⁴ 1796: 412.

⁵ For the first interpretation, see Williams (1978: 95); for the second, see Wittgenstein (1933: 100–1) and Katz (1990: 169); for the third, see Shoemaker (1963: 10) and Parfit (1984: 224–5).

to sustain any one interpretation. All are valid readings, which is as good as saying that none are.

TO WHAT DOES *I* REFER?

§ 2. Nevertheless some specificity attaches to the writings of earlier authors, if only because they argued about what kinds of thing *I* refers to, and a crucial task for any adequate account is to address precisely this question. By noting the various answers provided by foremost writers, we gain clues as to how they regarded the meaning of *I*, and thus retrospectively discover whatever precision and order lie implicit in their views.

Descartes evidently assumed that *I* cannot but be a referring expression, at least in the broad sense that its uses stand for or pick out some particular object (the self). He argued that what its uses single out are particular substances on the grounds that thoughts are acts, uses of *I* refer to the subject of the thoughts they express, and 'the subject of any act can be understood only in terms of a substance'.⁶ To claim that *I* refers to substances is not very precise since he acknowledged that many kinds of item belong to the class of substances, including God and matter. We can only make sense of his precision on this point if we equip him with a further assumption about *I*: that by the very meaning of the term, any use of *I* is guaranteed to have something, and the right something, to refer to. This is the doctrine we have called *The Guarantee*. This assumption enabled Descartes to narrow the field considerably. As he makes clear in the Second and Sixth Meditations, uses of *I* must refer to a substance that is entirely distinct from the body and other material objects, simple, indivisible, indestructible, immortal, and numerically identical through time.

Locke and Berkeley also assumed that uses of *I* single out objects by referring to them. But both complicate the way in which this occurs, demanding a more intricate role for *I* and thus a more complex notion of its meaning. It is Locke's linguistic sensitivity and choice of referent that complicate. With Berkeley, it is his choice of means.

Locke evidently thought that (due perhaps to superficial misunderstandings) *I* occasionally refers to 'the Man only', by which he means, roughly speaking, the human being one is. On most occasions, Locke uses the term to refer to what he calls a *Person*. A thinking item in which ideas are interrelated but which is neither a human being nor any other substance. It is something *annexed* to a substance. So Locke cannot be fitted into, and thus wrecks, William James's oft-repeated classification of historical theories of the self as *Substantialist*, *Associationist*, and *Transcendentalist*.⁷ *I* must be a peculiar term if we are to suppose its meaning

⁶ Descartes (1984: ii. 124).

⁷ Locke (1689: ii. xxvii. 20 ff., pp. 335–6). James (1891: i. 371).

sufficiently flexible to fulfil both referring tasks Locke assigns it and to pick out the kind of item he identifies as its referent.

Berkeley demands that *I* be viewed as a term which one can use to single oneself out, not by some simple means like perceiving oneself, but only by reflecting on one's agency. This might be acceptable if our agency is understood to include the usual abilities to manipulate our environment. But he limits the extent of our agency severely, allowing us little more than the ability to will ideas into existence by the imagination.⁸ If we are to suppose *I* competent to single out a referent in these straitened circumstances, then again, it must be a most unusual term. But no explanation of its peculiarities is forthcoming from these writers, nor could it be without the proper apparatus.

§ 3. Hume rejected all such views, arguing that the referent of uses of *I* cannot be 'something simple and continued . . . and individual'. If it were, it would have to be observable as such when one introspects. But he professes himself unable to reason with anyone who supposes that might occur. When *he* introspects, he observes only perceptions.⁹

Hume makes at least two assumptions about *I* and its capabilities here. First, it must be a term which one can use to express thoughts about an item (oneself) without the need to identify what is being referred to. This is the *Independence* doctrine. It tells us what must be true of *I* if it is to have an expressive use. That is, if the term is to be used to say something of the individual one is thinking about. These are cases where the individual needs to know that the individual spoken about is the same as the individual thought about. According to the doctrine, in using *I*, one can know that the individual thought of is the same as the individual spoken about without having to make an identity judgement.

Hume's second assumption is that *I* must be capable of referring to whatever unites a particular series of perceptions, determining them precisely as mine. And what is that? Relations of causation and resemblance cannot serve if, as he supposed, they provide no real connections themselves. There must be some prior explanation of how a set of experiences comes to be a set, of how discrete entities can be observed as causing and resembling each other in the first place. His inability to find a solution to this problem notoriously caused his own hopes to vanish. But there is a deeper set of questions that would have survived discovery of such a solution. What kind of referring term must *I* be, what meaning must it have, that it fulfil the task set for it: to refer independently of identification, and to stand for logical constructions made up out of causal connections between different perceptions?

Kant diverges more radically, focusing his attention on the use of *I* in the phrase *I Think*, which he considered 'the sole text' from which rational psychology develops its wisdom. This *I* which thinks is not to be considered as 'in the world' but rather as a phenomenal item, for it cannot be intuited or conceived

⁸ Berkeley (*3rd Dialogue*, 115–17; *Principles*, § 28).

⁹ Hume (1739: 252, 633; 252).

as either a material object with a spatio-temporal location, or a merely spatio-temporal object, or even a purely temporal one. But upon its existence depends the possibility of self-ascribing experiences, itself a precondition of experience under any conception which we can make intelligible to ourselves. What most clearly sets Kant apart, however, is his realization that taking any such position demands giving an account of that term, *I*, which expresses it. Called on to supply such an account, however, he can only really say what kind of term *I* is *not*. Its use depends on neither an intuition nor a concept, and it does not ‘distinguish a particular object’.¹⁰ Such are the obstacles arising.

It has been suggested that in his analysis of this use of *I*, Kant espoused the *Independence* doctrine: that there is no need to identify what *I* refers to in order to express thoughts using the term.¹¹ This is at least to suppose that the use of *I* in *I Think* refers. Yet he steadfastly denies that *I* fulfils the task required of a referring expression: to say which one thing is being spoken about. It may be said to ‘represent’ [*vorstellen*] or ‘designate’ [*bezeichnen*], but only ‘a thing of indeterminate signification’, ‘the mere form of consciousness’. It is, indeed, ‘a completely empty expression’.¹² Is the use of *I* in *I Think* then to be regarded as blank (having no function at all), or mere vocalization (a function outside structured language, like a grunt), or an expletive (like the semantically redundant ‘it’ in ‘it is raining’), or eliminable (contributing nothing that could not be expressed fully by other terms)? He leaves us with little clue.

The fundamental problem, which this last quandary reflects most graphically, is now familiar. Without some background organizing conception, we cannot license the assertion that *I* is or is not a referring expression. For without it, we are not able to specify with suitable precision what it is to be, or indeed *not* to be, such a thing. Nor can we license the assertion that *I* is the kind of term which could support any of the positions on offer. For without such apparatus, we have not begun to determine what kinds of term there are and into which category *I* properly falls.

IS *I* A NAME?

§ 4. It is because the now standard view conforms to an apparatus determining these issues that its simplicity may be said to hide immense philosophical effort. Two steps were crucial for progress.

The first was to explain the distinction between terms that are genuine singular referring expressions and those that are not, though they may appear to be. Compare two utterances, one of which contains a genuine singular referring

¹⁰ Kant (1787: A341–405; B399–432; A96–130; B129–69; A382; A 346).

¹¹ Strawson (1966: 165–9); Brook (1994); McDowell (1994: Lecture V).

¹² Kant (1787: A382; B407; B103; A382; A 436).

expression ('Abelard is hot') and one which does not ('It is raining'). Whether or not the first utterance is true depends on how it is with some particular. Namely, the one indicated by the meaning of the subject-term (in this case, a particular person). The second utterance contains no expression whose meaning indicates a particular, and it does not need to contain such a term to be true or false. So, in a move that is prominent and fruitful in Frege, a genuine singular referring expression comes to be characterized as a linguistic counter whose meaning indicates which one particular thing is relevant to the truth-value of the sentence containing it. Uses of such terms help determine the meaning of sentences in which they are embedded by pointing out what one thing, if the whole utterance is to be true, the predicate must be true of. It is an expression we use to indicate (and in communication to anticipate questions about) what, who, or which one individual we are thinking of or speaking about. Unlike plural expressions, uses of such terms take singular verb-forms.

I seems to conform to this model. Suppose both you and I say 'I am hot' but in fact I am cold (perhaps I am lying) and you alone are hot. Then one utterance of that same sentence is true and the other false. So the truth-value of each utterance depends on how it is with a particular individual. Namely, or so it seems plausible to say, the one indicated by the meaning of *I*. When used to issue such statements, then, sentences embedding *I* alter in truth-value depending on how it is with this indicated individual. Indeed, any statements of the form 'I am *F*' are true only when, but whenever, the one so indicated is *F*.

We can be misled by the grammatical name 'First Person Plural' into supposing that *I* is not necessarily a singular term. How can it be if it has this plural form? Grammarians sometimes assume that *I*/*We* are two forms of the same term that differ only in number. For example, when they describe *I*/*We* as morphologically unrelated number forms, in contrast to the typical regular formation of noun plurals—e.g. *boat/boats*.¹³ But if *We* were indeed the plural form of *I*, then sentences of the form 'We are *F*' would mean 'I and I and I ... are *F*' when evidently they mean that I and at least one other are *F* (typically people one would otherwise refer to as *You* or *He/She*). The speaker with the appearance of a man possessed by demons in Mark's Gospel who explained, 'My name is legion for we are many,' was using *We* to speak of what, at the time of utterance, he took to be himself and each other devil associated with him, not for more than one 'himself'. I and I and I do not amount to *We* but I, just as Abelard and Abelard and Abelard do not amount to more than Abelard. *I* no more has a plural form than does *Silver* or the number *One*. We are not just routinely fortunate when we correctly attend to exactly one individual on hearing the first person used. Perhaps no one ever seriously asserted that, qua plural of *I*, *We* can be analysed as 'I and I and I ...'. The point is still worth making. For that is how *We* would have to be analysed if the term were indeed the 'First Person Plural', the plural form of *I*.

¹³ e.g. Greenbaum et al. (1990).

Most theorists have been persuaded by these and like considerations that *I* is a genuine singular referring term. The evidence certainly seems sufficient to make this the 'default position'. By this, I mean the stance we would need to be argued out of rather than into; the position whose truth we are entitled to assume, unless and until there is sufficient counter-evidence to suppose otherwise. But the doctrines of *Independence* and *The Guarantee* present counter-evidence. The details will occupy us later, but the challenge can be expressed immediately to capture its intuitive appeal. Genuine singular terms single out the items to which they refer. So if *I* were such a term, one would have to identify that item to which it refers in order to express thoughts using it. Moreover, failing to single out that item from all other and like things would be a risk one would have to overcome. But, according to the doctrines, one's expressive use of *I* requires no such identification and courts no such failure. So there is a puzzle here awaiting those who try to combine the claim that *I* refers with adoption of the doctrines. Since this group includes those who advocate what it is now conventional to say about *I*'s meaning, a resolution for what is otherwise a deep inconsistency is urgently required.

Strawson's comment that '*I* can be used without criteria of subject-identity and yet refer to a subject because, even in such a use, the links with those criteria are not severed' has convinced some that a simple resolution exists.¹⁴ The idea seems to be that, no matter what *I* may or may not have to do when using *I*, *other people* still have to identify the item to which my use refers, and *other people* still have to overcome the risk of failing to do so. This is certainly compatible with the doctrines, which make no mention of others. It is less obviously congruent with what we plausibly require of a genuine singular term. For the fact that other people must try to identify the referent of my use of *I* if they want to understand it is in danger of seeming curiously irrelevant. It is not, after all, necessary that other people understand my use of *I* if I am nevertheless to use it. And if what I use *is* a genuine referring expression, then the problem is not resolved. How *can* what I use be a genuine referring expression if my use does not depend on successful identification of its referent? Moreover Strawson's manoeuvre confuses semantic with pragmatic issues. It may be true that uses of *I* have a referring role even if they are used without criteria of subject-identity in virtue of the fact that their audience applies empirical criteria of personal identity precisely to single out a referent. But this is an appeal to what an audience requires to understand a speaker's utterance, i.e. to issues of communication, of pragmatics. Whereas it is evidently a semantic issue whether or not uses of *I* have a referring role.

Wittgenstein is often said to have denied that *I* is a referring term. Since his interventions in debate on the meaning of *I* will play a significant role in what follows, it is worth establishing certain historical and interpretative points immediately.

¹⁴ Strawson (1966: 165); see also (1959: 99–100).

Commentators who find anti-referentialism in the early Wittgenstein focus their argument on one aspect of the *Tractatus* project: the apparent attempt systematically to reduce every occurrence of 'I believe that p' to the assertion of 'p'.¹⁵ This is to argue from redundancy to anti-referentialism. For if *I* can be eliminated in this way, so it is said, it must surely be redundant. Now the very definition of 'referring term' requires that a linguistic counter of this sort have a set of tasks to carry out: to single out individuals, forestall questions about which individual is in question, and so on. So if *I* is a referring term, it cannot be redundant. But if this were Wittgenstein's argument, it would hardly be persuasive. What is true for the logical analysis of present-tense belief-sentences need not hold for the analysis of non-belief-sentences, nor even for belief-sentences in other tenses. And it is notoriously difficult to see how his 'I *Vp*' \rightarrow *p*' schema (where *V* holds place for a verb) may be used to reduce sentences like 'I made a promise.' In short, the fact that *I* may be eliminable from a small sub-set of cases does not show it is redundant overall. Moreover, there is ample evidence that the early Wittgenstein regarded *I* as neither eliminable nor redundant. He continued to employ the term throughout the *Tractatus*, in both pronominal and substantive uses, and with particular reference to the 'metaphysical subject', an entity described in Kantian-Schopenhauerian terms as presupposed by the existence of the world. His purpose in analysing certain present-tense uses of *I* was certainly destructive, but of views about what *I* referred to, not of the view that *I* referred. He was rejecting the claim that *I* refers to a simple, enduring, empirically encounterable self.¹⁶

Many commentators have claimed that Wittgenstein held anti-referentialist views in various different forms and for various different reasons as his later views developed. But his strongest assertions commit him to nothing stronger than an ignorance theory. This is true not just of his published claims ('*I* does not name [*benennt*] a person'), but of his unpublished thoughts also ('*I* does not designate [*bezeichnet*] a person'). In brief: *I* may or may not refer. If it does refer, it is unclear what it refers to, though certainly not to a person.¹⁷

So why has the later Wittgenstein been misrepresented as an anti-referentialist? One explanation may be that those most closely associated with his programme themselves moved from an ignorance theory to anti-referentialism. Though differences of nuance separate these philosophers, all seem to have found this background thought compelling: it is so wholly unclear what *I* might refer to, surely it cannot refer at all. So Peter Geach has claimed that there is nothing for which *I* stands. And Elizabeth Anscombe has argued that *I* is 'neither a name nor another kind of expression whose logical role is to make a reference, *at all*'.¹⁸

¹⁵ Wittgenstein (1916: 118; 1921: v. 542; v. 421).

¹⁶ Wittgenstein (1921: v. 542; v. 5421; v. 63; v. 632; 1916: 79–80). For the contrary view, see Hacker (1993: 213–14) and Glock and Hacker (1996: 160).

¹⁷ Wittgenstein (1958: §410; 1936: 228). *Bezeichnet* is misleadingly translated 'refer to' by Glock and Hacker (1996: 95), who identify him as anti-referentialist.

¹⁸ Geach (1957: 117–20). Anscombe (1975: 154). See also Malcolm (1979) and Kenny (1989).

Another explanation may be that Wittgenstein himself is being mildly ironic. Perhaps his remarks are really to be interpreted as anti-referentialist. For it has generally been assumed that persons are the best candidates as referents for *I*. So if the term does not refer to a person, what else could it refer to? The question would expect the answer ‘nothing’, of course. But its rhetorical force is considerably weakened by the availability of at least one venerable answer. A Lockean might reply, ‘If not to a person, then *I* refers to a human being.’ For on their view, first, human beings and persons are not identical. And, second, usage provides evidence that we can use *I* to single out the human being *rather than* the associated person.

§ 5. So we have an account of what a singular referring expression is, one that give us reason, *prima facie*, to include *I* in the class. The second crucial step towards making *I*’s meaning and role tractable was to explain distinctions between different kinds of referring expression. This meant appreciating the subtly different conditions and means by which various kinds of expression fulfil the referring task—names; pronouns; descriptions; demonstratives; indexicals; and so on.

The grammatical characterization of *I* as a personal pronoun may cause one to suppose that it replaces or substitutes for proper names, and consequently should be grouped with them. Though tempting, the inference is hardly compelling, since personal pronouns also go proxy for whole noun phrases (as in: ‘The Parisian philosopher asked his young student whether *she* liked *him*’). Nevertheless, it is reasonable to hypothesize that, within the class of referring expressions, *I* belongs to the class of names. Mill and Frege certainly described it as such, though it should be admitted that in their hands the classification is not much sharper than ‘referring expression’. For the former specifically included in the set common and abstract nouns, descriptions, and adjectives, while the latter included any expression whose logical and semantic role it is to stand for a given object.¹⁹

Frege’s view of names is considerably sharpened by his notion of ‘sense’. He argued that such terms are to be associated not just with the object denoted, but with the sensitivity of a rational subject’s capacity for recognizing that object. Definite descriptions, demonstratives, and indexical expressions can be made to fall under the category of names, so conceived, with varying degrees of plausibility. *I* is a name, but of a special sort, at least in certain of its uses: it is the name which reflects the ‘particular and primitive way’ in which ‘everyone is presented to himself’ and ‘to no one else’. These particular uses are characteristically employed privately, in soliloquy. Frege argues that the public and communicative use of *I* is associated with a wholly different sense.²⁰

If *I* is a name, it must be proper rather than common, for it refers to individuals rather than kinds of individuals (like *City*; *Vegetable*; *Planet*). It is more akin to personal proper names like *Abelard* or *Heloise* than those proper names requiring an article (*The Reichstag*, *The Mediterranean*). But it was soon noted that there are significant differences between *I* and personal proper names. Every person can use

¹⁹ Mill (1843: I). Frege (1891; 1892*a*: 42–53).

²⁰ Frege (1892*b*; 1918: 24–6).

I and refer to himself, while only those called *Abelard* can use that name and refer to themselves. Every person who uses *I* can only refer to himself, while only those called *Abelard* can use that name and only refer to themselves. Every person who knows which proper name is at issue when someone says 'Abelard' can know which particular person is referred to without knowing certain facts about the context—e.g. who is speaking. No person who knows which singular term is at issue when someone says 'I' knows which particular person is referred to without knowing such facts. And whereas the context of proper names may be used to disambiguate which name is at issue (e.g. *Abelard of Paris* or *Abelard Jones*), the context of *I* is only ever used to determine which object the expression refers to.²¹

As is often noted, there is deep significance to these features of *I*: they give the term a unique role in manifesting self-reflexive self-consciousness.²² One may have a private name for oneself, *NN*. But this name will be unlike *I* in that no one else need even have a private name, let alone the name *NN*. Conversely, one has a public name by which one may archly refer to oneself. *Abelard* himself might say, 'I see *Abelard* has spilt the wine again.' But this name will be unlike *I* in that it may be used to refer to others also. Moreover, one cannot say, 'I am the *F* who is *G*' ('I am the fool who is spilling the wine'), without recognizing the direct bearing of the information conveyed by the statement on *oneself*. Not so with any name. *Abelard* may forget, for whatever reason, that his name is *Abelard* and therefore that the name refers to him. He may then say, '*Abelard* is the fool who is spilling the wine,' without entertaining as a possibility what simply could not fail to occur to him had he said, 'I am the fool who is spilling the wine'—i.e. that the information he is broadcasting is precisely information about *himself*.²³

These features of *I* give it the character of a device with varying referents. Such terms are often called 'one-off' or 'unrepeatable' devices in the literature. This is misleading because it suggests that the same individual could only once be the referent of uses of *I*, *You*, *He/She* and so on. In fact, of course, the same individual may be the referent of different uses of these terms regularly enough. The point is just that of no one individual is it true that it is *invariably* the referent. (Compare: bad weather in England is not constant. But nor is it a one-off, being regular enough. It is, rather, *variable*.) So I shall call such terms 'variant devices'. The contrast is with proper names, like *Abelard*. Adopting for now the usual convenient simplifying fiction that there is a single such name with which a single individual was baptized, *Abelard* invariably refers to the same individual. *I*, on the other hand, is without a constant referent. In this respect, it is like *You*, *He/She*,

²¹ For one way to draw this distinction between pre-semantic and semantic uses of context, see Perry (1997: 593).

²² Perry (1979).

²³ The gradual recognition and delineation of these features can be traced through Frege (1918: 24–6); Russell (1912: 27–8; 1914: 164–5); McTaggart (1927: 62–86); Strawson (1950); Shoemaker (1963); Castañeda (1966; 1968). Recent appreciation of their full significance is due particularly to Anscombe (1975); Perry (1979); Kripke (1972); and Kaplan (1989).

This, That, Now, and Here. These terms may pick out the same person or time or place regularly enough. But there is no single person, object, time, or place such that *it* is the referent *whenever* one of these terms are used.

These features distance *I* from proper names and associate it with the other singular personal pronouns. This dissociation may be recognized in different ways. So Saul Kripke explicitly regards *I* as a rigid designator, for example. But he does not claim that it is a proper name.²⁴ And there is good reason to reject such a view within his theory of reference. For what counts as reference-relevant context in this causal theory, and the role such context plays, differ markedly as between *I* and proper names like *Abelard*. The reference-relevant context of a proper name within an agent's idiolect is the causal history which links it to an original dubbing. Thus reliance on testimony replaces Fregean recognitional sensitivity in this part of the theory. By contrast, the context of *I* is its possible occasions of use. Moreover, as noted above, such context is not necessary to disambiguate which term is at issue when *I* is used, unlike proper names such as *Abelard*.

In brief: if *I* is a name, that must be so either in the very general sense that all singular referring terms are, or in some very specific sense that personal proper names like *Abelard* are not.

Russell at one time seized the possibility afforded by the second disjunct, classifying *I* as a *logically proper* name. McTaggart later advanced a similar position with considerable force.²⁵ Both rely on a background dualism: whatever 'things' (as opposed to truths) we can know and refer to, we must either know and refer to 'by acquaintance' or 'by description'. They argued that it is 'by acquaintance' that we know and refer to the self using *I*. This is to suppose that we are directly (non-inferentially) aware of the self in perception, that the self is 'before the mind', and that its existence is logically guaranteed. The self is a simple for which *I* stands. And to be adequate to stand for such a thing in the referring way, *I* must be a logically proper name. That is to say, in particular, a simple symbol, resistant to logical analysis, whose successful reference is logically guaranteed.

So the notion that *I* is a logically proper name requires one familiar doctrine, *The Guarantee*. But it flouts another, *Independence*. For it makes expressing oneself using the term depend on being acquainted with the item to which it refers. And that is a form of identification. So Russell soon renounced the notion, arguing for the only alternative consistent with his background dualism: that *I* is a descriptive term.

IS *I* A DESCRIPTIVE TERM?

§ 6. The claim that *I* is a descriptive term answers one question and raises another. For it tells us what kind of expression *I* is, but we want to know what uses

²⁴ Kripke (1980: 10 n. 2).

²⁵ Russell (1912: 27–8); McTaggart (1927: 62–86).

of the term mean. And in the case of descriptive terms (unlike proper names, probably), that requires saying what description *means the same as* the term in question. Evidently *I* cannot be an *indefinite* descriptive term—an expression to be understood in terms of the existential quantifier ('there is *at least* one thing which . . .'). For on each occasion of use it refers uniquely ('there is *exactly* one thing which . . .'). So if *I* is a descriptive term, we must ask 'what definite description, *the* ϕ , is such that it means the same as—is synonymous with—any use of *I*?' And it is this aspect of descriptivism about *I* that has attracted most attention. Russell suggested 'the subject of the present experience' and Reichenbach 'the person who utters this token'.²⁶ The basic position continues to appeal though the suggested descriptions change. So Peacocke offers 'The subject who has *these* conscious states' as a candidate. Parfit suggests '*this* subject of experience' instead. And, in response to perceived difficulties with these descriptions, Rovane has worked up the more complex candidate: 'the set of rationally related intentional episodes of which *this* one is a member'.²⁷

The suggested descriptions change consistently because each has consistently proved unsatisfactory. So consider only the most recent descriptivist option: Rovane's alleged solution to the problems raised by earlier attempts. What is wrong with her position is that the description she offers as a candidate falls foul of several basic requirements for the analysis of any singular referring expression: it is either circular or solipsistic, and falsifies the motivational role of the term in question.

The first set of considerations relates to the distinction between one's own states and states that are not one's own. Solipsism is a bad thing, as is generally recognized. And avoiding solipsism requires (minimally) that one have a use for this distinction. Now Rovane's description offers us such a use, but at a price she herself cannot afford. For it leaves nothing but reference to *myself* with which to distinguish what makes certain states *these* and not *those* (they fall under the category of states that are *mine*, directly experienced by *me*, and so on), or to legitimize combining two or more states as part of the set (it is because *that* state is also *mine* that it is part of the set of which *this* is a member). And if the candidate-description must make ineliminable use of *I* in this way, it cannot *give* the meaning of *I*. The description must instead *assume* it. Thus Rovane's candidate-description puts her position in a constructive dilemma. Either it is solipsistic (and thus immediately problematic), or it incorporates the very expression it is meant to analyse, escaping solipsism only by reneging on its appointed task.

The second set of difficulties with this most recent attempt at a viable descriptivism relates to the fact that, as many have noted, *I* has a crucial motivational role. It makes no sense to suppose that agents might know 'I am *F*' (where, in Perry's well-known examples, *F* stands for 'spilling sugar'; 'late for a meeting';

²⁶ Russell (1914: 163–9); Reichenbach (1947: 284).

²⁷ Peacocke (1983: 109–51). Parfit (1984: 252). Rovane (1998: 214).

‘approached by a bear’) without their recognizing the crucial bearing of that information on themselves. And this first-personal feat of self-conscious self-reference, by which the agent recognizes the subject and object of an intentional episode as identical, partly explains the way they are regularly motivated to do what they do. Now if Rovane’s candidate-description genuinely gives the meaning of *I*, it should at least preserve these crucial features of the term. But evidently it does not. For one might be in a position to know that the set of rationally related intentional episodes of which *this* one is a member is *F* without doing something about it, precisely because one does not know that this set is *oneself*.

Moving from this particular case to general descriptivist strategy in giving the meaning of *I*, it is notable that each proposed description contains a particularizing element: *Present*; *This*; *These*. This is just as well since descriptions without such elements—‘pure’ descriptions, like *The Subject* or *The Speaker*—fail to track uses of *I* in crucial respects. When I use *I*, I always refer to myself and never to you. When you use the very same term, you always refer to yourself and never to me. But I routinely use pure descriptions like *The Subject* or *The Speaker* to refer to you or to me just as you do. Unlike *I*, these pure descriptions need not refer to any one person rather than another on an occasion of use. They refer instead to some general property characteristically instantiated by satisfiers of the description—subjects, speakers, and so on. So it is by adding particularizing elements that descriptivists hope to capture these aspects of *I* with their candidate-descriptions. The underlying point here reflects a general claim (due ultimately to Russell): demonstrative or indexical reference as a whole is irreducible to descriptive reference.²⁸

§ 7. McTaggart and others helped formulate a well-known objection to descriptivism based around the following observation: it is possible for a subject to know that ‘*the ϕ is *F**’ to be true for any description ϕ without knowing what is the case, that ‘*I am *F**’. And if this is so, then *I* cannot be synonymous with any definite description *the ϕ* .²⁹ The observation can be illustrated with Russell’s candidate-description: ‘the subject of the present experience’. You might be on a fairground ride and, seeing one of your group’s reflected faces, know that ‘the subject of the present experience is terrified’, while not knowing (perhaps sincerely denying) that you are terrified and that it is your face you have seen.

There are manoeuvres that a Russellian can make in response and counter-moves to these. Moreover, the candidate-descriptions of other descriptivists like Reichenbach need to be tackled differently. But it is sufficient that the outlines of anti-descriptivist strategy be broadly clear to appreciate one crucial point. Either these strategies are hostage to fortune, or some underpinning reason exists for regarding uses of *I* as irreducible to descriptions. For at present, if we are

²⁸ Russell (1905). See Evans (1982: 51–60; 143–204).

²⁹ McTaggart (1927: 63–75). See also Castañeda (1966; 1968: 160–6); Perry (1979: 167–8); Evans (1982: 206–7).

persuaded of the truth of the observation, it is because all the cases we can think of exemplify it and none are counter-examples. And without some reason that would explain *why* this is so, we cannot say in advance that *no* candidate-description will ever be found that proves a counter-instance.

The obvious solution is to appeal to the doctrines of *Independence* and *The Guarantee*. They would serve as such underpinning reasons if they hold for *I* and do not hold for any candidate-description. And this is precisely the move made by theorists. This suited McTaggart, of course, who regarded *I* as a logically proper name and hence was already primed to suppose that its reference was logically guaranteed. Similarly, Castañeda argued that *The Guarantee* governs our use of *I* to refer to ourselves and does not regulate the use of descriptions. So the use of *I* must be 'ontologically prior' to any descriptive self-reference. Hence the meaning of *I* must be irreducible to any description.³⁰ This would explain why McTaggart's original observation turns out to be true for any case we have thought of in the past or might dream up in the future.

IS I A (PURE) INDEXICAL?

§ 8. The fact that *I* differs from proper names and descriptive terms is now commonly explained by first distinguishing such expressions from the group of so-called 'indexical' terms, and then assigning *I* to that group. What virtue is there in this programme?

The first step we may accept for the sake of argument and maintaining focus. Some question it. But the fact that they do is of much less interest than the reasons they offer for doing so. For these reasons will become a point of focus in Part II. Some claim we should regard proper names as indexical expressions because they are terms which depend for their literal content upon extra-semantic features of the contexts in which they are uttered.³¹ As we shall find, this profoundly misrepresents what it is to be an indexical expression. Logical character and inferential role reveal that the literal content of an indexical is fully dependent on intra-semantic features of the contexts in which it is uttered. That is how we should understand the function of devices like demonstration, for example, which make the referent of a particular use count as such by being salient.

With regard to the overall programme, we have certainly noticed signal differences between *I* and proper names like *Abelard*, and between *I* and definite descriptions like *The Subject of the Present Experience*. But that is the negative point. We need to know whether these differences should lead one positively to

³⁰ McTaggart (1927: 63–75). Castañeda (1968: 160–1). See also Perry (1979: 167–8) who appeals in addition to a third doctrine, *Rule Theory*, which we shall examine below.

³¹ See Pelczar and Rainsbury (1998: particularly 293).

identify *I* as an indexical. And in particular, we need to know what an indexical expression *is*, what makes a term count as such.

Indexicals are often defined as context-sensitive expressions. But what meaningful expression is not? No matter how one thinks words get their meaning—by stipulation; shared experience; convention; application to the way the world is; and so on—they are thus sensitive to *something* counting as their context. It follows that there must be some particular characteristic sense in which referring expressions that count as indexicals may be said to be ‘sensitive to context’. Some identify that sense, when pressed, by saying that the reference of indexical terms *shifts with* context, or is *identified via* context, or is *determined by* context. But proper names (*Abelard*) and definite descriptions (*The Author of the Logic*) satisfy these definitions. We might disagree about how some item counts as the unique referent of these kinds of term. For example: being the individual presented under a certain mode; baptized on a certain occasion; or uniquely satisfying a certain cluster of descriptions or conventions. But the fact that *something* must play this role on each occasion means that the reference of these kinds of term shifts with, is identified via, and is determined by, *something* that counts as context—on whatever view we take of what context is, be it the situation in which the statement is made, or the situation in which it is evaluated.³²

In brief: we might be prepared to define indexicality in such a way that proper names and descriptive terms are included. But the price would be high: we could no longer appeal to indexicality to explain what is special about *I*, what distinguishes it from other kinds of term. And that, recall, was our original purpose.

§ 9. The solution proposed for problems of this sort turns on the claim that indexicals are terms associated with rules. Their reference is governed by them. But the familiar objection returns. For rules can be supplied that govern descriptive terms also (e.g. ‘Any use of *the author of the Logic* refers to *x* if and only if *x* is the author of the *Logic*’), just as we can regulate the use of proper names (e.g. ‘Any use of *Abelard* refers to *x* if and only if *x* is *Abelard*’). So there must be some particular characteristic sense in which the reference of referring expressions that count as indexicals may be said to be ‘governed by rules’.

Rules give the meaning of indexical expressions and determine the reference of their uses in relation to context (their situation of use). Descriptive terms and proper names may be associated with rules, but the *meaning* of neither is a rule. For the former it is a synonym, as we have seen, and for the latter, plausibly, it is the referent. David Kaplan and others who have articulated this response regard such rule-governance as the essence of indexicality. It is not simply that rules give the meaning and determine the reference of indexicals in context. If some term is an indexical, then a rule is *sufficient* to give its meaning, and that rule is *sufficient* to determine its reference. *Now* and *Here* are examples in their central uses. (Cases

³² See Kaplan (1989: 494).

like that in which one points to a map and says *Here* are treated as eccentric.) So expressions like *He/She*, *This (F)* and *That (F)* which behave rather like indexicals—they are variant devices whose reference shifts from context to context—are not ‘Pure Indexicals’. For their reference is determined not simply by the rule which gives their meaning, but also by associated demonstration—a gesture or something of the sort. These *Impure Indexicals* are often called ‘Demonstratives’ or ‘Deictic Terms’.³³

So the history of debate brings us to this question: ‘what sort of indexical is *I*?’ If the distinction between *Pure Indexicals* and *Impure Indexicals* (henceforth *Deictic Terms*) is even roughly in order, to which group does *I* belong? Answering this question is the central task of the essay. On its solution depends the nature of *I* and of the thoughts it expresses—and hence, amongst much else (if Descartes or Kant are correct), the possibility of all our knowledge.

§ 10. *I* is a *Pure Indexical*. A rule in context is sufficient to give the meaning and determine the reference of the term. This has been the prevailing view for some time. Deep differences of approach and position may otherwise divide philosophers. But about this doctrine—call it *Rule Theory*—there is an impressive uniformity. No dissent is apparent in the community here.³⁴

So Sydney Shoemaker claims that *I* is ‘a referring expression whose meaning is given by the rule that it refers to the person who uses it’. According to Barwise and Perry, ‘whenever it is used by a speaker of English, [*I*] stands for, or designates, that person. We think that this is all there is to know about the meaning of *I* in English and that it serves as a paradigm rule for meaning.’ For Christopher Peacocke, ‘the reference rule for *I* fully determines its meaning in English’. John Campbell claims: ‘The term *I* is governed by a simple rule: any token of *I* refers to whoever produced it. To characterise the meaning of the term, this is all that is needed.’ For Kaplan, *I* is the term it is precisely because it is governed by a rule: ‘*I* refers to the speaker or writer.’ In context, the rule is sufficient to determine reference for all tokens of *I*.

Kaplan introduces a small complication: two supplementary rules are required, one to govern indexicality, and the other direct reference. But the caveat is irrelevant to the overall claim. For these additional rules are simply required to determine the way in which *the* rule is to be interpreted: *the* rule is the rule it is because of them. So the supplementary rules operate in an anterior way: they do not determine the reference of tokens of *I* alongside *the* rule. *The* rule, once established, is sufficient for this role. Thus Kaplan is in wholehearted agreement with the prevailing view. A simple rule is sufficient to give the meaning and

³³ See Kaplan (1989: 489–92).

³⁴ For advocacy in addition to that mentioned in the text, see Nozick (1981: 78). Rovane (1987: 147). Mellor (1988: 79). Strawson (1994: 210). O’Brien (1995*a*). Perry (1997: 598). Bermúdez (1998: 9). Soames (2002: 103). Reichenbach evidently thought *I* was a descriptive term *and* one the reference of whose uses is sufficiently determined by the rule (1947: 284).

determine the reference of *I*. Indeed, he is largely responsible for making it the current orthodoxy.³⁵

Unity on this issue is made that much more striking when we recognize the complete disunity about related matters. Even those who deny that *I* is a referring term, and consequently do not regard the rule as having a reference to determine, nevertheless regard the role as giving the meaning of the term.³⁶ But it would be wrong to treat *Rule Theory* as a unique unifier. For the armistice extends to embrace both *Independence* and *The Guarantee*.³⁷ What explains this extension of the truce is the fact that the three doctrines are logically related. Supporters of the doctrines often appeal to the existence and nature of these relations for the mutual support they afford.

Thus Strawson, taking the example of someone thinking ‘I am feeling terrible’, writes:

Anyone who is capable of formulating such a thought will have mastered the ordinary practice of personal reference by the use of personal pronouns; and it is a rule of that practice that the first personal pronoun refers, on each occasion of its use, to whoever then uses it. So the fact that we have, in the case imagined, a user, is sufficient to guarantee the reference, and the correct reference for the use.³⁸

And Campbell makes the point in the other direction in a characteristically understated way: ‘when we leave this rule behind, the datum that reference-failure is impossible is hard to understand.’³⁹

Why should these theorists suppose the doctrines offer each other such strong mutual support? Suppose that uses of *I* are sufficiently determined in context by the rule that it refers on each such occasion to whoever uses it, and it is guaranteed that there is exactly one user for every use. Then it seems that every use must be guaranteed to have a referent (the user), and that every use must express a thought *about* someone *by* that same someone (i.e. the user) without their needing to identify the one as the other. The endorsement is reciprocated, for if *The Guarantee* and *Independence* are true, then the rule must be sufficient in context to determine the reference of uses of *I*. For suppose *Rule Theory* were false and *I* a *Deictic Term* like *He/She*, its reference dependent on an associated demonstration.

³⁵ Shoemaker (1968: 91). Barwise and Perry (1981: 670). Peacocke (1983: 133–9). Campbell (1994: 73). Kaplan (1989: 493; 505; 520).

³⁶ For example, Anscombe (1975: 154).

³⁷ For endorsement of *Independence* (often discussed under ‘immunity to error through misidentification’), see in particular Woods (1968). Shoemaker (1968; 1986). Anscombe (1975). Evans (1982: esp. 179–91; 215–20). Rovane (1987; 1993). Brewer (1992; 1995). Campbell (1994: esp. 121–34). O’Brien (1995*a* and *b*). Cassam (1995). Bermúdez (1998: esp. 6–12). Wittgenstein (1969: 63–70) is a special case as we shall see. For endorsement of *The Guarantee*, see in particular Castañeda (1968: 160–1). Anscombe (1975: 149 and 151–2). Peacocke (1983: 175–9). Shoemaker (1986: 126–7). Rovane (1987: 151–5). Kaplan (1989: 509). Hacker (1993: 223). Strawson (1994: 210–11). Campbell (1994: 125). McDowell (1998). Bermúdez (1998: esp. 9).

³⁸ Strawson (1994: 210).

³⁹ Campbell (1994: 125). See also Bermúdez (1998: 9).

Independence would fail because it would be necessary to identify whichever someone the thought is *about* (i.e. the demonstrated one) with that someone whom the thought is expressed *by* (i.e. the user). And since it is not guaranteed that such a demonstration will accompany each use anyway, *The Guarantee* also would fail.

A crucial effect of the corroboration these three doctrines offer each other is to make the gulf between *I* and the other singular personal pronouns (*You* and *He/She*) unbridgeable. The distinction was originally drawn by reference to *Rule Theory*. It is true of *I* and not of these other terms that its reference is sufficiently determined in context by a simple rule. But *Rule Theory* brings in its train two doctrines which constitute two more ways of marking the cleavage which consolidates the divide and makes it comprehensive. For the reference of *You* and *He/She* is not logically guaranteed, and their use to express thoughts is not independent of identification. Hence, if these doctrines are correct, in every relevant way *I* is marked out by its meaning as radically different.

The triumph of *Rule Theory* should not surprise. In saying what kind of term *I* is, it also gives an explanation of why *I* is not the kind of term we have decided it is *not*—a proper name, for example, or a descriptive term. But there is something more peculiarly satisfying about *Rule Theory*. It is the recently acquired centrepiece of a comprehensive position to which we have seen debate about the first person long tending. For *Rule Theory* satisfies a need that had become increasingly stark: to justify two doctrines which dominated that history but which lacked both grounding and explanation. *Rule Theory* validates *The Guarantee* and *Independence*, and acquires validation by so doing. It seems appropriate to label the comprehensive position at whose heart these doctrines lie *Purism*. For, together, they set out an appealing conception of *I* as purified of the demanding features and requirements which make other terms so complicated. A 'simple rule' gives its meaning. No identification is necessary in the central cases. Each use is logically secured against failure.

So the history of debate has brought us to *Purism*, with its component doctrines strongly bound to each other by ties of reciprocal endorsement. The question remains, however, whether any of the doctrines in this mutual support group deserves our support. And one nagging anomaly looms up like Banquo's ghost, unsettling the triumph with reminders of a suppressed past. Regarding an expression simultaneously as a genuine referring term and as either logically guaranteed against reference-failure or independent of identification is a trick we have yet to master. Unless and until we do, *Purism* remains an unstable position and its three components call for careful examination.