***Agency and the First Person***

‘You cannot understand the capacity for personal agency and the capacity for first personal thinking separately from one another.’ One version or another of this claim has been made in philosophy often, and for a long time. In this paper I want to proceed by stepping back and asking why one might think that – whatever the particular details of distinct claims of dependence – that the problem of agency and the problem of first person thinking might be philosophical bedfellows. Why should we think that the illumination of one will go along with the illumination of the other? In this paper I want to suggest that understanding agency as the power of reflexive change helps us to understand the role it plays in our capacity for first person thought and reference.

*(i) Reflexivity* *problems*

One place to start, in answering the question of how the problem of agency relates to the problem of first person thought and reference, is to see that both – alongside other perennial philosophical problems – are in a category of what we can call intentional *reflexivity* problems.

Intentional reflexivity problems tend to have a common structure and development. We start with various forms of puzzling self-directed intentional relational phenomenon of:

self-consciousness,

self-knowledge,

self-directed thought and speech,

self-change,

self-love,

self-concern.

We then consider the relations that are involved, and what it is for a subject to stand in such a relation to an object. We consider relations such as the relation of being conscious of, having knowledge of, thinking and speaking about, changing, loving, caring about – and we try to understand what it is for some X to be conscious of, know of, speak or think about, change, love, care for – some non-identical Y. We make philosophical progress – or not – in clarifying what the nature of the general relation is, and how we as subjects stand in those relations to objects. Then, having some idea of the nature of the relation, and how two relata can be so related, we ask how we should think about the case where the relation obtains *reflexively*: we consider the case where the relational obtains but where the relata are identical, where X is identical to Y.

Put in a schematic form:

1. We consider a range of intentional phenomenon: φ1, φ2,…φn.
2. We ask what it is, in each case, for a subject X to φi Y.
3. We ask what it is, in each case, for a subject X to φi X.

At this point two very different types of case in which X is identical to Y emerge.

First, we have the *internally* reflexive case. We have the case in which for a subject X to φi X the subject X relates to themselves, *as the subject* of the phenomenon at issue. In such cases we can talk about the relation as a ‘subject-reflexive relation’: X speaks or thinks about herself as the *subject* of the speaking or thinking, refers herself as the referrer, changes herself, as the changer, loves herself, as the lover, cares for herself, as the carer, knows herself as the knower, perceives herself as the perceiver and so on.

Second, we have the *externally* reflexive case. We have the case in which the subject relates to herself reflexively in virtue of facts external to her φ1–ing – whether contingent or necessary – that secure her as the object of the relation. In these cases the relation is an ‘object-reflexive relation’: X speaks or thinks about herself as an object of her speaking or thinking, refers herself as an object referred to, changes herself, as an object changed, loves herself, as an object loved, cares for herself, as an object cared for, knows herself as the object known, perceives herself as an object perceived and so on.

Put schematically, again, we note that there seem to be two ways for a subject X, to φi X:

1. A subject X, can φi X externally – when facts external to her φ1–ing – whether contingent or necessary – that secure her, X, as the object of the φ1–ing.
2. A subject X, can φi X internally – when facts internal to her φ1–ing secure her, X, as the object of the φ1–ing.

We might then ask – what is the relation between the *internally* reflexive case and the *externally* reflexive case. Are there conditions or additions, attached to the *internally* reflexive case, but not to the *externally* reflexive case, or vice versa?

Consider, for example, in Anscombe’s discussion of first-person reference, when she asks whether to say:

‘“I” is the word each one uses when he knowingly and intentionally speaks of himself’ (Anscombe 1975: 22)

is to say enough to characterise first person reference. We can take her as asking whether adding internal conditions of knowledge and intention to the conditions that the reference be externally reflexive might give us what is needed to characterise the first person in English.

We might, similarly, taking our cue from Anscombe, ask whether there are added knowledge or intention conditions that we can, in general, add to cases of intentional phenomena that are *externally reflexive*, to give us *internally* reflexive cases. We might, for example, wonder whether we can get the subject-reflexive relation from the object-reflexive relation as long as the subject also knows or is conscious that she, the subject, in fact – or indeed of necessity – is the object that she relates to? Or we might wonder whether we can get the subject-reflexive relation from the object-reflexive relation as long as the subject also aims to secure, or intends that she, the subject, in fact – or indeed of necessity – is the object that she relates to? Or both?

Thinking about things in this way is to think of the internal reflexive case as just a version of the external reflexive case, with the added knowledge or intention on the part of the subject that she is, or will be, both the subject and the object of the relation. Subject-reflexive relations are, on this picture, self-conscious, self-directed object reflexive relations. But are subject reflexive intentional phenomenon of the same kind as object reflexive intentional phenomenon but with some consciousness of, or aim directed at, X φi-ing X added on?

Obviously, as Anscombe introduced her suggestion to show, things are not that simple. Internal reflexivity – the reflexivity in virtue of which a subject relates to herself *as the subject* of the relation – will not be secured simply by *adding* a separate knowledge or intention condition. Such a condition will inevitably either be insufficient to secure subject reflexive reference, or will already have to exploit the fact that we are trying to explain: that the subject *is* the subject of the relation. The resulting account will be insufficient or circular.

Suppose that it is true that certain external facts mean that the subject stands in one of these relations to herself as object – thinking about herself, for example. Suppose it is also true that the subject knows that this is so, or that the subject aimed to make it so. This will not be enough to secure the result that that thinking about herself is transformed into a form of *internal* thinking about herself – unless already what the subject knows is that *she, as subject*, is, in so thinking, thinking about herself as object, or that she intends that *she, as subject*, is in so thinking, thinking about herself as object. But to do *that* she already needs to stand in a subject-reflexive knowledge or intention relation. She needs to know that the object she has knowledge of is herself, the subject of the knowing. Or she needs to intend that the object she has an intention directed towards is herself. She needs, in short, already to stand in a subject-reflexive relation. If she knew or intended only that *x* stands in the relation φ to *x*, but did not thereby know that the *subject of that particular knowing or intending* – viz, herself – is *x*, then we would still have an external reflexive relation only. For us to try to understand the internal reflexivity case in terms of the external reflexivity case, plus knowledge or intention on the part of the subject that the relation is realised reflexively, that knowledge or intention would already have to be understand as involving an internally reflexive relation. If the subject knew only that the subject and the object of the relation where identical, but did not know that the subject and object of the knowing *are herself,* the subject who is *now knowing* them to be identical, we would be no closer to capturing what we need. If the subject intended only that the subject and the object of the relation be identical, but did not intend that the subject and object of the intending *be* the same subject who *now intends* them to be identical, we would be no closer to capturing what we need.

It is easy to misconstrue this point as a point that again claims that self-knowledge, or self-directedness in intention, is needed to characterise internal reflexivity – the subject needs to know that *she* is the subject who *knows* the subject and object to be identical; or that the subject needs to intend that *she* is the subject who intends the subject and object of her thought to be identical. And, indeed, it is almost impossible to say what is missing without invoking those locutions – they function as markers of certain internally reflexive relations. Moreover, a subject who stands in an internal reflexive intentional relation will be a subject, if she has the capacity to know and intend such things, able to know and intend that she stands in such a relation, by realising that relation. However, putting things in terms of a claim that self-knowledge or self-directedness characterise internal reflexivity can be misleading.

We can see this is not the most helpful way to put it by noting two things. One, to say that the subject needs to know that *she* is the subject that she knows relates reflexively to herself, or that *she* is the subject who intends the subject and object of her thought to be identical, can seem to be just making the same mistake again, and meet the same problem in relation to characterising that knowledge or intention. How are we to understand ‘she’ here – as expressing a subject reflexive relation or not. Two, we have not ruled out the possibility of internal reflexive intentional relations – and with them the corresponding reflexivity problems – that come without knowledge or intentions, as they are generally understood, as markers. Suppose for example we take self-love to involve *x* standing in an intentional loving relation to *x*, and think that self-love, where that involves a subject-reflexive relation, is importantly different from self-love where that involves an object-reflexive relation. Suppose that we think of self-love as coming in two kinds. One as self-love that is realised by a subject-reflexive *affective* relation that is fundamentally reflexive – as a relation of a kind that one could hold only to oneself. Two, as self-love that is realised by an object-reflexive affective relation that is de facto reflexive, and of a kind that one could also hold to others.[[1]](#footnote-1) We may resist the claim that subject reflexive self-love involves a subject knowing or intending that she stand in the *x* loves *x* relation – we may deny that love needs the capacity to have any such attitudes to the intentional phenomenon of loving someone: you can love without the idea of love. Certainly, it is not compelling to think that what makes the subject-reflexive form of self-love distinct from the object-reflexive form is just that that knows or intends that the subject of the love be identical to the object.

It may be true is that any subject capable of *reporting* that they stand in the *x* φi*x* relation, where the subject’s so standing does not rely on external facts, but only truths internal to them standing in the *x* φi*x* relation, will be a subject who has self-knowledge and distinct self-directed intentions of the kind we are concerned. But it would be a mistake to take the only *evidence* a subject can offer that she is thinking about herself internally reflexively, as a reason to believe that she can think internally reflexively only when she is in a position to offer such evidence.

If the above is right, the problem of characterising internal reflexivity is not solved by characterising the subject as standing in an *epistemic or intentional* relation *to* the fact that *x* stands in relation R to *x*. What we need to make sense of is the possibility of there being a condition of the subject’s φ1–ing that, in itself, *realises* the fact that *x* stands in relation φ1 to *x.*

The notions of internal and external reflexivity are, primarily, metaphysical notions. As we have understood it an intentional reflexive relation can be realised in two ways – internally or externally. Those two ways have been understood in terms of ways in which the relation may be realised; in terms of what kinds of facts makes it the case that *x*φi*x.* The distinction has not been unpacked in terms of facts about any independent intentional attitude of x’s – whether epistemic or intentional – towards *x*φi*x*. The kinds of facts that make it the case that *x*φi*x* maybe external facts about *x,* and her φi-ing, which she may or may not know about or intend, in virtue of which she comes to stand in the φi relation to herself. However, the facts that make it the case that *x*φi*x* may obtain *only* in virtue of condition of the subject’s φ1–ing. If we have the latter, we have an internal reflexive relation.

Suppose we agree that the internal reflexivity case is going to need to be explained differently from the external reflexivity case plus some epistemic or intentional add-ons formulated using precisely the notions – of knowledge, or intention, for example – in relation to which the puzzle arises. What we then need is for there to be a way of characterising what it is to come to stand as the subject of the intentional relation at issue, which *already secures* one coming to stand as the object of the relation. The kind of activity or process in relation to which one has the status as subject – as referrer, speaker, thinker, knower, carer, lover etc. – must be sufficient *in itself, already,* to secure one’s status as the object – the referred to, thought or spoken of, known, cared for, loved etc.

It is this demand – that we make sense of what the subject does as subject in a way that thereby, and already, secures her position as object – is what, in my view, makes the appeal to agency in offering us a way of unpacking these matters so compelling.

*(ii) Why agency?*

What role does *agency* play in an account of the way a subject, as subject, relates to herself, that thereby, and already, secures her position as the object of an intentional phenomenon? What role does *agency* play in characterising *an internal relation a self can stand into itself?*

There is one natural way of answering these questions. When a subject exercises her power to act, her power to φ, what she does, in φ-ing, is to self-change. If a subject X, φs, for any act type φ within the range of acts available to her as immediate choices then X will, in φ-ing, herself, change in a way concordant with her φ-ing. That is not to say that the power to act is not also the power to change other things: in being able to act, the subject X, may change Y or Z, but X will only do so by changing X. If the above is right, then the power of agency is a power of reflexive change: for X to *act* is always, already for X to change X. If φ-ing is the intentional act that X has elected to enact, then in being the *subject* of her φ-ing X cannot but, herself, in φ-ing also be an *object* of her φ-ing. Agency promises to give the acting subject her status as object for free.

This claim, that in action a subject is both the subject and the object of the act, can be thought of involving four claims:

1. To act intentionally must be, oneself, to change something.
2. That which is intentionally changed is an object of the act.
3. Nothing is intentionally changed by a subject without the subject, herself, being intentionally changed.
4. The subject’s intentional changing is the change to the subject.

Taken together suggest that all intentional actions are, in a specific and fundamental way, reflexive: to act is to be both the subject, and thereby be the object of the act.

We said that internally reflexive intentional phenomenon should be understood as follows:

If a subject X φis X internally – then facts about her φ1–ing itself secure her, X, as the object of the φ1–ing.

So, if *all* actions secure the subject as also the object of the action, in virtue of facts about the being an action at all, then agency in general, seems to meet a condition we put on internal reflexivity: facts internal to any φ–ing by a subject secure the subject, herself, as the object of the φ–ing. And if *x* having φ-ed just is the change to *x* in φ-ing then we have identified a relation that has the property that any subject of that relation must already be its object.

Now we might think that there are grounds to resist idea that the intentional changes to the subject concordant with her intentional action are enough to secure her status as the *object* of the act. We may, rightly, point out that I do not – in the usual run of things – act on myself as I act on other objects. As an agent I can act *on* a range of objects. But, in my own case, it is different. We might want to say I act *with myself,* not *on myself*. I only rarely identify myself as an object, one amoung others, to be acted upon. It is rather pre-suppositional to acting that it is with myself that I will carry out any act of φ-ing chosen, and that I will thereby have φ-ed.

We also might point out that the agent does not need to identify herself as the agent to carry out the act, as one object amoung many, and that there is difference to the way in which she is secured as the object of the act, to the way on which other objects are secured as objects of the act. She is secured as the object by no more than being the agent; her status as object is implicit, given, in acting at all.

However, these observations do not mean that the subject is not the object of her acts. The difference between talk acting ‘with oneself’, and ‘on oneself’ is not a difference that relates to which object the subject intentionally changes in acting. It is rather a difference in how the object figures in the *intentional form* of the act she carries out. When no subject, other than the agent figures in the form of what is done, we identify the act types that characterise what is done as having the form *x* ϕs. When a subject acts on an object we can identify the act types as having the form *x* ϕs *y.* When we identify a case in which the subject acts on herself, we can identify the act types as having the form *x* ϕs *x.*

Now there are many ways in which a subject can bear a non-reflexive relation to act-types - *x* ϕs, or *x* ϕs *y*, or *x* ϕs *x.*  A subject can, for example, believe that Bob enacts act types of the form *x* ϕs, or *x* ϕs *y,* or *x* ϕs *x,* and a subject can, for example, believe that many people enacts acts types of the form *x* ϕs, or *x* ϕs *y,* or *x* ϕs *x.*  Or a subject can intend that Bob, or many people, enact acts of the type *x* ϕs, or *x* ϕs *y, or x* ϕs *x.* However, if the subject *enacts* act types of the form *x* ϕs, or *x* ϕs *y,* or*, x* ϕs *x* then what she does, perforce, has a form of reflexivity.

Nevertheless, there is a special kind of reflexivity exhibited in enactions of the act-type *x* ϕs *x.*

Suppose we now ask whether, if what is intentionally done by a subject is correctly characterised as *x* ϕs, we can infer that what the subject does is also correctly characterised using the form *x* ϕs *x.*  I think the answer to this question brings to light an ambiguity.

On one understanding of the question the answer seems obviously to be ‘yes’. If S moves, S moves S, if S raises her arm, S raises S’s arm. The action that is carried out by enacting an act type *x* ϕs, will always be a case of *x* ϕing *x.*

Moreover, a reflexively self-conscious subject who intentionally enacts the act type *x* ϕs can also be taken to intentionally enacting the act type *x* ϕs *x*. She can always know that enacting an act type of the form *x* ϕs, is also enacting an act type of the form *x* ϕs *x.*  However, this is not, to say that act types of the form *x* ϕs, do not have a different form from act types of the form *x* ϕs *x.* A subject may be unable to conceptualise the reflexive form *x* ϕs *x,* but still able to enact an act type of the form *x* ϕs, and in so doing bring about a reflexive change.

What distinguishing the two ways of acting reflexively enables us to see is that if we combine (i) the kind of reflexivity inherent to all exercises of agency, with (ii) the power to enact act types that have a reflexive form – *x* ϕs *x* – we get a way of characterising a distinct form of reflexive intentional action. It is easy to make the mistake of thinking that one form is reflexive and the other not. But that is not right – one is what we might call reflectively, or first personally, reflexive and the other pre-suppositionally, or primitively, reflexive.

A first step in understanding reflexivity problems is to realise that there are at least three kinds of reflexivity that interact:

Subject-reflexivity: reflexivity in enacting any act type, *x* ϕs

Formal object-reflexivity: reflexivity in enacting an act type of the form xRx

De facto object-reflexivity: reflexivity in enacting an act type of the form xRy, when in fact x=y

We get first personal reflexivity when S enacts an act with the form xRx.

*(ii) First person thinking*

Let us now consider how these remarks about agency and reflexivity relate to first person thinking. First person thinking is intentional reflexive thinking. It is reflexive thinking of a kind that the subject of the thinking relates to herself as the object thought about, in virtue of being the subject of such thinking, and not in virtue of other external facts that she may or not stand in a subjective relation to. But what is it to do that?

To explain such thinking can seem to be impossibly hard. So, hard that some think that we can only be sceptics either about its existence or its explanation. Consider, again, Anscombe’s unpacking of our difficulty. First person thinking cannot merely be said to be reflexive thinking, thinking in which a subject thinks of herself. For there are many cases in which a subject thinks of herself, but does not think of herself first personally. I can think of myself when I think that NN is on earth, but to think that may not be to think of myself first personally. There can, in the terminology we used above, be cases of externally reflexive thinking in which the subject stands in an object-reflexive relation. Such thinking is not first person thinking. What we need to capture the nature of first person thinking seems to be internal reflexive thinking, where a subject stands in a subject-reflexive relation. But we cannot, as we saw, ‘upgrade’ object-reflexive thinking to the kind of thinking that we are after by insisting that it is also *known* reflexive thinking or *intended* reflexive thinking in relation to which the subject has the appropriate knowledge or intention – thinking in which ‘a subject knowing and intentionally thinks of herself’. For there are many ways of knowing and intentionally thinking of oneself that do not characterise the first person reflexive way of thinking. I knowingly think of myself when I think NN is in on earth. After all I know that I am NN, but that is not to think of myself in the first person way. Even if I in fact know that I am NN, and so intend to refer to myself in thinking ‘NN is in on earth’ the kind relation of ‘thinking of’ that is characterised as ‘thinking of NN’ is a kind of ‘thinking of’ relation that is consistent with my thinking of that which is in fact, or perhaps necessarily, myself, while failing to realise that I am both relata: the subject that is the thinker of, and the object that is the thought of. If I insist that the upgrade involves not any old *known* reflexive thinking or *intended* reflexive thinking, but a thinking on the part of the subject not only *of* herself, but *as* herself, we face circularity. The condition that the subject think of herself as herself can only be a condition that she thinks of herself first personally.

What we need is a way of characterising, in a non-circular way, internally reflexive thinking: thinking that has a form that does, in so thinking a thought with that form, put a subject in a subject-reflexive relation. As Nozick put it in his discussion of first person reference in speech, also using the spatial metaphor we used earlier, the reflexivity that characterises the kind of thinking we are concerned with is ‘internal to’ that which we aim to explain:

For a person X to reflexively refer is not merely for X to use a term that actually refers to X; this omits as internal to the act of referring that it is himself to which he refers.” (Nozick 1981, p. 71)

So, what is enough to give us internal reflexive reference in speech? Not only reflexivity – for X to refer to X – as we saw. Not even, as Nozick points out, rigid, and necessary reflexivity. (He considers self-reference secured by Godel numbering.) To get internal reflexive referring, the referring reflexively must be secured only by the act of referring to the referring subject. You get internal reflexive reference in speech when the instance of referring produces itself:

The reference is peculiarly internal to the act of referring since that act refers in virtue of a feature created by or produced in that very act itself…It follows from its sense that the term “I” refers to the producer of that very token and that the person is referred to in virtue of the property he acquires in the very act of referring or producing the token, the property of being the producer of that token, and that the person is referred to in virtue of the property he acquires in the very act of referring or producing the token. (Nozick 1981, p.75)

For any subject to internally reflexively refer in a speech act, the act itself – not any distinct attitude the subject has to the act – needs to be an act of reflexive reference. Moreover, it needs to be an act of the form ‘*x* refers to *x*’ where the value of ‘*x*’ is fixed simply by the agent of the act. The pronoun “I” is a device of reflexive reference – it is for referring to the agent of its use. Any agent who understands it as a device of reflexive reference can therefore use it in their act of reflexive reference. And, in knowing others understand it in the same way, they can use it to make their act of reflexive reference public.

However, such a possibility depends upon there being a general fact about agency: that in enacting any act-type the subject is also thereby the object: that in acting a subject alters themselves. Two things come together. First, the subject in intentionally using “I” – a term for any x to use for x to refer to x – realises a general act-type of ‘x refers to x’. Second, in realising any general act-type an agent makes it the case that they are the object as well as the subject of the act realised. If they enact the act type ‘x jumps’ they will be thereby be an object that rises. And, similarly, if they intentionally use “I” they enact an act type ‘x refers to x’ and will thereby be both the agent and the object reflexively referred to.

Moreover, this is a possibility that has been realised in a way that is manifest to both speaker and hearer. Both can know that the act-type ‘x refers to x’ has been realised by the speaker in her first person speech act. That is how, simply by being the person saying the word ‘I’ you are thereby the person spoken of self-consciously, and understood to be so by others.

We need to be a bit careful how we specify what it is to be the person saying ‘I’. It is not enough that that I am, for example, unintentionally the cause of a sound that would be identified as “I” by hearers. What the subject is intentionally doing in saying ‘I’ must be that she is enacting the act-type ‘x refers to x’. However, once we she acts the form of her act-type, and the reflexive nature of agency does the rest. As Nozick puts it the ‘act refers in virtue of a feature created by or produced in that very act itself’.

It might, prima facie, seem that this account for first person reference in speech cannot carry over very effectively to the case of thought. Nozick’s talk of ‘terms’, might for example, seem to bring with it a category mistake if we think about terms produced in thought. There is no *word* “I” the thinking of which in thought produces a feature or property – being the producer of the term – which, given the sense of the term, makes me its object, because there are no *words* in thought.

However, the worry, is not in the end be a fundamental one. We can see this coming at it from two directions.

First, speakers refer by speaking, in virtue of the possibility of speaking being an intentional act with a particular form. No production of a word can count (non-derivatively) as an instance of a speaker referring using that word unless the speaker (cf. Madden) is comprehendingly using that word with the intention of doing with that word what its meaning allows it to do. Speaking and referring are in, all core cases, themselves forms of thinking - albeit thinking with words. We would expect to find a problem with a concordant account of first person thinking only if it proved impossible to accord instances of thinking any intentional form without appealing to the use of words. But, if we found that, that would be a form of scepticism about unspoken thought, and that would mean that we could solve the problem of first person thinking by solving the problem of first person speaking, because first person thinking would be first person speaking.

Second, assuming we are *not* sceptical of the possibility of thinking without speaking, and think that there *is* a problem of first person thinking to be solved, then we will have to acknowledge that while we may not think with words our thoughts – in order to be thoughts at all – must have an intentional form such that it makes sense for us to say that a subject comprehends her own thoughts. And the form of first person thinking must be such that the thinking of it secures the thinker as the object the thinking is about. In both cases – of first person thinking, and of first person thinking without speaking – what we need for ‘internal’ reflexive reference is for the form of the subject’s thinking or speaking be such that in so speaking or thinking – with that form – she is thinking or speaking of herself. That is, the form of her thinking or speaking must be of the form ‘x refers to x’. What I have argued is that, given the reflexive nature of agency, if her thinking and speaking is her intentionally enacting a general act-type with the form ‘x refers to x’ – irrespective of the means by which she does it – she will be referring to herself personally in the fullest sense.

References

Anscombe, G.E.M. (1975) ‘The First Person’, in Guttenplan, S (ed.) *Mind and Language,* Oxford: Clarendon Press

Nozick, R. (1981), *Philosophical Explanations,* Cambridge: Harvard University Press

Wollheim, R. (1984). *Thread of Life*. Cambridge University Press.

1. Some call the former self-concern and the latter self- love. (See Wollheim 1984, Chapter VIII). [↑](#footnote-ref-1)