12

English Speakers Should Use "I" to Refer to Themselves

R. M. Sainsbury

I. Self-knowledge and self-reference

Self-knowledge is knowledge of oneself *as* oneself. The point of the qualification is made vivid in Perry's famous example (1979): In knowing that *the person whose shopping cart is trailing sugar is making a mess*, I know something about myself, for I am that shopper; but I do not have self-knowledge in the interesting sense, knowledge of myself *as* myself, since I do not know that *I* am that shopper.

It's useful to divide questions about self-knowledge into two roughly delimited groups, metaphysical questions and epistemic ones. Metaphysical questions center on the nature of self-knowledge: what makes a piece of knowledge *self*-knowledge, as opposed merely to knowledge about a person who is in fact oneself? Epistemic questions relate to the special directness or authority that, many believe, attach to our self-knowledge. Although one expects an intimate connection between the answers to both kinds of question, the early sections of the present paper relate to the metaphysical kind.

One who expresses self-knowledge typically does so by engaging in self-reference: she refers to herself *as* herself. It's not obvious that all self-knowledge involves self-reference. Perhaps there are cases in which a property *simply* presents itself to a subject: hungry, thirsty. What makes the property one which putatively applies to the subject is not that the subject is represented, and represented as having the property. It's the fact that the property is presented to the subject in a particular way, a way that ensures that the presentation of the property counts as accurate only if the property holds of the subject. Here's an analogy: a thermometer reads 85°F. It represents the property of being at 85°F. It does not represent a subject of the property. In fact the subject is the tip of the thermometer's probe. The representation of the property counts as accurate iff the tip is at that temperature. The tip itself, though present, and though essential

Thanks to referees for the Press for helpful comments, and to Michael Tye for discussions about IEM.

to accuracy conditions, is not represented. Likewise one can imagine cases of self-knowledge in which there is no representation of the self, and so no self-reference. Although it seems to me quite plausible that there are such cases, in this paper I set them to one side, focusing on those in which self-knowledge does involve self-reference.

Self-reference is typically achieved by exercising a concept corresponding to the first person pronoun.[1] So it's natural to think that, in José Luis Bermúdez's words, "There is a close connection between the sense of 'I' and a particular form of self-knowledge". In a tradition that goes back at least to Frege, the sense of "I" is explained in terms of a special ego-distinctive content, and this special content is what is metaphysically distinctive of self-knowledge. The positive claim of this paper is that this Fregean view is mistaken. I suggest, by contrast, that self-knowledge is knowledge of something that is publicly available, so the content known will not be distinctive of self-knowledge. What makes knowledge *self*-knowledge is not what is known or its subject matter, but the special relation in which the knower stands to the knowledge, reflected by the fact that only the knower can use the *I*-concept to express it. This concept has a content which can also be expressed (by suitably placed thinkers) using other concepts, for example the *you*-concept, concepts which cannot be used to express self-knowledge. The self-known content can be known by others; it's just that only one subject can *self*-know that content.

This position is motivated by the following reflection: Self-knowledge, as normally understood, is something that people can have only of themselves, but the sense of "I", as used on a specific occasion to express self-knowledge, is common to speakers and hearers, even though, from the perspective of hearers, the word is used to refer to someone other than themselves. If one thinks of self-knowledge as a distinctive sort of thing known, this reflection, in conjunction with some further premises, leads to a contradiction:

1 Let p be an item of self-knowledge for subject x.
2 No one other than x can know that p. [Hypothesis for refutation]
3 x can express p using an "I"-sentence, say s. Thus used, an utterance of s has the content that p.
4 Others can understand an utterance of s by x, and can come to know that it is true.
5 So others can come to know that p.

Premise (4) is contentious from a classical Fregean perspective, for Frege said that the sense of "I" is given by a distinctive way of thinking of oneself that is unavailable to anyone else. If understanding involves sharing ways of thinking, this precludes the understanding of "I"-utterances by others. But many see this as simply a problem for

[1] I take this to hold even of monolingual speakers of languages like Latin, in which a suitably conjugated verb form, rather than a personal pronoun, can be used in acts of self-reference. The relevant conjugation can be regarded as expressing the first person concept.

the Fregean position. The natural view, and the one to be adopted in this paper, is that one can understand the "I"-utterances of others. Once this is granted, the second part of (4), that others can (sometimes) know such utterances to be true, will be problematic only for solipsists. (4) embodies the reflection of the beginning of this paragraph: "I" is a word in a public, shared language, and so its meaning is common to speakers and hearers. If you express self-knowledge by using "I", and I understand you and believe you, I end up knowing about you that which you know about yourself; but what I know is knowledge of you, and so (for me) is not self-knowledge. This places limitations on the extent to which an account of "I" can throw light on self-knowledge: the knowledge it is used to express may be self-knowledge for one person, but not self-knowledge for another.

Bermúdez offers an account of self-reference that explicitly recognizes this limitation, and attempts to finesse it. Thinking of oneself as oneself involves thinking of oneself "as an object uniquely located in space". Likewise, a proper, "practical", understanding of a use of "I" requires that one locate (or at least have the capacity to locate) the utterer in space. The demands upon speaker and hearer are in one respect the same: each must be able to locate the speaker. In another respect the demands are different, for the speaker must be able to locate himself whereas the hearer must be able to locate someone other than himself, namely, the speaker. The difference does justice to the fact that one who understands an expression of self-knowledge by another cannot thereby acquire self-knowledge. The similarity does justice to the fact that speaker and hearer both have full access to a single and unambiguous sense of the first person pronoun employed by the speaker, so that a hearer can come to know what a speaker who expresses self-knowledge expresses: you can share your self-knowledge with me, for whom it is not self-knowledge but knowledge of you.

There is much to admire both in the way that Bermúdez sets up the issues and in the subtle way he handles specific points. In the end, however, I find his own proposal implausible, for reasons I set out in the two sections that follow. In §§IV–VI I develop an alternative account and indicate some consequences.

II. Bermúdez on locational abilities

According to Bermúdez, the ability to know the speaker's location can be the common element in an account both of her capacity for self-reference and of another's capacity for understanding. Bermúdez's starting point is Evans's "Russell's principle": thinking about a particular object involves having discriminating knowledge of it, "a capacity to distinguish the object of his judgment from all other things" (Evans 1982: 89). In the special case of a subject thinking of himself, Evans explains discriminating knowledge in terms of three components: an information component, an action component, and a location component. Bermúdez suggests that the third component by itself will provide a condition that's necessary and sufficient for understanding a use of "I", whether the understanding is on the part of the user or on the part of the hearer.

A hearer can bring this understanding to bear in deploying a token of "you": I appreciate that, when you utter "I am hot", what you said is that *you* (yourself) are hot; all being well, this token of "you" will have the same token sense as your token of "I". This applies what Bermúdez calls the symmetry constraint.

The idea that an ability to locate the subject explains both use and understanding of the first person pronoun is initially implausible. As Bermúdez recognizes, one may have no idea where one is when correctly exercising the first-person concept, so straightforward locational knowledge is not necessary for first-person understanding of "I". One may have no idea where an utterer of an "I"-sentence is (she is speaking on her cell phone, but you don't know so much as which country she is calling from) without this damaging one's ability to understand the utterance, so straightforward locational knowledge is not necessary for third-person understanding of "I". Finally, the insufficiency of straightforward locational knowledge for understanding may be suggested by such examples as writing in the sand: confronted with an inscription of "I am hot", I know precisely where the utterer was at the time of the utterance, but there's room to doubt whether this aids understanding (this case is discussed in more detail towards the end of §IV).

Bermúdez does not identify the required ability with the straightforward locational knowledge envisaged in the previous paragraph. The ability he considers is weaker, and this helps with problems for providing a necessary condition, though it exacerbates problems for providing a sufficient one.

The weaker conception of locational knowledge he offers is more hypothetical, and so less demanding:

> One needs to be able to identify the utterer . . . to know who uttered the sentence in question. But in what does this knowledge consist? It is not necessary to track down the physical source of the token, since it is frequently enough to know what one would have to do to track down the physical source of the token. (Bermúdez 2005: 189)

The contrast is between *tracking down* and *knowing what one would have to do to track down*. To illustrate, Bermúdez invites us to consider a situation in which I hear some "I"-sentence uttered by an unfamiliar voice in the next room. Bermúdez takes it for granted that one can understand the utterance, and this has some intuitive plausibility. Perhaps I cannot track down the physical source of the utterance (I am a prisoner in chains), so the strong necessary condition for understanding is not met. But, Bermúdez suggests, the weaker condition, of knowing what one would have to do to track down the physical source, is met: I know that "I need simply go next door". I'm not sure how seriously one should take this. If there are several people next door, I don't see how going there would enable me to "track down the physical source", so I don't see how I could know, in this case, that I need simply go next door. Likewise if the utterer had promptly left; and so on. In this case, it seems that the "I"-utterance can be understood even by one who can formulate no very specific plan for locating the utterer. So even

Bermúdez's weaker condition, as I understand it (in terms of being able to formulate a plan for locating) is not necessary.

The condition can be weakened still further. In some sense, we have an entirely general ability to locate people in space. No doubt we can understand "I"-utterances only if that ability is on tap. But if "being on tap" is not understood in terms of some further requirement, for example, as in the preceding paragraph, in terms of formulating a plan for locating, the ability being on tap places no interesting constraints on the conditions under which full understanding is possible. Then there would indeed be no counterexamples to the proposed condition, regarded as necessary, but only for the uninteresting reason that it would always obtain; by the same token, it would be unlikely to provide a sufficient condition.

Can the situation be improved by appealing to a distinction between full and various degrees of partial understanding? Perhaps the entirely general ability to locate people in space provides a necessary and sufficient condition for a minimal degree of understanding; higher degrees of understanding are correlative with more demanding conceptions of identification, where the degree to which the conception is demanding is correlative with how specific the ability is to the situation (more demanding, more specific). Bermúdez deals with an apparent counterexample to the necessity of a locational criterion by some such appeal: "Telephone conversations with people one does not know proceed perfectly well when one is in complete ignorance of the location of the other person". He suggests we should not rush to identify "proceeding perfectly well" with the attainment of full understanding; we must recall that understanding comes in degrees (Bermúdez 2005: 189 n. 15). The example is not fully developed and does not come from his most recent paper on this topic. But perhaps we could understand it as offering a case in which, though ignorant of location, I have a better than purely general ability to discover location, and so, on the proposed degree-theoretic version of his account, I have a more than minimal, though less than total, degree of understanding. My ability goes beyond the purely general ability since I could, for example, ask my interlocutor where she is located. But it falls short of a fully specific ability, which might perhaps be marked by having the speaker before my eyes.

It's unclear that this suggestion accomplishes much, since it seems that, at best, the distinction of degrees of understanding marks nothing other than the degree of specificity of the identificatory ability. Certainly, other markers of understanding, including ones endorsed by Bermúdez, are in place independently of degrees. For example, he suggests, in my view quite correctly, that "the canonical way of understanding a token of 'I' involves being able to put oneself in a position relative to the speaker such that one can think about him in second person terms…" (Bermúdez 2005: 189). One can certainly do this when speaking to the unfamiliar caller from an unknown location: "You should put me on your 'no-call' list", I can say; intuitively, I know perfectly well whom I am addressing, and my remark is in keeping with having perfectly understood their "I"-utterances. Likewise I can give reports of what was said by the caller that are entirely comparable to reports of the clearest and most uncontroversial cases in which "I"-utterances are understood. Compare, for example:

> A stranger called and said she was authorized to allocate me a free Mercedes.
>
> A stranger I met in the supermarket said she was authorized to allocate me a free Mercedes.

In both cases, my anaphorically dependent "she" corresponds to my successful interpretation of an "I"-utterance on the part of my interlocutor. Yet the first case is one in which the only identificatory ability is rather non-specific (or must at least be so counted by Bermúdez, if we take seriously his most recently cited words), and so my understanding should be awarded at best a medium degree; whereas the second case is one of perceptual identification, which is being taken (provisionally – see §IV below) as marking the most specific kind of identificatory ability, which means my understanding was of the highest degree. Yet intuitively there is no difference. Different degrees of understanding are artifacts of the theory, and not data in need of theoretical explanation.

Let's try another revision of Bermúdez's condition. In his telephone example, the hearer has no prior knowledge of the caller. But suppose the caller is my wife, whose current location I do not know. In fact, I know that she is keen to keep her location secret, and so will not reveal it if asked, and has used a complex routing system to make the call untraceable. This ensures that my ability to locate her is at best of the most general kind: I would be at a loss to formulate a plan for finding her. This has no impact at all on the fact that I can fully understand her "I"-sentences, which again shows that locational abilities at least sometimes have no bearing on understanding. However, the case is one in which another cognitive achievement is clearly in place: I know who my wife is. This suggests that the knowledge needed to use or understand an "I"-sentence, or to use or understand a correlative "you"-sentence (one which reports back to the utterer of an "I"-sentence what she said) is simply knowledge of who the utterer is. This indeed is Bermúdez's first suggestion in the displayed quotation above, and it is a very natural one. In the light of it, one could see locational ability as an attempt, not successful in general, to give a helpful gloss on what *knowing who* involves: knowing who is knowing where. Removing the locational gloss, we end up with the following claim: assuming the remainder of an "I"-utterance is understood, one understands the token of "I", whether as speaker or hearer, iff one knows to whom self-reference is made; this same knowledge will govern uses of other pronouns in reports of such speech. The suggestion entails that one does not understand the inscription in the sand despite knowing the utterer's location at the time of utterance and having resources to improve on that locational information, that one does understand one's wife's cell phone "I"-utterances despite not knowing where she is, and that one can understand one's own "I"-utterances while having no idea where one is and no means to improve on one's ignorance. It seems that we can endorse at least the overall character of Bermúdez's position, even if not the details: identification is essential, though actual or potential knowledge of location is only loosely related to identification. In the next section, I suggest that identification is not essential: the picture centered on *knowing who* is at best misleading.

III. Knowing who does not help explain the sense of "I"

Everyone agrees that English speakers should use "I" just to refer to themselves as themselves. More controversially, I'll suggest that this is all one needs to know to understand "I", whether as a user of it or as a hearer. If this is sufficient, then no independent requirement of knowing-who is necessary. I'll first illustrate this, and then show that the knowing-who requirement has problems of detail.

The following is a mickey-mouse presentation of how a hearer might move from identification of an "I"-utterance to a "she"-report of what was said.

1 She uttered the words "I am hot".
2 In so doing, she applied the rule for "I".
3 So she used "I" to refer to herself as herself.
4 She used "is hot" to ascribe hotness to the referent of "I", as she referred to it.
5 So she said that she herself was hot.[2]

Line (2) takes for granted that the speaker used English in conformity with its rules, and line (3) is an application of the general rule (to use "I" to refer to oneself as oneself) to the specific case. Line (4) is an application of a general rule for using subject–predicate sentences: one ascribes what the predicate says to what the subject-expression refers to, as it was referred to.[3]

The derivation leads to a good report in indirect speech of what was said, so, for Fregean reasons Bermúdez and I both subscribe to, one who can make the derivation can understand the utterance. Yet the derivation made no explicit use of an identification of the utterer. This suggests that identification is not a requirement for understanding.

It's natural to object that the premises presuppose identification of the utterer. Does not the very first "She" require precisely that? True, if "She" is understood deictically (as would be natural), then in some sense identification is involved. However, the derivation works just as well if the initial "She" in (1) is replaced by "Someone" (or by "Some female"). Then the hearer can achieve a correct report without identifying the speaker. The subsequent occurrences of "she" are anaphoric, and so make no further

[2] Although the "herself" in "she herself" would be fussy and unnecessary in many contexts, if the speaker is Suzy and her words were "Suzy is hot" it might be acceptable to report her as having said that she is hot, but unacceptable to report her as having said that she herself is hot.

[3] This gestures to the non-extensional character of semantic rules. An utterance of "Hesperus is visible" ascribes visibility to Hesperus, as referred to in the utterance; but not to Phosphorus, as referred to in the utterance, for nothing was referred to in the utterance as Phosphorus.

demand for identification. This reinforces the suggestion that identification is not required for understanding.

Bermúdez himself, to the extent that he accepts the symmetry constraint,[4] would seem also to be committed to the view that identification plays no significant part in understanding. Suppose one encounters an inscription in the sand of the words "I am hot". There's some inclination to say that one does not understand the token inscription in the sand (though of course one understands the type "I am hot"), but it seems that the inclination must be resisted by someone who accepts the symmetry constraint (as I understand it). Arguably, one can fully accurately report what is said by the inscription in the sand as follows: Someone, writing in the sand, said that (s)he was hot. The symmetry constraint entails that, if this is a correct and complete report, the token sense of "(s)he" is identical with the token sense of the sand-inscribed "I". Since the former does not require that one know who the utterer was (it's just a pronoun anaphoric upon an indefinite noun phrase), nor does the latter. So a token of "I" can be fully understood without the understander knowing, in any substantive way, who the utterer was.

In cases like this, there is a non-substantive or trivial sense in which one knows who made the inscription: the inscriber, or someone with rather small feet, or someone who handled a stick rather skillfully. But in this sense, a sense according to which one knows who the utterer is if one can give any correct answer to the question "Who is the utterer?", the requirement of identification or knowing-who places no constraints at all. The answer "The utterer" is always correct, however inadequate it may be.

This brings us to the heart of the reason why identification (knowing who) is not part of a proper account of understanding "I"-utterances: it is too sensitive to context, and sensitive in the wrong kind of way, to be an appropriate basis for a semantic achievement. Let's illustrate this with some examples.

Case 1: The woman in front of you at the meeting says in a loud voice "I'm hot". You nudge your neighbor:

> Who is that?
> You mean you don't know?
> No, I've no idea who it is.

Your last claim is sincere and true: you really have no idea who the person in front of you is. Yet it cannot be doubted that you understood the utterance, and can give an impeccable report: the woman in front of me said that she was hot.

Case 2: The letter contains the utterance "I need your help". There's a signature, Dora White, and an address. You have no idea who she is. It's one of a series of begging letters out of the blue, all signed Dora White. Intuitively, you understand the utterance. She has said she needs your help.

[4] I accept the constraint myself, but will not argue for it here. See Sainsbury (1998).

Case 3: I hear the words "I am hot" in the next room. I don't have identifying knowledge of who uttered them, but intuitively I understand them: the speaker said that she was hot.

In all three cases, it seems clear that we understand, but various different kinds of knowledge-who are absent. This shows that none of these kinds of knowledge-who is in general necessary to understanding.

In Case 1, we understand even though we "don't know who the speaker was" in the sense that we can't link our perceptual identification of her with prior knowledge (the link might be effected by naming the speaker: "It's Madonna!"). A whole range of variants can be considered. For example, two men go behind a screen, and are told to toss a coin and on that basis to determine which sits on the left and which on the right. A voice from the left utters "I am hot". I have auditory perceptual identification (I know that whichever was on the left said that he was hot), but I don't know which man was on the left, so in some sense I don't know who the utterer was. Yet this seems to me clearly a case of understanding, despite lack of some knowledge-who.

In Case 2, we understand even though there is no perceptual identifying knowledge. Hence perceptual identifying knowledge is not in general necessary for understanding. In some sense we do "know who the speaker was": we know her name. If someone came to the door and established she was Dora White, it would be natural to say: "I know who you are: the person who wrote those letters".

In Case 3 (from Bermúdez 2005), I lack both kinds of identifying knowledge of the utterer considered in the previous two examples, but Bermúdez says, and I agree, that this does not prevent me from understanding what was said.

Here's a dilemma for those wishing to say that some form of *knowledge-who* is necessary in order to grasp first person utterances. If the kind of *knowledge-who* is specific, I can find clear cases of understanding in which it is absent. Thus Case 1 refutes the idea that what's required is perceptual identification, Case 2 refutes the idea that what's required is name-based identification, and Case 3 refutes the idea that what's required is a disjunction of these. If it is pretty unspecific, there are no cases in which it is absent, so it's an idle constraint.

IV. The rule for "I"

I claim that there's no more to understanding a token of "I", whether as speaker or hearer, than being able to apply to the token the rule: English speakers should use "I" to refer to themselves as themselves.[5] That's why the derivation in §III worked as

[5] Bermúdez says that "I" refers to the utterer. That's a different idea, missing the reflexive element, and clearly does not fully capture the semantics of "I". Both rules require qualification, for it's well known that there are abnormal cases, as in role-playing. Bermúdez gives a particularly nice example: a lawyer who makes "I"-utterances in the course of reading the will of the deceased ("I, being of sound mind, ...") refers to the deceased and not to himself.

well as it did. Some identificatory work is presupposed, but that is identification of an utterance and not, in any substantive sense, identification of an utterer. Since token utterances have their utterers essentially, in a sense we have identified the utterer once we have identified the utterance; this exemplifies the "non-substantive" or trivial notion of identification of the utterer: he or she is identified simply as the utterer.

The claim is motivated in part by the failure of other views, as indicated in earlier sections, and in part by a reconsideration of the symmetry constraint. What a person who has self-knowledge thereby knows can be expressed by another, third-personally. You utter "I am hot", expressing your knowledge of yourself, as yourself, that you are hot. I have just stated what you know (that you yourself are hot) without using the first person pronoun and without expressing knowledge of myself as myself. I've expressed (as we might misleadingly put it) your self-knowledge. Self-knowledge is thus not some distinctive thing known, but a distinctive way of knowing something which others can know in a different way. These facts are perfectly captured in the reflexive rule: English speakers should use "I" to refer to themselves as themselves. This entails that I should use "I" to refer to myself as myself, that you should use "I" to refer to yourself as yourself, and that she should use "I" to refer to herself as herself. The immediate consequences of the rule already reveal that the same content can be accessed from different perspectives. When I know you have self-referred, I know you have done something I cannot do: you have used "I" to refer to yourself as yourself. The closest I can come to this achievement is to use "you" to refer to you or "I" to refer to myself. With the pronoun appropriately contextualized, neither of these acts falls under the act-type of using "I" to refer to yourself. In my view, there is nothing more than this to the "special way of thinking of oneself".

The notion of an "essentially indexical thought" might wrongly be taken to refer to a specific kind of content. The previous paragraph shows that this is not what's at issue. To self-refer is not to invoke a special ego-distinctive content, but to stand in a special relation to a content available to all: only you can use "I" to refer to yourself, but any of us can use other expressions to refer to you, and to express the very same content as you express.

Self-knowledge is not a distinctive thing known, so is not special at the level of content. What you know when you know that you are hot is just what I know when I know that you are hot. The special feature is the relation in which you stand to the content, a relation to it in which no one else can stand. Only you can express the content using "I"; others need to use other expressive resources. In this there is nothing more mysterious than the fact that if we are far apart your "here" is my "there".

I agree with Bermúdez that the symmetry constraint ensures that some "I" tokens have the same content as some "you" tokens. Whereas he feels the need to elaborate this in terms of a shared identificatory ability, I feel no need for such an elaboration. The identity is, it seems to me, ensured by the uncontentious reflexive rule for "I", whose consequences embody sameness of content accessed from different perspectives: I should use "I" to refer to myself, and you should do just the same, that is, use "I" to refer to yourself. When you do that, I may understand a token of "I" without thinking of myself at all.

Self-knowledge and self-reference are not absolute notions, but depend upon a perspective. What's self-knowledge for you (e.g. that you yourself are hot) is not self-knowledge for me (it's knowledge about you). In my view, it's a mistake to think that the reflexive iteration of a pronoun (as in "I myself", "you yourself", and so on) introduces a distinctive content. Rather, it serves to mark a distinctive and perhaps restricted relation to an ordinary, and open, content. You yourself are hot. That's the open content, accessible by anyone. What's restricted is who can express this content using the first person pronoun: only you can do that. And only someone else, someone who is addressing you, can express this content using "you". The reflexive helps keep track of these symmetric relations to content, rather than contributing a distinctive content.

No content is distinctive of self-knowledge. Suppose Suzy was hot but didn't notice. Then she comes to realize it. What is the *it*? That she is hot. Something can be known at a time when it is not self-known, and the very same thing can later become an item of self-knowledge. When it is known but not self-known, there is no appropriate reflexive mode of expressing it: she's hot all right, but "*she herself* is hot" makes sense only in some special context, e.g. "She doesn't realize that she herself is hot".

Consider a content of the form "S is F". Does the fact that S knows this content, in which S himself is referred to, ensure that S has self-knowledge? I think not. Suppose Suzy is hot and knows that Suzy is hot. Perhaps she knows this because, not knowing her name is "Suzy" or that the person pointed out to her in the mirror as Suzy is herself, she attains testimonial knowledge that Suzy is hot without knowing that she herself is hot. The same could be said of a deictic use of "she". There are limits on how to refer to someone so as to express a content which that person does, or may come to, self-know. The limits are captured in the rule for self-reference, and the consequences that flow from it through the adjustment in reflexive pronouns to perspective: English speakers should use "I" to refer to themselves as themselves. So you should use "I" to refer to yourself; and so on. Likewise I may use "you" to refer to you, and I need to do this (or use a third person reflexive) if I'm to express your self-knowledge.

There is a content Perry's shopper did not know, but it is not some special content only available to him. It's simply the content *that he was making a mess*, available to us, and which I have just expressed. If he had thought it (and there were no defeating circumstances) he would have had self-knowledge. What he didn't know was not something only he could know. We, reading this paragraph, know the very thing of which he was ignorant, namely, that he was making a mess, and we do not thereby have self-knowledge.

V. Two objections

the explanation of the word "I" as 'the word which each of us uses to speak of himself' is hardly an explanation! (Anscombe 1975: 48)

Anscombe reaches this conclusion on the basis of the fact that someone might refer to themselves yet not realize that this was what they had done; yet it seems that the cited rule does not block this happening, and so it does not capture what is special about the first person.

I have tried to pre-empt this objection by expressing the rule as: English speakers should use "I" to refer to themselves as themselves. But Anscombe objects that the "as themselves" requires explanation. I agree; and once the explanation is given, it will emerge that the qualification "as themselves" is not needed in an adequate statement of the rule.

We need to distinguish between an act that consists in referring to oneself, and an application of the rule: use "I" to refer to yourself. Every act in the second category also belongs to the first, but the converse does not hold. As Anscombe stresses, and as we have seen in discussing Perry's examples and others, one can refer to oneself without realizing that one has done so. However, applying the rule just quoted precludes referring to oneself in the ignorant way. To obey the rule explicitly, you need to think the rule, that is, think some thought you might express thus: I must use "I" to refer to myself. To obey the rule implicitly, you need to behave as if you were obeying it explicitly. Either way, to confront the rule with understanding is to confront yourself as yourself, and to act on the rule is to act on yourself as the rule presents you, that is, as yourself. There is no possibility of an act being an application of the rule, yet your being in no position to realize that you have applied "I" to yourself. That's what it is to refer to yourself *as yourself*. Accordingly, it makes no difference whether we formulate the rule for "I" with or without the final qualification "as yourself". If you use the rule to guide your performance, you will not merely refer to yourself with "I", but refer to yourself as yourself. We'll see in §VI that this has an epistemic consequence.

The second objection I wish to consider is this: Won't it follow from what's been said that I can understand the inscription in the sand? I can produce the impeccable reflexive report: someone inscribed these words in the sand, thereby expressing that she herself or he himself was hot. And is not this consequence enough to refute the view? There are two unclarities. One is what the right thing is to think about the example. The other is what the present view commits to.

Although at first sight one may suppose that I don't understand the inscription, further reflection may suggest that I do understand it. Consider a conversation that evening back at the hotel. A stranger says "Weird weather today. I wrote my response to it in the sand." It seems to me appropriate to respond: "I saw it: you wrote that you were hot." This suggests I had already understood the inscription.

It is slightly unclear what the present view says about whether or not the inscription is understood. That's because the present view requires that an understander identify the token utterance, but has so far said nothing about what that involves. An understander must see an identified token of "I" as exemplifying the rule that English speakers should use "I" to refer to themselves. But what is it to identify a token of "I"?

One view is that simply having the marks on the sand before my eyes is enough for me to identify the token utterance. Adding that view to this paper's main claim delivers the result that I can understand the utterance. Personally I don't find this very counterintuitive, though that may be because I think that it's quite easy to exaggerate what understanding, regarded as a purely semantic achievement, can deliver. Those who feel that this is definitely not a case of understanding may wish to accommodate their position to the present view by saying that an utterance token is an event, with a definite temporal span, that identifying such a token requires an at least approximate knowledge of when it occurred, and that merely seeing the marks on the sand gives me insufficient information about the utterance's temporal position.

One wishing to justify this more demanding view might make these suggestions. First, we are perhaps more disposed to say we understand the inscription when it is on a part of the beach washed by the tide (and so not more than eight hours old) than when it is on a part of the beach that is not; an explanation would be that we are more disposed to attribute understanding when we are closer to pinning down the time of the utterance. Secondly, we are possibly much less disposed to say we understand when we encounter a slip of paper, blown by the wind, of completely unknown origin and date, on which are written the words "I am hot"; an explanation would be that we are not even close to identifying the token utterance. I myself don't know what to say about these cases, and this line of thought is not part of what I wish to advance. As far as my main point goes, one can adopt either a more or a less demanding account of token utterance identification. For my main point is that, whatever the appropriate standards for identifying an utterance-token, understanding an "I"-utterance is a matter of identifying it and seeing it as an application of the general rule that English speakers should use "I" to refer to themselves. We can forget about any special modes of identifying or locating utterers, and we can forget about self-knowledge as a distinctive kind of fact known.

VI. An application to the epistemology of self-knowledge

Some cases of self-knowledge are said to be "immune to error through misidentification" (IEM). This phrase can be used of more than one property. But it's reasonable to link it to the following thought: I can't use the first person pronoun (in the normal way) yet erroneously refer to someone or something other than myself. There are, of course, abnormal projective or deferential cases, as already mentioned (see note 5 above). But setting these aside, a limited immunity to error seems to follow from the envisaged rule for "I". Someone trying to apply that in the normal way already needs to think of herself as herself since, in the fully explicit case, the rule presents itself in this way: "I must apply 'I' to myself". Once the rule is thus presented, the possibility of erroneous application is removed, since the rule-user must already have thought of herself as herself.

One might regard this consequence as a refutation of whatever leads to it, for is it not well known, and indeed obvious, that one can misidentify oneself? One can think of an arm as one's arm when it is not (you see it among a tangle of limbs after an accident). One can think of a person (seen in a mirror) as having a property one would be loathe to suppose one possessed oneself, even though the person is oneself. One can think one is Napoleon. Or one could be deluded by too much Zen talk of union, and think of oneself as the universe as a whole. So is not any view which has the consequence that one cannot misapply "I" obviously incorrect? More generally, "refer" is a success verb. But how could anything guarantee, beyond metaphysical possibility of failure, that a certain kind of attempt, the attempt to refer to oneself using the first person, would succeed?

I think the errors of self-identification mentioned in the previous paragraph are to be understood as the subject successfully referring to himself, but then making a faulty identification, rather than as cases in which the subject, in using the first person, succeeds in referring to someone or something other than himself. In the examples, the subject falsely thinks of himself that he is the person whose arm that is; that he is not the person in the mirror; that he is Napoleon; that he is the universe as a whole. In each case, he successfully refers to himself, but then makes a mistake about who or what *he himself* is. For example, it's not that when he thinks "I am decreasing in entropy" the subject refers to the universe rather than to himself. He refers to himself, and falsely believing himself to be the universe, believes his entropy is decreasing.

I accept the general principle that what can succeed can fail. So if referring to oneself constitutes success, there needs to be some possibility of failure. Such failure would be constituted by failing to understand or exploit the rule that one is to use "I" to refer to oneself. Grasp of this rule incorporates all that is required for its successful application; indeed, the grasp already involves the application, for it requires one to think of oneself as oneself. That's why one who knowingly applies the rule cannot fail to refer to himself. That's consistent with an attempt to apply the rule failing, through some failure to understand the rule. In the overall success, there is room for failure, but the failure does not consist in failure to apply a rule that is understood, but rather in a failure to understand the rule.

Can we rebut Anscombe's criticism of this position? She agrees that "Getting hold of the wrong object is excluded" but continues: "that makes us think that getting hold of the right object is guaranteed" (1975: 59). She thinks this last thought is incorrect; and that leads to her notorious claim that "I" is not a referring expression. The argument seems to run as follows:

1 If "I" was a referring expression, it would have to have a referent.
2 Its referent would be either a human body or else a Cartesian Ego.
3 The referent cannot be a human body, for I can refer to myself even in a state of sensory deprivation, when I can doubt that *I* have a body.
4 The referent cannot be a Cartesian Ego because (a) this cannot do justice to identity between distinct *I*-thoughts and (b) there are many unresolved disputes about what the Ego would be like and how it would function in thought.
5 "this is the solution: 'I' is neither a name nor another kind of expression whose logical role is to make a reference, *at all*" (1975: 60).

The argument raises several difficulties, but I will focus on one of them. Suppose that, for independent reasons, it was right to say that people are their bodies. Then a person who referred to herself would be referring to her body. It does not follow that she would be referring to herself as her body, and so it does not follow that there would be no room for her to doubt that *she* has a body. So (3) is false. Moreover, the dubious metaphysical claim at (2) need have no bearing on how a subject who uses "I" conceives of himself. Minimally, he conceives of himself as himself; that's consistent with his also having all sorts of other conceptions of himself. This has no bearing on whether "I" is a referring expression. Even if we accept (2), theses (3) and (4) are not even the kinds of theses that could follow.

Self-reference is assured of success; in some sense, it's immune to error. Clearly it does not follow that what is thought, in an act involving self-reference, is immune to error, or indeed has any further epistemologically interesting features. If self-knowledge has such features, they will need to be found elsewhere than in the nature of self-reference.

References

Anscombe, Elizabeth (1975), "The First Person", in S. Guttenplan (ed.), *Mind and Language* (Oxford: Clarendon Press), 45–65.

Bermúdez, José Luis (2005), "Evans and the Sense of 'I' ", in Bermúdez (ed.), *Thought, Reference and Experience: Themes from the Philosophy of Gareth Evans* (Oxford: Clarendon Press), 164–94.

——(this volume) "Self-Knowledge and the Sense of 'I' ".

Evans, G. (1982), *The Varieties of Reference* (Oxford: Clarendon Press).

Perry, J. (1979), "The Problem of the Essential Indexical", *Noûs* 13: 3–21.

Sainsbury, R. M. (1998), "Indexicals and Reported Speech", *Proceedings of the British Academy* 95: 45–69; repr. in *Departing from Frege* (London: Routledge, 2002), 137–58.