Self-Consciousness and Its Linguistic Expression

Rory Madden

Abstract Which linguistic actions are expressions of self-conscious states of mind? I defend a certain answer to this question. Having presented problems for a simple view of the connection between self-conscious states of mind and first person language, and for a slight modification of the view, I go on to distinguish two, more promising, ways of getting a linguistic handle on first person thought. These two positions—which I call the Knowledge View and the Intention View—are not explicitly distinguished in the existing literature on the subject. My aim is to argue that the Intention View is the superior view. One reason for preferring the Intention View is its capacity to furnish a noncircular route to the identification of first person thoughts. This advantage accrues from the way in which objects of intention contrast with objects of propositional knowledge. Another reason for preferring the Intention View is that it diagnoses what is going on in certain persuasive counter-examples to the Knowledge View. In the final section of this chapter, I consider whether the Intention View is subject to some counter-examples of its own. Clarification of the relevant notion of linguistic expression reveals the counter-examples to be merely apparent.

1 Introduction

How is self-consciousness expressed in language? Which utterances are expressions of self-conscious states of mind? I want to defend a certain answer to this question.

Here is a familiar example to illustrate the notion of a self-conscious state of mind.

> As John rounds the aisles of the supermarket he spies a trail of spilled sugar, and comes to believe that the shopper with a torn bag is making a mess. In fact John is unwittingly depositing the trail of sugar from a torn bag in his own supermarket trolley. After some minutes of futile pursuit of the shopper—the trail of sugar growing thicker as he goes round and round the same shelves—it finally dawns on John that he himself is making the mess. He stops the trolley and rearranges the bag.

R. Madden (✉)
Department of Philosophy, University College London, London, UK
e-mail: r.madden@ucl.ac.uk

Tzu-Wei Hung (ed.), *Communicative Action,* DOI 10.1007/978-981-4585-84-2_5,

Since John himself was the shopper with the torn bag all along his belief that the shopper with a torn bag is making mess was, in fact, a belief about himself. So he was in one sense thinking about himself from the outset. When finally he comes to realize that he himself is making a mess he is still thinking about himself, but there has been a shift in the functional role of his psychological state. His belief now disposes him to stop pushing the trolley and to begin fixing his possessions. He feels embarrassed. He resolves to be more careful in future. He has entered a state one is disposed to enter upon tracing a trail of sugar back to roughly the origin of one's own egocentric representations of the world. States of mind with this distinctive, self-conscious, kind of functional role philosophers are apt to call "first personal" states of mind. After his realization, John thinks of himself first personally.

But the term "first person" primarily effects a *grammatical* categorization, of pronouns, possessive determiners, and verb forms. Why should philosophers use a grammatical term to classify states of mind?

Evidently, philosophers' use of the term here is metonymic; the states of mind in question are states of a kind that are expressively associated with language grammatically categorized as first personal. For example, after his realization, and not before, John will be disposed to express his state of mind using the first person pronoun. As an English speaker, he will now be disposed to say such things as "I am making a mess."

But how straightforward is the connection between first personal states of mind and uses of the first person pronoun and cognate expressions? It is sometimes assumed that first personal states of mind can simply be identified with those states of mind that would be expressed using the first person. This simple view, as I shall shortly argue, is mistaken. In order to more accurately identify first personal states of mind by their means of linguistic expression, we need to keep in mind that expressive language use is a rational, intentional, activity. I shall argue that first personal states of mind are states of mind expressible by linguistic actions performed with a certain intention. This position explains why first personal states of mind correlate imperfectly, though still closely, with uses of first personal language.

In the next section, I present problems for the simple view of the connection between first personal states of mind and first personal language, and for a slight modification of the view. I go on to distinguish two, more promising, ways of getting a linguistic handle on first person thought. These two positions—which I shall call the Knowledge View and the Intention View—are not explicitly distinguished in the existing literature on the subject. My aim is to argue that the Intention View is the superior view. One reason for preferring the Intention View is its capacity to furnish a noncircular route to the identification of first person thoughts. This advantage accrues from the way in which objects of intention contrast with objects of propositional knowledge. Another reason for preferring the Intention View is that it diagnoses what is going on in certain persuasive counter-examples to the Knowledge View. In the final section, I consider whether the Intention View is subject to some counterexamples of its own. Clarification of the relevant notion of linguistic expression reveals the counter-examples to be merely apparent.

2 A Simple View

It is helpful first to reflect on the shortcomings of a simple view of the connection between first personal states of mind and first personal language.

(Simple View) One linguistically expresses a first person thought if and only if one uses first personal language.

The Simple View faces counterexamples, to both the necessity and the sufficiency of first personal language.

Against necessity it is enough to point out that first personal states of mind can be expressed by the arch, or pompous, use of one's own name, à la General De Gaulle (certainly in contexts in which it is common knowledge that it is one's name). Equally, one might use more complex descriptive expressions to give voice to first personal states of mind, such as "the subject of these experiences" or "this speaker." First personal devices are not necessary for the linguistic expression of first person thinking.

Is it promising to regard the use of first personal language as at least a *sufficient* condition for the expression of first person thought?

Some nonstandard uses of "I" suggest otherwise. Suppose that a helpful PA has the bright idea of recording a computerized answering machine message on behalf of The Boss, who is presently incommunicado on a delayed flight. The PA, a competent English speaker, chooses to produce the sentence "I am away from my office."[1] The PA thereby comprehendingly uses first personal language. But the PA does not thereby express a self-conscious, first personal, state of mind. If it is right to say that the PA expresses a thought at all, the PA expresses a "third personal" thought, which would in more normal circumstances be expressed by "The Boss is away from his office." The use of first personal language is not a sufficient condition for the expression of first person thought.

A notable feature of the case is that the PA does not use "I" to refer to himself. It might be thought that the simple view can be adapted accordingly, to give a sound criterion for the expression of first person thought:

(Simple View) One linguistically expresses a first person thought if one uses first personal language *and one thereby refers to oneself.*

The counterexample can be adapted accordingly. In a variant case it is not the PA who records the message but The Boss himself. For The Boss suffers an amnesiac-delusional episode on the airplane, during which he forgets entirely who he is and comes to believe that he is the earthbound PA to the airborne Boss. He uses a satellite phone to record a first personal message "I am away from my office"—as he conceives of it, on behalf of someone else.

[1] Romdenh-Romluc (2008) gives roughly this example, in the service of a general skepticism about using language to identify first person thought. The conclusions of the present essay imply that such skepticism is unwarranted.

The Boss is the speaker's intended referent of "I." The Boss is the person whom listeners of the message will take to be the referent of "I." The Boss is the producer of the word "I." So it is extremely plausible that The Boss thereby refers to himself. However, he no more expresses a first person thought than the PA in the original case. Thus, one can use first personal language and thereby refer to oneself, and yet not express a first person thought. The Simple View* is no improvement over The Simple View.

Can we do better? In her essay "The First Person," G. E. M. Anscombe considers, only to reject, the following proposed explanation of the meaning of the English first person pronoun.

> "I" is the word each one uses when he knowingly and intentionally speaks of himself. (Anscombe 1975, p. 22)

As it stands, this proposal is vulnerable to the examples just considered. In De Gaulle-type cases, the speaker knowingly and intentionally speaks of himself but he does not use the word "I." Conversely, in both the PA cases and their variants the speaker uses "I" but does not knowingly and intentionally speak of himself.

These observations, however, strongly suggest an improved approach to the connection between first person states of mind and their linguistic expression: one linguistically expresses a first person thought if and only if one knowingly and intentionally speaks (writes, etc.) of oneself. None of the recent counterexamples threatens this way of getting a linguistic handle on self-conscious states of mind. Just in so far as he knows that "De Gaulle" is his own name and intends to speak of himself, De Gaulle is using his name to express his first person state of mind. Equally, a speaker who expresses first person states of mind by means of "the subject of these experiences" is a speaker who also thereby knowingly and intentionally refers to himself by means of that descriptive device. We can also note that the approach accords nicely with the opening messy shopper case. Before the onset of his self-conscious belief John will in fact refer to himself, with the expression "the shopper with a torn sack"; what is notable is that he does not refer to himself knowingly and intentionally. When self-awareness dawns he switches to the first person pronoun "I" and now he knowingly and intentionally refers to himself.

I think that something along these lines is correct. However Anscombe's wording "knowingly and intentionally" fuses two strands of thought. Separating these strands yields two views of the connection between first person thought and language:

(Knowledge View) One linguistically expresses a first person thought if and only if one uses language in the knowledge that one is thereby referring to oneself.

(Intention View) One linguistically expresses a first person thought if and only if one uses language with the intention thereby to refer to oneself.

These views are not obviously incompatible. For all the examples so far have shown they are both true.[2]

What I want to argue, though, is that the Knowledge View is incorrect. The Intention View is much more promising. In the next part of the chapter, I shall explain an explanatory advantage of the Intention View over the Knowledge View. I shall then turn to more straightforward counter-examples to the Knowledge View, which the Intention View can handle quite comfortably.

3 Objects of Knowledge and Objects of Intention

Why did Anscombe reject the explanation of the first person pronoun "I" as "the word each one uses when he knowingly and intentionally speaks of himself?" I shall briefly summarize her extended argument, which can be seen as posing a trilemma. This argument is prima facie threatening to both the Knowledge View and the Intention View.

First, note that the ascriptions of knowledge and intention on the right-hand sides of the Knowledge View and the Intention View, must, in order for the two theses to have any plausibility whatsoever, be understood as opaque. In particular, the occurrence of the reflexive pronoun "oneself" after the attitude words cannot be understood as a transparent reference to the speaker, substitutable *salva veritate* with any designation of the speaker. For, on a transparent reading of the ascription, it is true that the messy shopper, even before he enters a first personal state of mind, knows that he refers to, and intends to refer to, himself. For he knows that he refers to, and he intends to refer to, the messy shopper—and the messy shopper is he.

On the other hand, if the attitude ascriptions are opaque, then how should their truth conditions be construed? A natural thought is that the occurrence of "oneself" is an indirect report of a first personal reference made by the speaker to whom the attitude is ascribed. The ascription of knowledge is true only if the speaker knows that he refers to the speaker, while conceiving of the speaker under a first personal guise, or mode of presentation. Similarly, the ascription of intention is true only if the speaker intends to refer to the speaker, conceiving of the speaker under a first personal guise or mode of presentation. With the ascriptions read in this way the Knowledge and Intention Views accord with the datum that the messy shopper does not express a first personal state of mind before his realization. The messy shopper knows that he refers to, and intends to refer to, what is in fact himself—but he does not conceive of the object of reference in first personal terms.

[2] Those attracted to the Knowledge View include O'Brien, who holds that "reflexive reference can only be first-person reference if one knows that one is referring to oneself" (2007, p. 9), and with qualifications, Peacocke: "fully self-conscious uses of 'I' are those in which the thinker knows that he is referring to himself" (2008, p. 78). Nozick (1980, p. 79) is another advocate. Evans appears to be sympathetic to both the Knowledge and the Intention Views (1982, pp. 258–259). Although Rumfitt's (1994) discussion of conventional linguistic meaning of "I" is suggestive, I am aware of no explicit defense of the Intention View. The present essay aims to remedy this situation.

In response to this natural thought, however, one might well follow Anscombe in raising a question about the noncircularity, or explanatory power of the views read in this way. For once driven to understand the relevant occurrences of "oneself" in essentially first personal terms, it seems we can no longer give an informative linguistic identification of first person thought in independent terms. We are distinguishing cases of the expression of first person thought from other cases precisely by reference to first person thought on behalf of the subject.[3]

Anscombe assumes that the only avenue of escape from this circularity is to embark upon the quest for some independently specifiable mode of presentation, or conceptual guise, by means of which the speaker singles himself out in the content of the knowledge or intention in question. Anscombe is skeptical about the prospects for such a quest. Neither descriptive nor demonstrative modes of presentation seem adequate. In cases of total descriptive ignorance or error about one's properties—and in sensory deprivation cases in which one is in no position to demonstrate oneself—one can surely still knowingly and intentionally refer to oneself in the way that is expressive of first person states of mind.

Thus, the advocate of the general proposal that first personal thoughts are those linguistically expressible using language by means of which one knowingly and intentionally refers to oneself faces a trilemma. (a) If the occurrence of "oneself" is understood to make a transparent reference to the subject then the proposal is clearly extensionally inadequate. (b) If the occurrence of "oneself" ascribes a first personal way of thinking of the subject then the proposal is extensionally adequate but circular. (c) If the occurrence of "oneself" ascribes a nonfirst personal way of thinking of the subject then, again, the proposal is extensionally inadequate.

How should one respond to this trilemma? While it would take a lot more argument than can be given here, I think Anscombe is certainly right to be skeptical about the possibility of locating some independently specifiable, descriptive or demonstrative mode of presentation of the subject. The weak point in her argument is rather a presupposition of the whole trilemma. In presenting the choice between interpreting the relevant occurrence of "oneself" as a transparent reference to the subject, a first personal reference to the subject, or some other reference to the subject, Anscombe is clearly presupposing that the reflexive pronoun "oneself" within the scope of the attitude verbs has a referential function of some kind.[4] There is a better way of understanding the occurrence of the reflexive pronoun. The right-hand sides of the Knowledge View and the Intention View are ascribing attitudes to a *reflexive act*.

To get a feel for this view, consider the following example. When on a particular occasion Alice washes Alice there is whole range of things that Alice thereby does. She uses water. She uses soap. She washes Alice. She washes someone. The single event in which she does all these things we can call an *action*, which is an unrepeatable event, occurring at a particular time, with a particular agent. The various things she *does*, we can call *acts*. Acts are not unrepeatable particulars like action.

[3] This is the "paradox of self-consciousness" which animates Bermudez's eponymous 1998.

[4] I owe this astute diagnosis to Rumfitt (1994).

Different agents at different times may engage in the very same act. For example, when on some other occasion *Bob* washes Alice, he does one of the very same things that Alice does on this occasion: he performs the act of washing Alice. Acts may be specified by predicates, or open sentences, such as "*x* washes Alice." The predicate specifies the act any *x* engages in when *x* washes Alice.[5]

Now suppose on a particular occasion that Bob washes Bob. There is something Bob does on this occasion that Alice does on an occasion that Alice washes Alice. This act we specify using the *reflexive* predication "*x* washes *x*." The act is the act any *x* engages in when *x* washes *x*. We may call this act the act of self-washing or reflexive washing.

The way out of Anscombe's trilemma, then, is to understand the ascriptions of knowledge and intention as ascribing to the speaker an attitude toward an act of just this sort. It is the act specified by the predicate "*x* refers to *x*": the act of self-referring or reflexively referring. In order to think of such an act, the speaker need to employ no first personal concept: They need only to be able to discern the common feature present in such situations as the situation in which Alice refers to Alice, Bob refers to Bob, Carol refers to Carol (and absent from the situation in which Alice refers to Bob). When the Knowledge and Intention Views ascribe attitudes toward referring to oneself, the "oneself" here is not making a reference, first personal or otherwise. The views ascribe to the speaker attitudes toward the act any *x* does when *x* refers to *x*.

At this point, however, we can begin to discern an advantage of the Intention View over the Knowledge View. The Knowledge View claims that a first person thought is expressed just in case the one knows that one thereby refers to oneself. Let it be agreed that the occurrence of the reflexive pronoun "oneself" is not to be understood as making a reference to the speaker; the condition is that the speaker knows that he thereby performs the act any *x* performs when *x* refers to *x*. An obvious question remains: how should we understand the "one" that heads the sentential complement of the attitude verb?

Here, I think, there is no choice but to regard the "one" as referential. Thus Anscombe's trilemma can again be pressed, now with full force. First, if the "he" is understood transparently then the Knowledge View is false. A complex example is needed to illustrate the point.

Again, John is in fact the messy shopper although he has not yet realized this. John expresses a third personal thought using the words "the messy shopper is reflexively referring." He says this because he lip-reads the words of an evidently messy shopper using the term "the messy shopper." He knows by visual means that the messy shopper performs the act of reflexively referring: He knows that the messy shopper refers to the messy shopper. What he does not know is that he is in fact looking at himself in a mirror. Thus, with the "he" understood to make a transparent reference, it is true that John knows that *he* is reflexively referring. For he knows that the messy shopper is reflexively referring—and the messy is shopper is he. But John does not linguistically express a first person thought.

[5] See Rumfitt (1994) for a more detailed presentation of this way of identifying acts.

It is clear then, that for the Knowledge View to have plausibility, the "he" heading the complement of the verb "knows" cannot be transparent; indeed the view is surely only plausible if the "he" ascribes a first personal thought, which would be expressed by the words "I am referring to myself." The nonreferential reading of the original reflexive does not defuse this point. Thus, the Knowledge View remains accused of helping itself to the ascription of first person thought in identifying cases of the linguistic expression of first person thought.

Why, though, should the Intention View be supposed to have any advantage here? Does not exactly the same problem arise for the Intention View?

The same problem would arise just in case one made the admittedly common "propositionalist" assumption that the objects of intention, like the objects of propositional knowledge, must be propositional in nature. On this common assumption, an intention to ϕ must be understood as an intention *that* one will ϕ. In the present case, the view would have it that the intention to reflexively refer must be understood as the intention that one will reflexively refer. If this propositionalist view of the objects of intention is correct, then the same difficult question arises for the Intention View, how to understand the occurrence of "one" heading the sentential complement of the intention ascription. It is again plausible that the view will be false unless the "one" is understood as making a first personal reference to the speaker. So no less than the Knowledge View, the Intention View must help itself to the ascription of first person thought in order to identify first person thoughts by their conditions of linguistic expression.

Now, one reaction to this would be to question the supposed viciousness of this element of circularity. One might still find the Knowledge and the Intention Views interesting and informative claims even if they fail to provide an impeccably non-circular route to the identification of first person states of mind via their linguistic expression.

But in fact an advocate of the Intention View need not retreat to this less ambitious stance. Instead the advocate of the Intention View may refuse to accept that an intention to ϕ must be understood as an attitude to a proposition in the first place. The alternative view is to regard the objects of intention as simply *acts*. Whereas, the things we know are such things as *that the Battle of Hastings was in 1066,* the things we intend are naturally regarded as belonging to the same category as the things we do. And the things we do are simply acts, such as washing Alice, buying a ticket, or, as the case may be, reflexively referring. On this view there need be no first personal element to the object of intention itself. While perhaps intentions to act are always in some sense "directed" or "sent" to oneself, a reference to oneself need be no part of the *content* of what is intended.

An analogy with memory may be helpful. A report of *episodic* memory, of the form "*S* remembers ϕ-ing" does not on the face of it appear to ascribe an attitude to a proposition. There is such a thing as propositional first-person memory, ascribed by such reports as "*S* remembers that he himself ϕ-ed" but there is no evident reason to assimilate the former to the latter. A more plausible view is that episodic memory relates the subject not to a proposition but to a past experiential *event*.[6] As things

[6] See Martin (2001) for defense of this view of episodic memory.

are, we only stand in this remembering relation to events in our own past. For this reason, episodic memory is an immediate basis for first person statements of the form "I ϕ-ed." But this does not show that the memory itself has a propositional object, any more than the fact that awareness of a pain is an immediate basis for the statement "I am in pain" shows that the object of awareness is a first person proposition, and not simply a pain.

Although I am myself attracted to an act-directed view of the objects of intention, this is not the place for its full defense. So I cannot claim here to have conclusively shown that the Intention View has the upper hand over the Knowledge View in responding to Anscombe's worries about circularity.[7] Fortunately, there is a far more straightforward way to demonstrate the advantage of the Intention View of the conditions of linguistic expression of first person states of mind. As I shall argue in the next section, there are counterexamples to the Knowledge View. The Intention View handles these cases quite elegantly.

4 Problems for the Knowledge View

There is a tradition in the philosophy of action, inspired by Anscombe's *Intention*, which takes the view that intentional actions are constitutively those events in which the agent knows, by a certain means, what she is doing.[8] A theorist in this tradition will have little inclination to prise apart the "knowingly" and "intentionally" strands in the way I have here.

There is, I think, a general difficulty with the claim that intentionally ϕ-ing requires knowingly ϕ-ing. It is of course perfectly normal in intentionally doing something to know what one is doing. Indeed, it would be practically near impossible to carry through an extended course of intentional action in the absence of successful self-monitoring. But the strong constitutive claim struggles with the point that the conditions for knowing that one is ϕ-ing seem to be *stronger* than the conditions for intentionally ϕ-ing. Here is an example to illustrate the point.

> Scott's toes are totally numbed by the cold and hidden from sight in his hiking boots. Scott believes that he will wiggle his toes whenever he tries. He is wrong: his toes are so paralyzed by cold that the chance that he will wiggle his toes if he tries is extremely low; the chance of a motor signal making it all the way to his toes is less than one percent. On a certain occasion Scott tries to wiggle his toes and, against the odds, he succeeds. So he intentionally wiggles his toes.

[7] A full defense would need to engage with the position in theoretical linguistics, supported by high-level syntactic considerations, that the underlying syntactic structure of the infinitival clause of an intention ascription contains a covert unpronounced subject-term "PRO," and thereby ascribes a complete first personal propositional thought about the intender. A defender of the present response to Anscombe's circularity worries should question whether the underlying syntactic form of a mental state ascription reflects in any straightforward way the metaphysics of the objects of the state.

[8] Anscombe (1957); Velleman (1989); Setiya (2003).

Does he know that he is wiggling his toes? He believes that he is wiggling his toes, and he is right. But in very similar close possible situations he would have been mistaken in his belief that he is wiggling his toes. So it is plausible that he does not know that he is wiggling his toes. He has a merely luckily true belief. Similar examples can be used to elucidate the superiority of the Intention View to the Knowledge View.

> Gertrude is completely deaf, and her vocal apparatus—chest, throat, mouth, and lips—is completely numb. Although she does not know it, she is also paralysed in such a way that attempts to vocalize will in fact succeed only very rarely. Gertrude thinks that she may have overdone the Novocain. As an English speaker she tries to give voice to this thought, and against all the odds, succeeds, uttering the sentence "I may have overdone the Novocain".

It is plausible that Gertrude gives linguistic expression to her self-conscious state of mind. According to the Knowledge View, then, she must be using language in the knowledge that she is thereby self-referring. However, while her belief that she is self-referring by means of "I" is true, it is not plausible that she knows that she is self-referring by means of "I." Given the improbability of motoric success, and the absence of any alternative auditory or proprioceptive means of self-monitoring, Gertrude's belief is extremely unsafe. She could very easily in close possible situations have been mistaken in believing that she is self-referring by means of "I." So she does not know that she is self-referring. She has a merely luckily true belief to that effect.

The case is quite unlike the situation of the ignorant messy shopper who uses "the messy shopper" without knowing that he is thereby self-referring. Unlike the ignorant messy shopper Gertrude linguistically expresses a first person state of mind. The Intention View respects this difference. Gertrude, unlike the messy shopper, speaks with the *intention* to self-refer.

Could the defender of the Knowledge View individuate methods in such a way that basic motoric success is required for the method distinctive of Anscombian "practical knowledge" genuinely to be operative? On the basis of this externalist individuation of methods, it might be argued that the close possible cases of error, while subjectively similar to the actual case of successful wiggling, cannot be cases in which the same method is used to form the belief that one is ϕ-ing.[9] The safety of Gertrude's true belief that she is self-referring, based on one method, is not undermined by the close possibility of making a mistake on the basis of some distinct method.

However this may be, the shortcomings of the Knowledge View can be illustrated with examples in which there is no such breakdown in basic motor control.

> Edmund has a shaky grasp of English vocabulary. He defers to the English linguistic community in his use of English words, intending them to have

[9] See Nozick's "Jesse James" and "Sick Grandmother" cases for the general point about the need to relativize reliability conditions to methods (1980, p. 179 ff).

> whatever is their conventional meaning. Edmund has two further peculiarities. First, he is under the misapprehension that the English word "I" is a proper name that English speakers conventionally use to refer only to him. Second, he is in the pompous De Gaulle habit of using what he takes to be his own proper name to refer to himself. He intones self-importantly to his English audience "I will not tolerate this sniggering".

Edmund uses the word "I" with deferential semantic intentions. The word "I" is conventionally a device of self-reference in English. So Edmund self-refers using "I." He believes that he self-refers using "I" and he is in fact correct in believing this. However, his basis for believing that he is self-referring using "I" is the false lemma that "I" is his own proper name. Thus, he does not know that he is self-referring using "I." This is a Gettier case, for knowledge of linguistic self-reference.

Edmund is nevertheless linguistically expressing a first person thought, about what he himself will not tolerate. Therefore the Knowledge View is mistaken. The Intention View does better. Edmund, despite his semantic confusion, still speaks with the intention to self-refer.

The cases of Gertrude and Edmund make it plausible that the knowledge that one is self-referring is not necessary for the linguistic expression of first person thought. Is it nevertheless *sufficient* for the linguistic expression of first person thought that one should use language in the knowledge that one is thereby self-referring? Perhaps, this direction of the Knowledge biconditional is still plausible.

As the following kind of example brings out, not even this much is clear. The example serves to illustrate again the merits of the Intention View.

> Alf is writing up, from his notes, the minutes of a meeting that he attended along with others. He writes "Alf was present. Betty was present. Gary was present…". Alf knows that Alf is his name. Thus he knows that in writing "Alf was present" he is thereby self-referring. However, Alf records these minutes in a very cool-headed, detached, state of mind. He has at the time of writing no particular interest in the question of whether he himself was among the attendees. He simply expresses in writing the thought that Alf was present. Since he does know that he is Alf, he is in a position to draw the consequence that he himself was present. However, he need not do so, and given his present interests, he does not do so. It is not a first personal thought that he expresses but a third personal thought. Nonetheless he does, in detached amusement, note the semantic fact that he is referring to himself in writing "Alf was present".

The possibility of such a case appears to show that it is not sufficient for the linguistic expression of a first person thought that one should know that one is thereby self-referring.

It might be thought that the case threatens the Intention View no less than the Knowledge View. Does not Alf intend to self-refer using "Alf"? He is intentionally using a word, which he knows refers to him. If so, then the case shows that using language with the intention to self-refer is also not sufficient for the linguistic expression of self-conscious states of mind.

In fact this worry does not stand up to scrutiny. The element of "detachment" that makes it plausible that Alf does not express a first person thought also makes it plausible that Alf does not intend to self-refer. In order to see why, one should observe the distinction between what is intended and what is merely a foreseen side effect of doing as intended. For example, suppose that I intend to indulge my taste for fast food. I believe that I will thereby put on weight. Do I thereby intend to put on weight? No, putting on weight is a foreseen consequence of acting as intended; but it is not itself an aim I have in acting.

Is there a way of testing for this distinction? An intuitive test is to ask what would have happened under counterfactual changes in the agent's beliefs. Suppose I had come to believe that indulging my taste for fast food would not in fact result in my putting on weight. Would I have sought alternative means of putting on weight? No. But if I had genuinely intended, and not only expected, to put on weight, then I would have sought alterative means upon learning that indulging my taste for fast good would not contribute to that in the end.

In the case as described, self-referring is known to Alf to be consequence of using the name "Alf." The intuitive test makes it plausible that self-referring is not an aim Alf has in using the name "Alf." Suppose, counterfactually, that while writing up the minutes Alf had come to believe that "Alf" was not his own name. Would he have sought alternative means of self-referring? No. Insofar as he is giving expression to a detached third personal thought about the participants he would go on to use "Alf" just as before, unmoved by the belief that he is no longer thereby self-referring.

Suppose, instead, that upon learning that his name was not "Alf" but, say, "Omar," Alf would have chosen to write "Omar was present" instead. That counterfactual truth is evidence that it was a genuine aim of his, in his actual use of "Alf," to self-refer. But it is equally evident that he was not really thinking of the participants of the meeting in a detached, third personal, state of mind. Why the interest in self-referring?

The examples of Gertrude, Edmund, and Alf, each demonstrate the superiority of the Intention View to the Knowledge View. The linguistic expression of first person states of mind seems to march in lockstep with the use of language with the intention to self-refer. On the other hand, the knowledge that one is self-referring in using language, while no doubt a typical accompaniment, is neither necessary nor sufficient for the linguistic expression of first person thought.

5 Linguistic Expression

It is time to consider some potential counterexamples to the Intention View. While ultimately ineffective, these cases will allow us to get clearer about the notion of linguistic expression as it features in the view.

> Bert is hiding in a foxhole. The enemy search party approaches the foxhole. Bert thinks the first person thought that he doesn't want to die. He has no intention to speak his mind, for he does not want to give away his position. However, he loses control of himself. Bert blurts out aloud "I don't want to die!"

Why might this be thought to be a counter-example to the Intention View? Suppose that Bert linguistically expresses his first personal state of mind when he produces the words "I don't want to die." Then the left-hand side of the biconditional Intention View is true. Now suppose that it is agreed that Bert does not intend to speak at all. A fortiori he does not intend to use words to refer to himself. Then, the right-hand side of the biconditional is false. The Intention View is mistaken to claim that one linguistically expresses a first person thought only if one uses language with the intention thereby to self-refer.

In order to see the way through this case, the defender of the Intention View should note first that the case is underdescribed in an important respect: The mechanism of speech production is not specified. Suppose that the case works as follows: The intense shame of entertaining such a cowardly thought as the thought that he does not want to die causes Bert to have a seizure. The neural mayhem randomly stimulates his motor cortex in such a way that he vocalizes the English sentence "I don't want to die."

In such a case, it is perfectly obvious that Bert does not act with the intention to self-refer.On the other hand, his noise-making is not an intentional action at all. To that extent, his noises cannot seriously be regarded as the linguistic *expression* of his state of mind. Of course, one could define the notion of "expression" so loosely that any causal effect of a state of mind counts as its expression. But in that sense there can be no interesting connection between kinds of mental state and their linguistic expression: Any subject could be wired up so that arbitrary states of mind cause arbitrary vocalizations. Claims like the Knowledge View and the Intention View are of interest only if linguistic expression is taken to be a kind of rational intentional activity. Events of speech and writing are events for which it is appropriate to ask what were the reasons, purposes, intentions of the agent's so acting.

So let it be supposed that Bert does not have a seizure but genuinely acts in using the sentence "I don't want to die." The case still appears to be troublesome for the Intention View, for then he *is* engaged in linguistic expression when ex hypothesi he did not intend to speak at all.

However, we need to distinguish prior intention from intention in acting. The case is one in which Bert had no immediately prior intention to self-refer. It does not follow that in speaking he has no intention to self-refer. If his spontaneous action is a rationalizable action at all, then it can sensibly be asked what were his intentions in acting, even if, as with other cases of sudden or spontaneous action, these intentions do not reflect any prior process of explicit deliberation on the part of the agent. Now, suppose Bert is asked during a reflective debriefing session after the event: Why did you use the word "I"—rather than the word "you," or the word "Hitler," or the word "sausages." His plausible answer, revealing of his intention in action, would be: “to speak of myself.” Why else would he have chosen the word “I” to express his thought?[10]

A different kind of potential counter-example to the Intention View involves, not unplanned outbursts of speech, but rather more calculated linguistic action. The case is another opportunity for clarification of the notion of linguistic expression.

> Winnie has beaten Louis in a two-player game of cards. A third party, interested in the outcome of this game, puts to Winnie the question: so who won? Winnie is a very modest person, and is averse to any explicit first personal bragging about herself. So she does not linguistically express her first personal thought that she herself won by means of the English words “I won”. Instead—it being common knowledge that Louis was her sole opponent—she chooses a more oblique strategy of answering the question, uttering the words “Louis lost”.

The potential difficulty for the Intention View is as follows. It might be thought that, in the circumstances, Winnie intentionally linguistically expresses the first person thought that she herself won the game. However, she does not use language with the intention to refer to herself. The only referring expression she chooses to use is the name “Louis,” with the intention of referring to Louis. Thus, one might conclude, the Intention View is wrong to claim that the linguistic expression of first person states of mind requires one to speak with the intention to self-refer.

The right response to this example is, again, to get clearer about the notion of linguistic expression. On an occasion on which a speaker uses words expressively, there may be any number of thoughts causally antecedent to the action. It is not plausible to say that the speaker gives voice to each one of these thoughts when he speaks. It is not even plausible to say that the speaker gives voice to all those thoughts that he expects an audience, in the circumstances, to be able to figure out that he thinks. The direct linguistic expression of a thought is only one such communicative strategy.

Can we say more about the identification of those thoughts that are genuinely expressed by a speaker on an occasion—as opposed to those insinuated or implicated by other communicative means? This is not the place for a proper treatment of this large question, but expectations on the part of the speaker, about what an audience would know on the basis of his utterance, must still be central. However, the relevant expectations here are in a certain sense purely semantic ones: One linguistically expresses the thought that p when one expects one’s audience to know that one speaks truly if and only if p—where the audience’s knowledge of this biconditional is expected to derive solely from what the audience would know about the (interpreted) words one uses.[11] The basic, and intuitive, idea is that one gives voice to those thoughts that one communicates by just exploiting the meaning of one’s words—as opposed to other contingencies about the situation in which one speaks.

[10] See Hornsby and Stanley (2005) for more evidence that even fast and spontaneous speech involves rational word choice.

[11] See Rumfitt (1995) for a compelling and detailed defense of roughly this approach.

How does this work in the present case? When Winnie produces the sentence "Louis has lost" she speaks in the expectation that her audience will know that she speaks truly if and only if she herself has won. However, the audience's knowledge of this truth condition is not expected to derive solely from their knowledge of the meaning of the name "Louis," the predicate "_ has lost," and their mode of combination. While she expects that they will know that she speaks truly if and only if she herself has won, this is only because she expects them to bring to bear some supplementary circumstantial information about the character of the game played, viz., that Louis has lost if and only if she herself has won. All that she can expect them to know solely on the basis of knowledge of the words she uses is that she has spoken truly if and only if Louis has lost. Thus, it is the thought that Louis has lost which is the thought she linguistically expresses. This is the thought she actually puts into words.

In order to linguistically express her first person thought, then, Winnie must overcome her modesty. She must speak in the expectation that her audience will know that she has spoken truly if and only if she herself has won, where she expects them to know this just on the basis of what is known about the semantic values of the words she chooses. And in order to do this, she will need to choose a word she thinks is commonly known to refer to her. She could, in the circumstances that everyone thinks that "Winnie" is her own name, speak a la De Gaulle and produce the sentence "Winnie has won." Or, more likely, she will just choose to use the first person pronoun "I," a word typically known in English conversation to be a device of self-reference. Either way, she can give voice to her first person thought only by choosing to use a word to refer to herself. So, just as the Intention View claims, the linguistic expression of first person thought requires speaking with the aim of self-referring.

I will close by considering a final potential counterexample to the Intention View. The right response to this example further illustrates the notion of linguistic expression, and provides an opportunity to clarify the conditions for intending to do something.

> Arthur is highly influenced by Schopenhauer. He is especially persuaded of the following pronouncement: "that the subject should become object for itself is the most monstrous contradiction ever thought of". Accordingly he develops some strange theoretical views about the pronoun "I". He believes that it is not a device whereby each speaker x may refer to x—for that would be a monstrous contradiction. Rather, each speaker x is a fusion of two proper parts, a body and a spirit, and when a speaker x uses "I", x refers not to x but to the spirit of which x has a form of introspective awareness. He has a correspondingly deviant view about the semantics of predicates such as "_ is making a mess". He believes that the predicate is true of an individual just in case the individual is the spiritual part of a person who is making a mess. Arthur, in all other respects, and outside his study, is perfectly well immersed in the ordinary communal use of the word "I" to give voice to first personal states of mind. For example when he realizes that he himself is making a mess, he will, without any second thoughts, say "I am making a mess".

The difficulty for the Intention View arises as follows. While it is questionable whether it is a condition of intending to ϕ that one must believe that one will *succeed* in ϕ-ing, the following weak belief constraint at least seems plausible: one intends to ϕ only if one does not believe that it is impossible for one to ϕ. Though we may idly wish for what we think is impossible, we cannot intend to do what we think is impossible.

Arthur believes, on the authority of Schopenhauer, that reflexive reference is impossible. According to the weak belief constraint, then, Arthur cannot intend to reflexively refer. Nonetheless, in virtue of his practical immersion in the ordinary use of "I," he does linguistically express his first person states of mind. So it is not necessary for the linguistic expression of first person states of mind that one use language with the intention to reflexively refer.

What is the right reaction to this case? It seems to me that the case calls not for the abandonment of the Intention View but for the refinement of the belief constraint upon intending. Let it be agreed that Arthur's theoretical opinions about the nature of first person reference are sincerely held. The right thing to say about the case is that in so far as Arthur, outside his study, is using the word "I" with the rest of the community to express first person thoughts, his theoretical opinions must in a certain sense take a back seat.

For let it be supposed that on an occasion when he thinks that he himself is making a mess, Arthur says "I am making a mess" while keeping firmly in the forefront of his mind his opinions about the impossibility of linguistic self-reference. Can he linguistically express his first person thoughts while consciously in the grip of this philosophical opinion?

Recall, for Arthur linguistically to express his thought that he himself is making a mess, he must produce an utterance in the expectation that the audience will know that he has spoken truly if and only if he himself is making a mess, where their knowledge derives solely from knowledge of his interpreted words. In order to determine whether this condition is met, we need to distinguish two variants of the case. In one case, Arthur believes his community shares his semantic opinions. In another case, Arthur believes that he is the only enlightened one; he thinks that the rest of the community mistakenly believes that "I" is a device whereby x may refer to x. In both cases, his present and lively opinion is that "I" does not, and cannot, refer to its speaker. It refers to the spiritual proper part of the speaker.

Take the first variant. Does Arthur, in using "I," meet the conditions for linguistically expressing his thought that he himself is making a mess? In this variant, what Arthur can expect his audience to know solely in virtue of their understanding of his words is that he speaks truly if and only if *this spirit* is part of a person who is making a mess. For, they share his deviant opinions about the semantics of the term "I" and the predicate "is making a mess." He may well expect them to be in a position to draw the further conclusion—perhaps on the basis of the supplementary metaphysical view that each person is able to demonstrate introspectively only his own spiritual part—that he speaks truly if and only if he himself is making a mess. But the thought he puts into words is a thought about his spirit. He does not express his first person thought, even if he would appear superficially to observers with our understanding of "I" to be someone expressing first person thoughts.

The second variant is less clear-cut. This is in effect a case in which Arthur does not expect his audience to know the real meaning of his words. He does expect them to believe that he speaks truly if and only if he himself is making a mess, but he expects them to believe this only because he expects them to have the beliefs—false beliefs, as he sees it—that the term "I" refers to the speaker, and the predicate "is making a mess" is true of individuals making a mess. But this feature of the case makes it implausible to say that he is genuinely putting his first person thought to them. As Arthur sees it, only linguistic incomprehension on the part of his audience leads them to believe that he speaks truly if and only if he himself is making a mess. But one cannot be said to express a thought to an audience if one speaks in the expectation that they do not know the meaning of one's words. In so far as Arthur, in his splendid isolation, genuinely expresses a thought, then he speaks either to himself, or to an imagined audience who would understand his words. In either case, the second variant collapses into the first variant of the case, the variant in which Arthur believes his audience to understand the meaning of his words. But in that variant, as explained, Arthur expresses a nonfirst personal thought, about his spirit.

The foregoing shows that if Arthur's opinions about the impossibility of self-referring with "I" are fully engaged when he speaks, then it is to say the least unclear whether we really have a case of the expression of first person thought. Of course, a more likely scenario for someone with Arthur's deviant philosophical beliefs will involve a character whose seminar-room opinions do not really impinge upon his psychology when he is immersed in the practice of ordinary speech. A rationalization of his habitual linguistic actions will ascribe to him, as much as to anyone else, the intentional choice of a word of reflexive reference when expressing first person states of mind.

This in turn shows, what should be clear anyway, that the weak belief condition on intention is at best an idealization. Perhaps it is really a psychological impossibility to form the intention to ϕ while at the same time clear-headedly endorsing the thought that ϕ-ing is impossible. However that may be, if an opinion about the impossibility of ϕ-ing is not engaged and occurrent at the moment of action, only manifesting itself in theoretical reflection, then there is no obstacle to acting with the intention to ϕ. Arthur is, at worst, mildly irrational in so far as he is caught up in the expression of first person thought in the usual way while at the same time retaining the philosophical opinion that reflexive reference is impossible.

So the case is not a counterexample to the claim, made by the Intention View, that the linguistic expression of first person thought requires the intention to self-refer. If Arthur's deviant opinions *are* operative at the time of speech, then he does not express first person thoughts even if he is superficially conformal with those who do express first person thoughts. If his opinions are *not* operative at the time of speech, then they are no obstacle to his intentionally self-referring.

I began with the question of how self-conscious states of mind are expressed in language. The simple view that such states of mind are exactly those expressed by uses of first personal language was shown to be a mistake. Of the two more promising accounts of the conditions for linguistic expression of self-conscious states of mind, the Intention View has emerged as superior. The Knowledge View, for all its eminent advocates, seems not to be correct.

Acknowledgments For comments and discussion, thanks to Ray Buchanan, Hans Kamp, Daniel Morgan, Lucy O'Brien, Lea Salje, and participants in the Taipei IEAS conference.

References

Anscombe, G. E. M. (1957). *Intention*. Oxford: Blackwell.
Anscombe, G. E. M. (1975). The first person. In *Collected philosophical papers: Vol. 2. Metaphysics and the philosophy of mind* (pp. 21–36). Oxford: Basil Blackwell.
Evans, G. (1982). *The varieties of reference*. Oxford: Clarendon Press.
Hornsby, J., & Stanley, J. (2005). Semantic knowledge and practical knowledge. *Proceedings of the Aristotelian Society, Supplementary Volumes, 79,* 107–145. http://onlinelibrary.wiley.com/doi/10.1111/j.0309-7013.2005.00129.x/abstract.
Martin, M. G. F. (2001). Out of the past: Episodic recall as retained acquaintance. In C. Hoerl & T. McCormack (Eds.), *Time and memory: Issue in philosophy and psychology* (pp. 257–284). Oxford: Clarendon Press.
Nozick, R. (1980). *Philosophical explanations*. Cambridge: Harvard.
O'Brien, L. (2007). *Self-knowing agents*. Oxford: Oxford University Press.
Peacocke, C. (2008). *Truly understood*. Oxford: Oxford University Press.
Romdenh-Romluc, K. (2008). First-person thought and the use of "I". *Synthese, 163,* 145–156.
Rumfitt, I. (1994). Frege's theory of predication: An elaboration and defense with some new applications. *Philosophical Review, 103*(4), 599–637.
Rumfitt, I. (1995). Truth-conditions and communication. *Mind, 104*(416), 827–862.
Setiya, K. (2003). Explaining action. *Philosophical Review, 112*(3), 339–393.
Velleman, J. D. (1989). *Practical reflection*. Princeton: Princeton University Press.