

## **FIRST-PERSON THOUGHT AND THE USE OF 'I'**

### 1 Introduction

Many theorists take the following claims to be true of 'I':

- [A] An utterance of 'I' always refers to the utterer (U).
- [B] Every utterance of 'I' will have a referent.
- [C] The intended and the actual referent of each utterance of 'I' are the same.

These claims play a crucial role in the traditional account of first-person thought (hereafter, 'TA'). According to some recent work in linguistics, however, they are false. My purpose in this paper will be to determine the impact of this work on TA.

### 2 The traditional account of first-person thought

TA argues that first-person thoughts involve a special kind of self-conception: the conception is formed in such a way that it always successfully 'latches onto' some worldly entity, and this entity is always the thinker. Different versions of this central thesis have been put forward.<sup>1</sup> TA reaches this conclusion about first-person thought on the basis of claims made about the first-person pronoun. I will begin by briefly outlining TA, explaining how claims [A] – [C] figure in it.

---

<sup>1</sup> See, e.g., Bermúdez (1998), Evans (1982), and Shoemaker (1968).

It is generally agreed that to have a first-person thought is to think about oneself in a distinctive way in which one cannot think about anything else. Since such thoughts are about oneself, they involve a self-conception. One can, however, have thoughts about oneself that are not first-personal. The task facing a theory of first-person thought is therefore to provide an account of what is distinctive about the self-conception involved in first-person thinking. To do this, some means of identifying which thoughts are first-personal is required. One way to identify thoughts is by using the language that expresses them. TA adopts this strategy. Examples like Perry's messy shopper suggest that first-person thoughts – and only first-person thoughts – are expressed in English using 'I'. The messy shopper follows a trail of sugar on the supermarket floor, meaning to inform the person with the burst sugar bag that he or she is making a mess. Eventually the shopper realises that it is *his* bag of sugar that has burst and stops to adjust it. Before and after this realisation, the shopper thinks "*a* [where *a* is a concept that picks out the shopper] is making a mess". But the thought the shopper has after his realisation is a first-person thought whereas the thought he has before is not. It seems natural for the shopper to express his second thought, but not his first, using 'I':

(1) I am making a mess.

Cases like that of the messy shopper lead proponents of TA to identify first-person thoughts on the basis of this claim:

[I-t] First-person thoughts are all and only those that can be expressed in English using 'I'.

An utterance (a use) of a sentence expresses a thought when the proposition encoded by that utterance is identical to the propositional content of the thought. Notice that the truth of [I-t] depends in part on the truth of [A] – the claim that an utterance of 'I' always refers to U. According to [I-t], the fact that I can express a thought of mine by uttering an 'I'-sentence *guarantees* that it is a thought about me. For this to be the case, any 'I'-sentence I utter must be about me. This is ensured by the truth of [A], which states that an utterance of 'I' always refers to U. Hence [I-t] depends on [A].

TA adopts a broadly Fregean account of propositions which takes them to be composed of something like concepts, rather than worldly individuals. Propositions are individuated by their cognitive impact, which is a function of their conceptual components. When someone expresses a first-person thought by uttering an 'I'-sentence, the propositional content of his thought is identical with the proposition encoded by his utterance. It follows that the self-conception involved in his first-person thought, and the conception associated with his use of 'I' on this occasion are one and the same. [I-t] states that first-

person thoughts are all and only those that can be expressed in English using 'I'. TA holds, therefore, that the only conception associated with the use of 'I' is the self-conception involved in first-person thinking (although this will be a different conception for each 'I'-user). It follows that an examination of how 'I' behaves – i.e., an examination of its cognitive significance – will reveal facts about the self-conception involved in first-person thinking. TA takes the behaviour of 'I' to be described by claim [B], which states that every utterance of 'I' will have a referent, and claim [C], which states that for each utterance of 'I', the intended and the actual referent will be the same. TA understands these claims as having implications for the basis upon which I form the self-conception involved in my first-person thoughts. It is taken to be such that a conception formed on it will always pick out some worldly entity (this is taken to follow from [B]), and this entity will always be me (this is taken to follow from [C]).

### 3 A counterexample to [A]

Veronique is about to start her new job, but there is a transport strike and she will arrive in England two weeks late. Knowing that her new colleagues will try to contact Veronique, Steve – the secretary – records the message, "I am not here yet", on the answering machine in Veronique's office. He types the message into the machine, and it is read out by a computerised voice whenever someone calls. Veronique knows nothing about the message. When Steve types the

message into the machine, we can understand him as arranging to make an utterance(s) at a later time – when a caller hears the message.<sup>2</sup> Someone rings and hears the message. We have an utterance,

(2) I am not here yet.

It would be completely implausible to hold that *Veronique* is the utterer/user (U) since she did not type the message, or even arrange for it to be typed on her behalf. Instead, Steve is U, since he types the message into the machine, and so arranges for the utterance(s) to be made. Nevertheless, it seems that *Veronique* – not Steve – is the referent of 'I' in (3). The person calling the office will understand on the basis of (3) that *Veronique* is not there yet. Furthermore, it is *Veronique's* whereabouts that determine the truth or falsity of (3) – it will be true just so long as *Veronique* is not in her office at the time when the message is heard by a caller. Thus we have a counterexample to the claim that a token of 'I' always refers to U.<sup>3</sup>

---

<sup>2</sup> This proposal is due to Sidelle (1991).

<sup>3</sup> See Corazza et al. (2002), and Romdenh-Romluc (2006) for more counterexamples to [A].

#### 4 An initial response

TA may respond to cases such as that of Veronique in the following way. TA identifies which thoughts are first-personal by using [I-t], which claims that first-person thoughts are all and only those that can be expressed in English using 'I'. [I-t] also allows TA to draw conclusions about first-person thinking from claims about 'I'. [I-t] presupposes [A] – the claim that an utterance of 'I' always refers to U. Since [A] is false, [I-t] must be rejected. An obvious response TA might make at this point is this. Examples like the messy shopper show that there is some connection between 'I'-use and first-person thinking. Given that my first-person thoughts are about me, one might suppose that when I express a thought using 'I' to refer to myself, the thought I express is first-personal. Hence, one might endorse the following thesis:

[I-t\*] First-person thoughts are all and only those that could be expressed in English by the thinker, with a use of 'I' that refers to U, i.e., to the thinker.

Clearly, whether or not a thought is so expressible should be taken to depend on the content of the thought, not the ability of the thinker to speak English. [I-t\*] can be used to pick out first-person thoughts, and providing that [B] and [C] are true of those uses of 'I' that refer to U, TA can take these claims to explain what is distinctive about the self-conception involved in our first-person thinking as before. However,

the counterexamples to [A] cannot be ignored. TA's conception of the way in which 'I' is used and behaves must, therefore, be compatible with a theory that can accommodate the counterexamples to [A].

There are three central proposals for dealing with cases such as that of Veronique. I will first consider the prospects for TA if one of these proposals is accepted. To anticipate somewhat, the prospects for TA are not good. The obvious move for TA is therefore to reject all three proposals and offer an alternative account of how the reference of 'I' is fixed. Space here prevents me from discussing all possible alternatives, however, I will argue that the discussion of the three central proposals strongly suggests that it will not be possible to provide an alternative theory of how 'I' refers that preserves the two referential guarantees for the right cases.

#### 5 Accounts of indexical reference and counterexamples to [A]

The three central proposals for dealing with counterexamples to [A] take Kaplan's basic theory of indexical reference as their starting-point. Kaplan holds that the reference of an indexical is determined by applying a rule of use – i.e., a *character* – to a context (Kaplan 1977). The character for 'I' yields the agent of the context as the referent of this term. Kaplan assumes, however, that the context to which the relevant rule should be applied is always the context of utterance. If the counterexamples are to be accommodated within his framework,

this assumption needs to be rejected, since it entails that U is always the referent of 'I', which – as the case of Veronique illustrates – is false. An account of which context determines the reference of an indexical is therefore required. One might appeal to conventions to fix the context which determines the reference of an indexical. Corazza et al. (2002) offer an account along these lines, according to which the context is fixed by conventions delivered by the setting in which the utterance takes place. The utterance setting includes such non-semantic features as “speaking English, belonging to a given community, hearing an answering machine message, sarcastically imitating someone, acting in a piece of theatre” (Corazza et al. 2002, 12). The setting for Steve's utterance (2) includes the fact that it is heard on an answering machine; the conventions governing the use of answering machines thus apply to (2). These conventions deliver a context in which Veronique is the agent as the one that determines reference. Veronique is thus the referent of 'I' in (2) according to Corazza et al.'s analysis.<sup>4</sup> An alternative is the intentionalist account given by Predelli (1998) who suggests that the context in question is the one with respect to which U *intends* her utterance to be interpreted. On his analysis, Steve is the utterer of (2), but the context in which he intends his utterance to be interpreted is one in which Veronique is the agent. It is

---

<sup>4</sup> Romdenh-Romluc (2006) raises difficulties for Corazza et al.'s proposal.

thus Veronique who is the referent of 'I' in (2).<sup>5</sup> Finally, one might appeal to the responses of an audience to fix the relevant context. I make a suggestion along these lines (Romdenh-Romluc 2002). I argue that the context which determines the reference of an indexical is the one that a competent and attentive audience (Ac) would identify on the basis of cues that she reasonably takes U to be exploiting. There are many cues that may be used to indicate a context – U can exploit A's history, beliefs, interests, etc.; previous conversations they have had; conventions that they share; and so forth. Since U will sometimes indicate the reference-determining context by exploiting A's desires, beliefs, history, etc., Ac must be a competent and attentive version of the person or people it is reasonable to take U to be addressing. On this view, the Veronique case should be treated as follows. Steve (U) can reasonably be understood as addressing the people who telephone Veronique's office. A competent and attentive caller (Ac) will reasonably take U to be indicating the reference-determining context on the basis that the number dialled is that of Veronique's office. On this basis, Ac will interpret (2) with respect to a context where Veronique is the agent. Thus Veronique is the referent of 'I' in (2).<sup>6</sup>

---

<sup>5</sup> For objections to Predelli's account see Corazza et al. (2002), Romdenh-Romluc (2002), and Romdenh-Romluc (2006). He responds to some of these worries in Predelli (2002).

<sup>6</sup> Weatherson (2002), and Predelli (2002) discuss some problems for my account.

## 6 Claims [B] and [C]

To defend TA, one must show that there is a satisfactory account of how 'I' refers in all cases, including the counterexamples to [A], which nevertheless renders [B] and [C] true of uses of 'I' that refer to U. In this section, I will examine whether such restricted versions of [B] and [C] can be maintained on any of the three accounts outlined above. [B] is the guarantee that every utterance of 'I' will have a referent. If we restrict the scope of this claim to cases where 'I' refers to U, it is clear that the guarantee holds on any account, since U must exist for the utterance to be made. [C] is the guarantee that the intended referent and the actual referent of each utterance of 'I' will always coincide. I will first consider this claim with respect to my account and that of Corazza et al., before discussing it with respect to Predelli's proposal.<sup>7</sup>

[C] does not hold for any uses of 'I' on my accounts and that of Corazza et al.<sup>8</sup> Their proposals allow for the possibility of wrong reference as follows. On Corazza et al's account the intended and the actual referent will diverge if the context fixed by the relevant

---

<sup>7</sup> This is not, however, because I agree with Predelli (2002) when he claims that my account collapses into that of Corazza et al. See Romdenh-Romluc (2006) for a discussion of the differences between them.

<sup>8</sup> This is not to say that there are no cases where the intended and the actual referent of 'I' are the same. It is rather, the claim that this is not *guaranteed* for any uses of 'I' on their proposals.

conventions is different from the context in which U expects her utterance to be interpreted. Similarly, the intended referent will differ from the actual one on my account if U expects her utterance to be interpreted with respect to one context, but Ac would identify a different one as the one that determines reference. It follows that for [C] to be true on Corazza et al's account, U's beliefs about which conventions the utterance setting delivers and which context is fixed by these conventions would have to be infallible. But there is no reason to think that U's beliefs about these things are infallible, even if the scope of [C] is restricted to those uses that refer to U. Similarly, for [C] to hold on my account, U must have infallible beliefs about the method of identifying a context that Ac can reasonably be expected to use and the context identified by this method. Again, there is no reason to take U's beliefs about these matters to be infallible, even if the scope of [C] is restricted as suggested above.

To illustrate these points, consider this case. Simeon presents both a history programme and a holiday programme. One week he gets confused and thinks that he is presenting the history programme when in fact he is presenting the holiday programme. Believing that the television audience have just been told that the year is 1066, and seen a re-enacted version of William the Conqueror's army landing on a beach, Simeon utters,

(3) I have arrived in England!

Intuitively, Simeon intends to convey information about William the Conqueror; he intends to inform the television audience that William the Conqueror arrived in England in 1066. The best way to capture this intuition is by understanding Simeon as intending to say something about William the Conqueror, and so intending 'I' to refer to this individual.<sup>9</sup> He consequently believes that the reference-determining context for (3) is one in which William the Conqueror is the agent.

---

<sup>9</sup> Alternatively, one might hold that Simeon intends to refer to *himself* using 'I', but intends his audience to pretend that he is William the Conqueror and infer information about William from the pretence. However, this is problematic. There does not seem to be any relevant difference between Simeon's use of 'I' in (3), and Steve's use of 'I' in (2) – both intend to use 'I' to convey information about another person. Thus it seems that both cases should be treated in the same way. It looks counterintuitive to hold that Steve intends to refer to himself, but intends his audience to pretend he is Veronique and on this basis infer information about Veronique, since it will surely be obvious to Steve that callers may obtain information about Veronique whilst being oblivious to the fact that the message was recorded by Steve. Instead, it seems that Steve should be understood as simply intending to refer to Veronique. Similarly, it appears that Simeon should be understood as intending to refer to William the Conqueror. Of course, Simeon may in fact intend his audience to understand that he is pretending to be William the Conqueror, and so identify the reference-determining context for the occurrence of 'I' in (3) partly on this basis, but for the reasons given above, we should not take this fact to mean that Simeon intends to refer to himself using 'I'.

However, the opening credits state that it is the holiday programme. The setting in which Simeon utters (3) thus delivers conventions which fix a context in which he is the agent as the one that determines reference. The actual referent of 'I' is therefore Simeon on Corazza et al's account. My account yields the same result. Simeon can reasonably be understood as addressing the television audience, who can reasonably be expected to identify the reference-determining context on the grounds that they have just been informed by the opening credits that it is a holiday programme. The competent and attentive television audience would therefore identify the reference-determining context as one in which Simeon is the agent. He is therefore the referent of 'I' in (3). Thus the restricted version of [C] does not hold.

The restricted version of [C] *does* appear to hold on Predelli's account. Assuming that U is always in a position to know for certain which context she intends as the context of interpretation, it seems that U's intended referent will always be the actual referent of her utterances of 'I'. Thus [C] appears to hold for *all* 'I'-uses on Predelli's proposal, including those that refer to U. It seems, therefore, that on Predelli's account, [B] and [C] will be true of 'I'-uses where U is the referent, and so one might think that TA may be saved by endorsing his theory of indexical reference.

But this strategy for rescuing TA is unsuccessful. First, it is not clear that [C] is preserved on Predelli's account. When I intend to refer to

some entity *a*, there is more than one way for my intention to 'latch onto' it. I may, e.g., think about *a* in virtue of seeing it in front of me at the time of thinking. Or I may think about *a* in virtue of knowing some fact which is true of it – e.g., that it is the Queen's pet kangaroo. In some cases my conception of *a* may involve more than one way of thinking about it – e.g., I may see *a* in front of me and think that the entity I now see is the Queen's pet kangaroo. In these cases I could, however, be wrong in supposing that it is the same entity which is at stake – the entity I now see in front of me may not be the Queen's pet kangaroo. It is plausible to suppose that there are cases where *U* utters 'I' intending to refer to an entity, which she thinks of as that which is *F*. *U* thinks that the thing which is *F* is also *G*. However, she is wrong in supposing that it is one and the same entity which is both *F* and *G*. When she utters 'I', she intends her utterance to be interpreted with respect to a context where the entity that is *G* is the agent. Her utterance of 'I' thus refers to the entity which is *G* on Predelli's account. Since *U* intends to refer to the entity which is *F*, this looks like a counterinstance to [C]. Notice further that this problem is not alleviated if the scope of [C] is restricted to uses of 'I' that refer to *U*. For the restricted version of [C] to hold, all of *U*'s beliefs about herself must be infallible. But there is no reason why this should be the case, since one often has false beliefs about oneself. The point can be illustrated with the following case. Bert and Ernie are absent-minded academics. Bert rings Ernie and tells him he will not be in work and asks

him to put a note on his door to inform his students. Ernie wrongly believes that Bert is the current Head of Department. In fact, Ernie took over the position earlier in the year, but has forgotten. Ernie writes the note,

(4) I am not here now.

and pins it to the door marked 'Head of Department'. In this case, Ernie's intended referent is Bert. But since Ernie has pinned the note to the door marked 'Head of Department' to facilitate communication with his audience, we can understand him as intending that (4) be interpreted in a context where the Head of Department (who he wrongly believes to be Bert) is the agent. The context which determines the reference of 'I' in (4) in Predelli's view is that where the Head of Department is the agent. Since Ernie is the Head of Department *he* is thus the actual referent of 'I', even though the intended referent is Bert.

One may be unpersuaded by the above argument, and reason that since the Head of Department is the agent of the intended context of interpretation, there is at least a sense in which Ernie intends to refer to the Head of Department. Thus there are some grounds for taking the Head of Department to be the intended referent of Ernie's utterance. I think that this is at *least* questionable, but will leave the matter undecided since there is a further problem for TA. As we saw

above, it is how one thinks of oneself when one has a first-person thought that distinguishes this kind of thinking from non-first-personal thought. Under the present proposal, TA endorses thesis [I-t\*], which states that first-person thoughts are all and only those that could be expressed in English by the thinker with a use of 'I' that refers to U. TA then aims to explain what is distinctive about first-person thinking by appealing to claims [B] and [C]. It is clear that for this strategy to work, what [B] and [C] claim about the reference of 'I' must be *uniquely* true of those uses that are relevant to [I-t\*], i.e., cases where 'I' refers to U. This is because these properties of reference are supposed to explain how I think of myself when I have a first-person thought. If uses of 'I' that refer not to U, but to some other entity *a* also carry the same referential guarantees, it follows that if I express a thought about *a* using 'I', I must be thinking about *a* in the same way that I think about myself when I have a first-person thought. But first-person thoughts are supposed to involve thinking of oneself in a distinctive way in which one cannot think about anything else. Thus [B] and [C] must only hold of those uses of 'I' that refer to U if TA is to be defensible. The problem is that if we take the fact that U always knows for certain which context she intends as the context of interpretation to mean that U's intended referent will always be the actual referent on Predelli's account, then as suggested above, [C] will be true for *all* uses of 'I' on Predelli's proposal.

TA may respond by holding that even though [C] is true of all uses of 'I' on Predelli's account, [B] is not. It is perfectly conceivable that the intended context of interpretation may be one with no agent. Consider this case. Tarquin is found dead on the sofa. The rather zealous, local constabulary are convinced that he was murdered. Anxious to apprehend the murderer before she or he strikes again, they call in MaGuckla the world-renowned criminal psychologist to build a profile of Tarquin's killer. After feverishly considering the case, MaGuckla reaches his conclusion about the sort of person who murdered Tarquin. He bursts out of the office and says to the local constabulary,

- (5) I am a white Caucasian female, about twenty-five years old, who drives a red pick-up truck.

MaGuckla intends to convey information about Tarquin's killer to the local constabulary. We can therefore take him as intending that his utterance be interpreted with respect to a context in which Tarquin's killer is the agent, thus securing it as the reference-determining context on Predelli's account. However, contrary to what the local constabulary believe, Tarquin was not murdered; he died peacefully in his sleep. It follows that there is no murderer and MaGuckla's utterance of 'I' fails to refer. Although [B] is not true of all uses of 'I', it is true of those uses which refer to U, since U must exist for the utterance

to be made. TA could, therefore, argue that [B] and [C] can explain what is distinctive about first-person thinking because it is only those uses of 'I' that carry both referential guarantees together.

However, U is not the only individual who must exist for an utterance to be made. U's biological parents must also exist for an utterance to be made. Thus if U utters 'I' intending that her utterance be interpreted in a context where one of these individuals is the agent, her use of 'I' is guaranteed to have a referent, and so uses of 'I' that refer to U's biological parents also carry both of the referential guarantees [B] and [C]. It might be objected that the existence of U is *logically or conceptually* necessary for the making of an utterance, whereas the existence of U's biological parents is only *physically* necessary. A noise, or some marks on paper just wouldn't be an utterance unless they were produced by an utterer, but it is logically possible for there to be beings who are spontaneously generated out of nothing, and so have no biological mother or father. Even so, we are not forced to accept that only uses of 'I' that refer to U carry guarantee [B]. The existence of a language is a necessary condition for the making of an utterance, moreover, a necessary condition for the existence of a language is the existence of a community of language-users. The sense in which the existence of the community of language-users is necessary for an utterance to occur is the same as the sense in which the existence of U is required for the making of an utterance. 'I' could plausibly be used to refer to the community of

language-users – we personify nations so why not communities of language-users? It follows that [B] will hold for uses of 'I' that refer to the community of language-users, as well as those uses that refer to U.

### 7 An alternative account of indexical reference?

Corazza et al, Predelli, and I put forward accounts of how 'I' refers that cause trouble for TA. TA could reject all three accounts and offer an alternative. In this section, I discuss whether an alternative account is available that preserves the referential guarantees for the right cases, i.e., those and only those where 'I' refers to U. There is no space here to explore every way in which one might try to account for indexical reference. Nevertheless, the foregoing discussion brings to light some considerations that strongly suggest the kind of account TA requires cannot be given.

[B] claims that every utterance of 'I' has a referent. On any account of how 'I' refers, this guarantee holds for uses of 'I' that refer to U, since U must exist for an utterance to take place. But U is not the only entity whose existence is necessary for the making of an utterance. The community of language-users, and perhaps U's biological parents must also exist for an utterance to take place. Uses of 'I' that refer to any of these entities will also carry referential guarantee [B], no matter how we take the reference of 'I' to be fixed. Although the existence of a referent is guaranteed for more uses of 'I' than just those that refer to U, TA could still appeal to [B] and [C] to

explain what is distinctive about first-person thinking if it could be shown that the two guarantees *together* hold of just those uses of 'I' that refer to U. Thus it must be shown that [C] holds of all and only those uses of 'I' that refer to U.

[C] is the guarantee that the intended referent and the actual referent of each utterance of 'I' will be the same. This guarantee will hold if the mechanism that fixes reference is such that it gives U complete authority over the reference of the term in question, i.e., if the factors that determine reference are entirely within U's control. Let us call this an authoritative mechanism. [C] will not hold if the reference of the term is fixed by factors over which U does not have authority. Let us call this a non-authoritative mechanism. The three accounts examined each posit just one mechanism of reference for all uses of 'I'. Thus [C] either holds for all cases (as on one reading of Predelli's account), or for no cases (as on the second reading of Predelli's account, and the my proposal and that of Corzazza et al). To defend TA, however, two referential mechanisms must be posited: an authoritative mechanism must determine reference when 'I' refers to U, whilst a non-authoritative mechanism must determine reference when U is not the referent.

There are various difficulties with supposing that such an account can be given. An immediate difficulty is that a theory of how 'I' refers that posits just one mechanism of reference for all uses of 'I' is simpler and more elegant than a theory that posits two, and is thus preferable.

But setting this problem aside, there are further difficulties. The proposal requires some principled means of distinguishing between uses of 'I' whose reference is fixed by the authoritative mechanism, and uses of 'I' whose reference is fixed by the non-authoritative one. Moreover, if the proposal is to help TA, all and only those uses whose reference is fixed by the authoritative mechanism must have U as their referents. Perhaps the most obvious way to cash this out is to try and distinguish between fictional and non-fictional uses of 'I'. The thought would be that the two uses have different semantics, and whilst non-fictional uses of 'I' refer to U, fictional uses do not. However, it is highly doubtful that the distinction can be drawn in the way TA requires. Clearly, TA cannot simply distinguish between the two uses on the grounds that non-fictional uses refer to U, whereas fictional uses do not. There may be reasons for classifying uses of 'I' in this way, but it is a distinction that is made *after* reference has already been fixed. The fact that a use of 'I' qualifies as fictional on this understanding cannot, therefore, be one of the factors that *determines* reference.

One might suppose that fictional uses of 'I' are simply uses of 'I' that pertain to fiction. But this suggestion does not clearly distinguish fictional from non-fictional uses of 'I'. Simeon's utterance

(3) I have arrived in England!

could be classed as fictional because it occurs as part of a storytelling, but the storytelling concerns events that really occurred, and in this sense (3) is non-fictional. Moreover, there are uses of 'I' that clearly occur as part of fiction, but which nevertheless refer to U – e.g., whilst telling my son a bedtime story about my imaginary visit to Santa Claus I might utter,

(6) I went to visit Santa Claus in his castle last night.

The story is an obvious fiction, but it is a story about me, and so intuitively, 'I' in (6) refers to me. Conversely, there are uses of 'I' that clearly do *not* occur as part of a fiction, where the referent is not U – e.g., Steve's recorded utterance,

(2) I am not here yet.

In this case, 'I' refers to Veronique, not Steve, but (2) is an informative message on an answering machine, not a piece of fiction.

Even supposing that fictional uses of 'I' could be properly distinguished from non-fictional uses, and that all and only non-fictional uses of 'I' had U as their referent, there is yet a further difficulty. To sustain TA, the reference of non-fictional uses of 'I' must be fixed by factors that are entirely within U's control. On the proposal we are exploring, uses of 'I' are divided into fictional and non-fictional uses,

and this classification determines which mechanism fixes the reference of a use of 'I'. In other words, one of the factors that determines the reference of some use of 'I' is whether or not that use is fictional or non-fictional. It follows that for U to have control over all the factors that determine the reference of a non-fictional use of 'I', she must have control over whether her use is fictional or non-fictional. But it is extremely implausible to suppose that this is within U's complete control. Fiction, e.g., often involves storytelling. Intending to tell a story is perhaps necessary for one to engage in storytelling – and whether or not one has this intention is within one's control – but it is surely insufficient. Telephoning a takeaway and ordering food, e.g., does not constitute telling a story, even if one intends it to. Indeed, it is difficult to see quite how one *could* intend to tell a story by telephoning a takeaway and ordering food, which points to the fact that storytelling is a rule-governed activity, and, of course, one cannot follow a rule merely by intending to do so. It is very likely that the same points will apply, no matter how fiction is defined.

TA cannot be saved by distinguishing between fictional and non-fictional uses of 'I'. Moreover, it is plausible to suppose that similar difficulties will be encountered by any attempt to save TA by distinguishing between different uses of 'I'. Thus the prospects of providing an account of how 'I' refers that can sustain TA are bleak.

### 8 'I' as the canonical expression of first-person thought

I suggested that TA might adopt the following thesis as a means of identifying first-person thoughts:

[I-†\*] First-person thoughts are all and only those that could be expressed in English by the thinker, with a use of 'I' that refers to U, i.e., to the thinker.

TA could then use [B] and [C] to explain what is distinctive about first-person thinking, providing that [B] and [C] are true of all and only those uses of 'I' that refer to U.

I have argued that the chances of providing an adequate account of 'I'-use that preserves [B] and [C] for the right cases are slim. It should also be noted that [I-†\*] is false. Consider, e.g., the case of Tarquin who is found dead on his sofa. Imagine this time that Tarquin really has been murdered. In a strange twist of fate, MaGuckla is the murderer, but suffers from amnesia and has no memory of carrying out the ghastly crime. After spending some time piecing together the evidence, MaGuckla announces,

(5) I am a white Caucasian female, about twenty-five years old, who drives a red pick-up truck.

We saw earlier that Tarquin's killer is the referent of 'I' on Predelli's account. We get the same result on my proposal and that Corazza et al. The setting for (5) includes the fact that it is spoken by someone who has just spent several nights building a profile of Tarquin's killer, to the people who have employed him to do so. This setting delivers conventions that fix the relevant context as one in which Tarquin's murderer is the agent. Hence Tarquin's killer is the referent of 'I' on Corazza et al's view. The competent and attentive local constabulary are Ac. They reasonably take MaGuckla (U) to be indicating the reference-determining context on the basis that they have employed him to profile Tarquin's killer. On this basis, they identify the context in which the person who murdered Tarquin is the agent, and so on my account, this person is the referent of 'I'. Thus on all three accounts, 'I' in (5) refers to Tarquin's killer. When MaGuckla utters (5) he expresses a thought – he expresses his sudden realisation that the person who killed Tarquin is a white Caucasian female who drives a red pick-up truck – and since MaGuckla is the killer, he expresses a thought about himself. But it is clearly not a first-person thought.

Neither [I-t] nor [I-t\*] correctly reflect the relation between 'I'-use and first-person thought. Nevertheless, there appears to be some connection between the two. This is illustrated by the fact that it seems natural to express the first-person thought the messy shopper has after he realises that it is *he* who is making a mess (but not the thought he has about himself before this realisation) using 'I'. One might wonder

whether a correct conception of the relation might yet save TA. The arguments in the previous section suggested that TA is beyond saving, since it is highly unlikely that an account of how 'I' refers is available that preserves the referential guarantees in the right cases. But setting this issue aside, it also seems that 'I'-use is not related to first-person thought in the kind of way that TA requires either. Here is a suggestion as to how they are related.<sup>10</sup> To use 'I' is to speak from a point of view; it is to speak from the perspective of the referent. Thus to use 'I' is to speak about the referent from the referent's own perspective. I take it that this is what Kaplan is trying to capture when he states that the character of 'I' picks out the *agent* of the context. First-person thoughts are a kind of self-conscious thought about oneself. There is much disagreement over what self-consciousness is, however, it is generally agreed that to be self-conscious is tied up with having a point of view. Thus the following is uncontroversial, if somewhat vague: to have a first-person thought is, in some sense, to have a thought about oneself from one's own point of view. Since to have a first-person thought is to think about oneself from one's own point of view, and since to utter an 'I'-sentence is to speak about the referent from the perspective of the referent, 'I' is particularly suited to the expression of first-person thoughts. Hence one might say, as many theorists do, that 'I' is the canonical linguistic expression of first-person thought.

---

<sup>10</sup> This suggestion is not particularly new or original. Perry (2000) explores an account of first-person thought along these lines.

There is, of course, far more to be said on this matter. Space here prevents me from exploring these ideas in more detail. But it is obvious from even these brief comments that this conception of the relation between 'I'-use and first-person thought cannot help TA. TA attempts to identify first-person thoughts by using 'I'. By hypothesis, it is canonical uses of 'I' that express first-person thoughts. Thus one might suppose that first-person thoughts can be identified by the fact that they are those thoughts expressed by canonical uses of 'I'. The problem is that canonical uses of 'I' are themselves identified by the fact that they express first-person thoughts. Canonical uses of 'I' do not, therefore, offer a non-circular way of identifying first-person thoughts.

## 9 Conclusion

TA claims that first-person thoughts involve a self-conception, which is formed in such a way that it always successfully 'latches onto' some worldly entity, and that entity is always the thinker. TA bases this thesis on claims about the first-person pronoun, [A], [B], and [C]. According to recent work in linguistics, these claims are false, which spells trouble for TA. One might suppose that TA can be defended by focusing just on uses of 'I' that refer to U, and trying to show that [B] and [C] hold of these uses. I have argued that this strategy is unsuccessful. It follows that if the work in linguistics is accepted, then TA must be abandoned.

It is unlikely that proponents of TA will be convinced by my conclusions, although I am by no means the first to suggest that this approach to first-person thinking is flawed. Anscombe (1975), e.g., makes this point albeit on different grounds to those explored here, and TA has largely ignored the problems she raises. A line someone could take is to hold that the analyses of first-person thinking offered by proponents of TA are largely correct. The problem, of course, is that they are based on false claims about 'I'. Thus it might be supposed that these analyses can be saved by simply divorcing them from the problematic claims about the first-person pronoun. This would make them 'free-standing' accounts of first-person thought, rather than accounts of thought grounded in observations about linguistic phenomena. But since TA bases statements about what this kind of thinking entails – and so what an account of first-person thinking needs to explain – upon the problematic claims made about 'I', if TA takes this line, and simply jettisons its faulty conception of the linguistic data, we are entitled to ask why we should accept that first-person thinking has the features that TA takes it to have.

### References

Anscombe, G. E. M. (1975). The first person. (In S. Guttenplan (Ed.), *Mind and language: Wolfson College lectures 1974* (pp. 45—64). Oxford: OUP.)

Bermúdez, J. L. (1998). *The Paradox of Self-Consciousness*.  
(Massachusetts: MIT Press)

Corazza, E., Fish, W. & Gorvett, J. (2002). Who is I? *Philosophical Studies*,  
107, 1—21

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

Kaplan, D. (1977). Demonstratives. (In J. Almog, J. Perry and H.  
Wettstein (Eds.), *Themes from Kaplan* (pp. 481—563). Oxford: OUP.)

Perry, J. (2000). Myself and I. (In J. Perry, *The problem of the essential  
indexical and other essays* (pp. 325—339). California: CSLI Publications.)

Predelli, S. (1998). I am not here now. *Analysis*, 58, 107—12

Predelli, S. (2002). Intentions, indexicals, and communication. *Analysis*,  
62, 310—16

Romdenh-Romluc, K. (2002). Now the French are invading England!  
*Analysis*, 62, 34—41

Romdenh-Romluc, K. (2006). 'I'. *Philosophical Studies*, 128, 257—283

Shoemaker, S. (1968). Self-reference and self-awareness. *The Journal of Philosophy*, 65, 555—567

Sidelle, A. (1991). The answering-machine paradox. *Canadian Journal of Philosophy*, 21, 525—39

Weatherson, B. (2002). Misleading indexicals. *Analysis*, 62, 308—310