Ergo AN OPEN ACCESS JOURNAL OF PHILOSOPHY

Selfish Reasons

KIERAN SETIYA
MIT

It is a commonplace that we have reason to care about ourselves, to pursue our own interests, to do what benefits us, and to save ourselves from harm. What is contentious is the strength of these reasons, and what else we have reason to do. Details aside, most would accept the following claim:

> SELF-CONCERN: The fact that an event will benefit or harm me is a reason for me to want, or not to want, that event to happen. This reason derives from the effects of the event on my well-being, not its effects on anything else. And its force as a reason turns on its first person character.

This principle can be clarified in three ways. First, that benefit and harm provide me with reasons of this kind is a generic proposition, which may admit exceptions. The fact that I would benefit from an immoral act may be a reason to prefer it that is outweighed by moral objections, or it may be silenced by morality, so not a reason at all.[1] Both views are consistent with Self-Concern, which speaks to the ordinary case. Second, the reasons in question do not derive from the instrumental significance of my well-being for some further end: the general happiness, say, or the fulfillment of my obligations. Benefiting me may contribute to the greater good, and harm may prevent me from acting as I should. If these facts are reasons for preference, they are not reasons of self-concern, whose normative explanation stops with my well-being.[2]

Finally, reasons of self-concern are facts about me, not just in being facts about Kieran Setiya, this particular human being, but because they are reasons in which I am represented as myself. One way to bring this out is to imagine that I have forgotten my name. As I wake in the hospital ward, surrounded by other patients, I notice a medical chart that indicates a painful operation for Kieran Setiya later

1. On the silencing of self-interest by morality, see McDowell (1979).

2. Self-Concern is thus opposed to desire-based theories on which facts about my well-being provide me with reasons only if I care about what happens to me, as in Williams (1979: 21).

http://dx.doi.org/10.3998/ergo.12405314.0002.018

that day. I feel sorry for him and I hope he does not suffer too much. The doctor then walks by and informs me that I will have the same operation. I shudder in dread, much more distressed than I was before. Self-concern is addressed to my well-being from the first person perspective: to my well-being considered as mine, not just that of one among many. The fact that Setiya will be harmed may be a reason for me, as for anyone, to care; the fact that *I* will suffer goes beyond this.

My topic is the rationality of self-interest and the truth of Self-Concern. More broadly, it is the significance of first person thought for practical reason. I will argue that Self-Concern is false, and that the justification of self-interest is not essentially first-personal. Self-interest involves an attitude of love towards oneself that is justified in the same way as love for anyone else.

In recent philosophy, the question of Self-Concern has been approached through the metaphysics of personal identity. Philosophers ask what it is for a future event to harm or benefit me: what it is for me, now, to be identical to the person who suffers or benefits at a future time. In a seminal discussion, Derek Parfit held that, given the true metaphysics of our persistence through time—a "Reductionist" view of what it is for a future person to be me—it is rational to reject Self-Concern. According to what he calls "the Extreme Claim," if Reductionism is true, the fact that a future will be mine is not a reason to care about what happens in it. Parfit believes that the Extreme Claim is defensible, though it can also be defensibly denied (1984: 310–11).

Although it has been extensively discussed by others, I think it is worth returning to Parfit's book. There are questions of interpretation in his argument that remain obscure, even in more recent work, and the fundamental problem with his approach is not well-understood. As I argue in Section 1, what matters to the rationality of self-concern is not what it is for an event to benefit or harm me, but what I am thinking when I believe that it will. The metaphysics of personal identity may tell us what it is for my thought to be true or false; it does not specify the content of that thought: what I believe when my belief is about me. This reflects a general point about reasons, and so about the methodology for investigating Self-Concern: when the fact that *p* consists in the fact that *q*, the proposition that *p* may be distinct from the proposition that *q*; and it is propositions that are the objects of attitudes and the reasons for action and desire.

In order to make progress with Self-Concern, we must turn from the metaphysics of persistence to the metaphysics of thought.[3] Instead of asking what it is for a future event to harm or benefit me, we should ask what it is for me to *think* that it will harm or benefit me, conceived as the first person. What is it to believe that *I* will suffer through a painful operation, as opposed to believing that Setiya

3. I set aside here the metaphysics of solipsism and its contemporary descendents, as in Fine (2005: §12), Hare (2009).

will? And why should this difference matter? In Section 2, I argue that first person thought is thought that is about its object in virtue of its being the object of immediate knowledge, which includes the knowledge we have of our own intentional actions. It is through this relation to myself that I think about myself in the first person. In Section 3, I draw out the implications for practical reason, disputing Self-Concern. The content of the first person does not justify non-instrumental interest in myself or the events that benefit and harm me. I end by sketching an alternative conception of self-interest, as a form of self-love, and by considering, briefly, the wider significance of first person thought.

1. Reductionism

On Parfit's explicit definition, Reductionists claim "(1) that the fact of a person's identity over time just consists in the holding of certain more particular facts" (1984: 210). They may also claim "(2) that these facts can be described without either presupposing the existence of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists" (Parfit 1984: 210). According to the first claim, there is a non-circular completion of the formula,

> What it is for person X, at *t*, to be identical to Y, at *t**, is for such-and-such conditions to obtain.

The right-hand side of this principle does not appeal to facts that are themselves explained in terms of personal identity over time. According to the second, stronger claim, it does not appeal to facts about people and their properties at all: it is thoroughly impersonal.

Like many, I struggle to make sense of the final thought. Parfit's strong Reductionist explains what it is for people to exist, and to persist through time, in terms of mental states that are not ascribed to subjects.[4] But the only grip I have on the concept of a mental state is through that of a mental property—intending, believing, desiring—that must be a property of someone. So I do not understand the conception of mental phenomena on which the reduction rests. For this reason, and because it makes no difference to the arguments below, I set claim (2) aside.

For our purposes, what figures on the right-hand side of the formula is also insignificant. Parfit contrasts a "Physical Criterion" of personal identity as the spatiotemporal continuity of a functioning brain with a "Psychological Criterion" that turns on psychological connections of memory, intention, belief, desire, and

4. See Parfit (1984: 223–5, 250–1).

character (Parfit 1984: §§77–78). Psychological connectedness consists in being related by a sufficient number of such connections. Psychological continuity is the ancestral of psychological connectedness. According to the

> PSYCHOLOGICAL CRITERION: What it is for me, now, to be identical to X, at *t*, is for us to be related by non-branching psychological continuity, with the right kind of cause.

"Non-branching" means that there is no time between now and *t* at which more than one person is related in this way to either of us.[5] Following Parfit, I will assume that Reductionists accept the Psychological Criterion. My argument is unaffected by problems of circularity in this criterion—is personal identity presupposed by connections of memory?—and by cases that conflict with it.[6] The issue is about Reductionism in general. Adopting a specific criterion helps to prevent our discussion from being too abstract.

The crucial point is that, for Parfit, Reductionism has practical implications, implications for "what matters in survival" (1984: 298). According to Parfit, "*Relation R* is what matters" where "R is psychological connectedness and/or continuity, with the right kind of cause" (Parfit 1984: 262). Although it draws on elaborate consideration of cases, Parfit's main contention is straightforward. On the Psychological Criterion, "personal identity just consists in the holding of relation R, when it takes a non-branching form. If personal identity just consists in this other relation, this other relation must be what matters" (Parfit 1984: 262–3). It is this reasoning I want to address.

The first step in doing so is to sort out an ambiguity in "what matters" that is unresolved in *Reasons and Persons*.[7] "What matters in survival" could refer to reasons of self-concern: the non-instrumental interest we have in what benefits and harms us in the future. Or it could mean what we want, or ought to want, from survival, what would make survival good. At times, Parfit clearly intends the former, asking "What is the relation that would justify egoistic concern about this resulting person?" (Parfit 1984: 283) At others, he suggests the latter, as when he argues that most of us care not just about psychological continuity but about connections with our future selves (Parfit 1984: 301). But the two may come apart. I

5. This is how I take the relevant clause of the criterion in Parfit (1984: 207). This reading implies that, if a person psychologically continuous with me comes into existence far away, I cease to exist, even if life here continues as usual. A bad result, perhaps, but the obvious alternative is worse. If we require that there be no branching at any point in the future—as in Parfit (1984: 267)—we get the result that, if there will ever be two people psychologically continuous with me, I cannot persist from moment to moment even in the time before the split.

6. As, for instance, the case from Williams (1970) discussed in Parfit (1984: 229–30).

7. Something like it is marked in "The Unimportance of Identity" (Parfit 1995: 28, 44 n. 2), to which we return below.

may regard a life of constant suffering, or radical psychic disconnection, or permanent coma, as being worse than death. It does not give me what I want from survival. But it may sustain the basis of self-concern. It is in part because *I* will endure this fate that I find it so terrible. Conversely, I can ask what makes it good to survive even if I deny that there are reasons of self-concern. Some of Parfit's arguments may equivocate between the two, as for instance when he claims that egoistic concern may be proportioned to degree of psychological connectedness (Parfit 1984: 313). Close connections may be part of what we want from survival; it does not follow that they justify self-concern.

The best way around this is to avoid "what matters" and to be explicit in our normative claims. Thus, Parfit presents the following argument against Self-Concern.[8]

> (a) What it is for me, now, to be identical to X, at *t*, is for us to be related by non-branching psychological continuity, with the right kind of cause.
>
> (b) Being related to someone by non-branching psychological continuity, with the right kind of cause, is not a reason for non-instrumental interest in his well-being.
>
> So: (c) Being identical to X, at *t*, is not a reason for non-instrumental interest in his well-being.

Parfit's attitude to this argument is subtle. He thinks it is reasonable both to accept and to reject the second premise. But he does not question its validity. If the Psychological Criterion is right, that an event will benefit or harm me is a reason for preference only if it is equally a reason for preference that this event will befall someone to whom I am related by non-branching psychological continuity. That is why Parfit infers from the Psychological Criterion that relation R is what matters: it is facts about relation R that justify self-concern, if it is justified at all.

But the inference is puzzling.[9] It is not in general true that, when the fact that *p* is a reason for me to φ, and for it to be the case that *p* is for it to be the case that *q*, the fact that *q* is a reason for me to φ. When I am thirsty, the fact that there is water in this glass is a reason for me to drink it. For there to be water in the glass is for there to be liquid in the glass with a certain chemical composition. But if I am ignorant of chemistry, facts about the chemical composition of the liquid in the glass will not provide me with reasons to act. It would not be rational for me to

8. See Parfit (1984: §102) on the Extreme Claim.

9. For a similar critique, see Johnston (1997). Johnston traces the defect to a "fallacious addition of values" (1997: 167–8), where I trace it to a mistake about the nature of reasons; but we agree that the argument is invalid.

drink what is in the glass on the basis of beliefs about its chemistry as I do when I believe that it is water.[10]

Reasons are facts in that they are true propositions: they are the sort of things that can be known or believed, and by which it is rational to be moved. In my view, this follows from a conception of reasons as premises of sound reasoning.[11] But my objection does not rest on this. What it assumes is more abstract: that reasons are not individuated by what it is for them to obtain; they are more fine-grained. The fact that *p* can be a reason to φ while the fact that *q* is not, so long as it is one thing to believe that *p*, another to believe that *q*. The distinction between these beliefs is not effaced when its being the case that *p* is its being the case that *q*. The argument above is thus invalid. That an event will benefit someone to whom I am non-branchingly R-related may not be a reason to prefer that it happen. It does not follow that Self-Concern is false, even on the Psychological Criterion of personal identity. The fact I took to be a reason, that the event will benefit me, is still a fact.

What if I discover that its truth consists in facts about non-branching R-relations? A future person's being me is their being related to me by non-branching psychological continuity, and I know it. Am I now permitted to take a non-instrumental interest in my future well-being only if being non-branchingly R-related to someone is a reason for non-instrumental interest in theirs? No. What I have discovered is that benefiting someone to whom I am thus related is a constitutive means to benefiting me. If I take a non-instrumental interest in my own well-being, I should take an instrumental interest in what is conducive to my well-being, including benefits to my R-relatives. I need not adjust my interest in my own well-being to the independent weight of R-relations—independent of the fact that they secure my persistence through time.[12] By way of analogy, if I want to avoid pain, and I learn that pain is C-fibres firing, I should want my C-fibres not to fire. I should not lose my aversion to pain on the ground that I have no prior reason to care what my C-fibres do. Instead, I should gain an instrumental interest in C-fibres.

Things might be different if reasons of self-concern were metaphysically prejudiced: if they presupposed a view about the nature of personal identity. Imagine a replacement for Self-Concern on which the reason for preference is that an event will benefit or harm me, and that this fact is irreducible, or cannot be reduced to non-branching psychological continuity. It is no mystery how the Psychological Criterion would undermine such reasons. But it is implausible to locate such commitments in ordinary self-concern. The reasons to which I respond when I am concerned about myself are propositions about me, not about the metaphysics of

10. We should not be misled here by the presence of related reasons. There is the conjunctive fact that the liquid has a certain chemistry *and* that this chemistry makes it water. This fact may be a reason for me to drink what is in the glass, though its first conjunct is not.

11. I defend this view in Setiya (2014a).

12. Again, I agree with Johnston (1997: 167).

persistence. As Jennifer Whiting complains, "It has always mystified me how reasons for concern are lost in the move from a pain's being unanalyzably mine to its being analyzably mine" (1986: 552). What matters in self-concern is whether I will suffer, not whether that fact can be explained in other terms.

In making these objections, I have ignored a metaphysical argument that may appear to work around them. This is the argument from the possibility of fission. On the face of it, the relation described by the Psychological Criterion could hold in branching form between me, now, and two distinct individuals in the future, A and B. This possibility might figure in a direct argument against the truth of Self-Concern. The thought is that, although I cease to exist in fission, there is reason for me to take a non-instrumental interest in the well-being of both A and B. This interest resembles self-concern, except that its object is someone else. If we assume that the reason in this case is the same as the reason that justifies self-concern, it follows that the force of this reason does not turn on its object's being me.

This argument is limited in several ways. For one thing, it is not clear that there is reason to care about A and B in quite the same way, or to the same degree, that I care about myself. For another, the description of the case is in dispute. Perhaps I survive branching in scattered form, or there were two of us all along.[13] And the conclusion of the argument is much less radical than we feared. Unlike the argument above, and the arguments below, the appeal to fission allows for reasons of self-concern that essentially involve the first person. These reasons cite R-relations to *me*. It is just that the object of concern, the person so related, may be someone else.

Even if we set these points aside, I doubt that the argument goes through. It is defective in much the same way as the argument above. Once we see that reasons are as fine-grained as the objects of belief, there is no pressure to conclude from the existence of a reason to care about A and B that this reason is what justifies interest in myself. Presumably, these reasons are related: there is a connection between reasons of self-interest and reasons for interest in A and B. But it does not follow that the metaphysical common factor is the reason in each case. Instead, we can hold on to Self-Concern and regard my interest in A and B as a sensible generalization of concern for my future self.[14] More strongly, we can explain why the holding of relation R is a reason for concern in terms of its resemblance to personal identity. It is because relation R approximates the metaphysical basis of survival that there is reason to care about A and B in something like the way I care about myself. When I learn the truth of the Psychological Criterion and reflect on the possibility of fission, I may begin to care about R-relatives as such. Unlike the concern described above—a concern for what happens to non-branching R-relatives on the ground

13. For extensive treatment of these options, and more, see Johnston (1989).
14. See Johnston (1997: 169–70).

that it will happen to me—this concern is not instrumental. But it is derivative: I would not think to care about R-relatives, except for the fact that relation R is part of what constitutes my identity over time. Metaphysically speaking, when I have reason to care about my future self, and about A or B, relation R is the common factor. But commonality at the metaphysical level does not dictate commonality at the level of reasons. Nor does it dictate the rational order of explanation. We can insist that the basic reason is first-personal. What branching suggests is a rational way to extend one's self-concern through R-relations, not a mistake in Self-Concern.[15]

If this is right, Parfit's arguments misfire because they move illicitly from the nature of the facts that provide us with practical reasons to the content of those reasons. In "The Unimportance of Identity," Parfit returns to his book in light of objections like these. He states the central argument as follows (Parfit 1995: 29):

> (1) Personal identity just consists in certain other facts.
>
> (2) If one fact just consists in certain others, it can only be these other facts which have rational or moral importance. We should ask whether, in themselves, these other facts matter.
>
> So: (3) Personal identity cannot be rationally or morally important. What matters can only be one or more of the other facts in which personally identity consists.

According to Parfit, "if one fact just consists in certain others, the first fact is not an independently or separately obtaining fact. And, in the cases with which we are concerned, it is also, in relation to these other facts, a merely conceptual fact" (Parfit 1995: 31–2). On this interpretation, he contends, the argument goes through.

What should we make of this reasoning? That will depend on how we read its crucial claim. What does it mean for one fact to be, in relation to others, "a merely conceptual fact"? In the essay to which Parfit is replying, Mark Johnston draws a contrast between "analytical" and "ontological" reductionism (Johnston 1997: 151–2). For the analytical reductionist, the Psychological Criterion is a conceptual truth in that it is both necessary and knowable *a priori*. For the ontological reductionist, persisting people are constructed or composed from the entities described by the right-hand side of this criterion, which need not be a conceptual truth. Since this is the context of Parfit's discussion, it is natural to interpret him as defending analytical reductionism, as Johnston does in his most recent response (Johnston 2010: 312–6). But this cannot be right. For one thing, Parfit is clear about the em-

15. This argument of this paragraph draws on Gendler (2002), which makes related points at greater length.

pirical nature of the argument for Reductionism against a metaphysics of irreducible, immaterial souls.[16] For another, it belongs to the logic of the new argument that, unlike the relation of analytic or conceptual equivalence, the relation of one fact to another when the first consists in the second is not symmetric. It would otherwise make no sense to claim that, "if one fact just consists in certain others, it can only be these other facts which have rational or moral importance" (Parfit 1995: 29). If the relation were symmetric, it would follow that, when the fact that *p* consists in the fact that *q*, it can only be the fact that *p* that has rational or moral importance, not the fact that *q*, and that it can only be the fact that *q* that has rational or moral importance, not the fact that *p*. Finally, even when it is *a priori* that its being the case that *p* is its being the case that *q*, the proposition that *p* may be distinct from the proposition that *q*, so that one provides a reason that the other does not. It is not clear how the analytic or conceptual status of the Psychological Criterion makes a difference to the problems raised above.

A more generous reading picks up on the repeated claim that "questions about personal identity should be taken to be questions, not about reality, but only about our language" (Parfit 1995: 33).

> Even without answering these questions, we could know the full truth about what happens. We would know the truth if we knew the facts about both physical and psychological continuity. (Parfit 1984: 23–4)

> We are discussing cases where, relative to the facts at some lower level, the higher-level fact is, in the sense I have sketched, merely conceptual. My claim is that such conceptual facts cannot be rationally or morally important. What matters is reality, not how it is described. (Parfit 1995: 33)

Assuming that what exists belongs to reality and is included in the full truth about what happens, these passages deny the real existence of persisting people. The 'fact' that they exist is merely conceptual: insofar as it is legitimate, talk of personal identity is not committed to their existence. On this interpretation, Parfit does not concede that there are persisting people, and then ask how, in light of their constitution, their existence could matter. Instead, he admits the existence only of what is described on the right-hand side of the Psychological Criterion, which is not literally true. We can think of this criterion as a rule for the use of language, a recipe for how to talk as if there were persisting people—when, in reality, there are not.[17]

16. According to Parfit, while "such a view might have been true . . . we have no evidence for thinking that it is, and some evidence for thinking that it isn't" (Parfit 1995: 16; see also Parfit 1984: 227–8).

17. This kind of proto-fictionalism is to be distinguished from the stage theory, on which personal names and terms like "I" refer to instantaneous person-stages, and claims about what will happen to me, or Setiya, are true in virtue of what happens to our temporal counterparts. This view

The downside of this reading is that Parfit apparently rejects it, dismissing the "eliminative" view on which people do not really exist (Parfit 1995: 16–18). But this is not decisive. The Eliminative Reductionist may be one who believes that ordinary talk of personal identity is committed to the real existence of people, and is therefore illegitimate. It must be revised or reconstructed in light of the true metaphysics. Parfit's alternative is not that people exist as part of reality, but that the legitimacy of ordinary talk does not demand this. The contrast between Eliminative and Non-Eliminative Reductionism thus corresponds to the difference between revolutionary and hermeneutic fictionalism.[18]

Whatever its textual difficulties, the upside of this reading is considerable. It makes sense of the asymmetry in premise (2). When one 'fact' just consists in others, their relation is not symmetric. The former reflects a way of talking in which we are not committed to the literal truth of what we say—as with the existence of persisting people. The latter reflects the factual basis on which we say what we say. This reading also makes sense of Parfit's argument about what matters. If propositions about personal identity are strictly false, they cannot be reasons for anything. If self-concern is justified, it must be justified by the facts on the right-hand side of the Psychological Criterion, or by merely conceptual facts. But if merely conceptual facts are facts about what is true according to a fiction, or 'facts' reported in that fiction, Parfit seems right to insist that they cannot be rationally or morally important.

Reading Parfit as a proto-fictionalist fits well with some otherwise puzzling remarks in *Reasons and Persons*. Discussing teleportation, he writes, "If personal identity does not involve a further fact, we should not believe that there are here two different possibilities: that my Replica will be me, or that he will be someone else who is merely like me" (Parfit 1984: 242). Concerning brain bisection and double transplant, "when we know that each resulting person would have one half of my brain, and would be psychologically continuous with me, we know everything" (Parfit 1984: 258).[19] If the Psychological Criterion is taken literally, as an account of the existence of persisting people that tells us when propositions about their identity are true, these claims will seem confused. Just as we distinguish two possibilities, that pain is C-fibres firing and that it is not, we should distinguish the possible truth of the Psychological Criterion from the possibility that it is false. If there are true propositions about our persistence through time, we do not know everything until we know what they are. By contrast, if such propositions are strictly false—though we speak as though they were true according to the Psycho-

is proposed in Sider (1996). Since it accepts the truth of propositions like those in Self-Concern, it will not give Parfit what he needs. For objections to the stage theory, see Johnston (2010: 68–74).

18. The terminology derives from Burgess (1983), via Stanley (2001: 36).

19. See also Parfit (1984: 248, 265, 282–5).

logical Criterion—we are not ignorant of anything, except the content of a fiction or the use of words, when we know the facts on its right-hand side.

Have we found, at last, a cogent argument against Self-Concern? This argument relies on (1) to (3) above. Personal identity consists in facts about non-branching psychological continuity. Relative to these facts, the 'fact' of our persistence through time is merely conceptual. The relevant propositions – for instance, that an event will benefit or harm me in the future – are strictly false, though we legitimately speak as if they were true. Since the propositions are false, they cannot be reasons for anything.

Unlike the argument from (a) to (c), the inference here seems valid. It is less clear that Parfit gives strong arguments for its premise, as opposed to premise (a). Because he is not careful to distinguish the Psychological Criterion as a species of realism from the proto-fictionalist view I have just described, it is hard to be sure which of these positions, if either, his reasoning supports. But there is a more important point to make. Even if Parfit is right that the propositions cited by Self-Concern are false, that would not matter to its truth. According to Self-Concern, the fact that an event will benefit or harm me is a reason for me to want, or not to want, that event to happen. This reason derives from the effects of the event on my well-being, not its effects on anything else. And its force as a reason turns on its first person character. This means that, if there are such facts, they provide me with reasons of a certain kind, in a certain way. This could be true even if, as it happens, there are no such facts. Our topic is not whether I will in fact persist through time but whether and why propositions about my future have the rational significance they are given by Self-Concern.

These questions are addressed by the invalid argument from (a) to (c); they are not addressed by the valid argument from (1) to (3). In fact, however, both arguments miss the central feature of self-concern, that it is essentially first-personal. They consider what will happen to me as a possible reason for preference. But what they ask is whether, and why, a future individual is identical to someone who exists right now. These questions could be raised in the third person. What is it for a future event to harm or benefit Setiya, the one who is typing these words? Is "Setiya" the name of a persisting person, or do we just talk that way? That is why, if they work, Parfit's arguments have moral implications, implications for autonomy, compensation, and distributive justice.[20] They are about the place of persisting people in our ethical scheme, not about the special status, for me, of the person I am.[21] This is not an objection to the arguments in their own terms. But it is a reason to think that they cannot answer the question I have posed. In order to

20. These implications are explored in Parfit (1984: Ch. 15), and to different effect in Jeske (1993).

21. For a similar point, see Wolf (1986: 706).

do that, we must turn to the distinctive nature of first person thought. What does it mean to think not just about the future of someone who happens to be me, but about myself, as such? And how does the fact that I think of myself in this way justify self-concern?

2. The First Person

We all have different ways of thinking about things, that is, of referring to particulars in thought. I can think that *I* am slow, or that *Setiya* is slow, or, as I watch a video of someone moving slowly—a video that is in fact of me—that *he* is slow, or I can think that this is true of the moral philosopher who was born in Hull. These are all distinct thoughts in that they involve distinct propositions or objects of thought. This comes out in the fact that it is possible to believe any one of them without believing any other.

At least, that is how things look on a natural, broadly Fregean, picture of thoughts.[22] On this conception, we can ask what it is to think of an individual as me, to refer to myself in the first person. In doing so, we aim to characterize part of what I think, an aspect of the proposition I believe when I believe that *I* am slow, as opposed to believing this about myself, referred to in some other way. When I think about someone, I do so in virtue of standing in a certain relation to him. What relation do I exploit in first person thought?

Some philosophers resist this frame. They insist that the proposition I believe when I think that I am slow is the same proposition I believe when I think that Setiya is slow or, watching the person in the video, that he is slow. This proposition is Russellian: its constituents are the individual in question and the property ascribed to him.[23] How do Russellians explain the data that motivate the Fregean view? Can't I believe, of the person in the video, that he is slow, without believing that I am slow? The reply is that I believe the proposition under a demonstrative guise, not under the guise of the first person. Believing is a relation between a thinker, a proposition, and a guise under which that proposition is entertained. Different guises play the role in psychological explanation that Fregeans assign to differences in the proposition believed. The question above can be adapted here, as a question not about the content of thoughts but about the guise under which they are entertained: what is it to think them in the first person?

Finally, David Lewis held that the objects of thought are not true or false propositions but properties self-ascribed by a subject (1979). All thought is in this

22. See Frege (1892/1997a; 1918/1997b); my understanding of Frege is indebted to Evans (1982).

23. For versions of this approach, see Marcus (1961), Kripke (1980), Salmon (1986), Kaplan (1989). The discussion in the text most closely follows Salmon.

way first-personal. The motivations for this claim are not important here, nor are the technical details. What matters is that our question is not erased. Again, it can be adapted. When I self-ascribe a property, what I believe is true just in case that property is instantiated by a certain individual. In virtue of what relation to me is this individual the focus of my self-ascriptions? What relation do I exploit in centering my thoughts on him?

In what follows, I assume that the Fregean approach is basically right. I will investigate the relation to myself that is involved in first person thought, a relation that is partly constitutive of the propositions I believe. As I have emphasized, however, versions of this project can be pursued in other frameworks, and similar questions arise.

An obvious starting point is the fact that first person thoughts are about whoever is thinking them. When I believe that I am slow, my subject is the thinker of that very thought, the one who instantiates that psychological property. Philosophy being what it is, there is controversy even here. On a metaphysics of temporal parts, persisting people are composed of short-lived *person-stages*.[24] A persisting person has a property at time *t* just in case that property is instantiated by the stage that exists at *t*. Where the property is psychological—believing that I am slow—there are thus two candidates to be the thinker of the thought. In fact, there may be more. If arbitrary sums of person-stages compose a persisting thing—perhaps not a person, but a thinker all the same—there will be many thinkers for every thought. Nor do we need temporal parts in order to raise a puzzle. It is enough to ask how we relate to human animals, animals who exist *in utero* and may survive the end of our mental lives. Some deny that we are identical to such animals. But then, if animals can think, there are again two candidates to be the subject of first person thought. Do both refer to themselves, one mistakenly believing that it is a person when in fact it is not?[25] Or do they refer to the person, not the animal that coincides with him?[26] For simplicity, I will ignore these complications; I am in any case dubious of the metaphysics behind them. When I instantiate a psychological property, I am not spatially coincident with something else—an animal or person-stage—that also instantiates it. Those who think otherwise will need to adapt my discussion to their views.[27]

---

24. See Lewis (1976), and for a similar metaphysics with a different take on the referent of the first person, Sider (1996).

25. Olson exploits this puzzle in arguing for "animalism" (1997). Shoemaker responds by denying that animals can think (1999).

26. As suggested by Noonan (1998).

27. Doing so will not affect the arguments to come. Further complexity, mediation, or arbitrariness in the relation by which we think of ourselves in the first person would only make Self-Concern more difficult to defend. An exception might be made for views on which the relation in question is ethical. Suppose the threat of indeterminacy is resolved by how I organize my concerns—do I care about people, animals, or specific sums of person-stages?—where the answer settles the reference of

Even if we grant the simplification, the characterization of first person thought as referring to the thinker is inadequate. It would, for instance, be a mistake to equate first person thoughts with ones in which I am identified descriptively as the subject. To believe that I am F is not to believe that the thinker of this thought is F. If it were, there would be no logical incoherence in believing that I am F, believing that I am G, and believing that no-one is both F and G. That these thoughts are about the same person would not follow from the propositions they involve, but from the fact that two distinct episodes of demonstrative reference—to one thought, then another—pick out thoughts of the same thinker. There may be reasons to believe this, and in the circumstance, no prospect of being wrong, but that is not guaranteed by the contents of my thoughts, if they identify me descriptively as their thinker. By contrast, it is a matter of logic that the proposition that I am F and the proposition that I am G together entail that I am both F and G.

Are these remarks enough to characterize first person thought? When I think of myself in the first person, I refer to the one who is thinking this thought, though not under that description. Instead, my thoughts involve a non-descriptive concept, *I*, that is the same when I think that I am F as when I think that I am G. It is internal to first person thought that the propositions it takes as objects are about the same person.

Although I accept all this, I think there is more to say. This comes out when we ask how I relate to my own beliefs. Suppose you have a concept that fits the criteria above: its reference exploits the relation of being the thinker of a thought; it is non-descriptive; and it enters identically into many thoughts about yourself. So far, we have said nothing about the grounds on which you apply this concept or form beliefs involving it. We can thus imagine, or try to imagine, that you do so only by inference or on the basis of testimony. In particular, when you believe that *p*, you regard it as an open question whether you have that belief, a question to be settled by further evidence: "Yes, it is true that *p*; but do I believe that *p*?" Although your concept, *I*, refers to the thinker of your thoughts, you are completely oblivious to this fact. As Gareth Evans observed, however, from the first person perspective our beliefs are transparent to the world: "I put myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*" (Evans 1982: 225); if the answer is yes, I conclude that I have that belief. In this way, I gain immediate knowledge of my own mental state. It is this capacity you lack in our imagined case. If you lack this capacity, however, you lack the capacity for first person thought. When you relate to the referent of *I* in this alienated fashion, you do not think of him as yourself.

I said that we should *try* to imagine this peculiar case because I am not sure

first person thought. But this is quite implausible. I can engage in first person thought while taking uniqueness utterly for granted, so that there is no fact of the matter as to how I would organize my concerns if I were to face a plurality of candidates for being me.

that it is possible. But this does not affect the point. If you could have a concept that meets our criteria, yet be unable to tell that you believe that *p* when you believe that *p*, it is not a concept of the first person. If this is not in fact possible, that is because possession of such a concept requires the capacity for self-knowledge: a concept could not refer to you as the thinker of these thoughts unless you are able to exploit that fact in reporting your beliefs. Either way, the capacity for self-knowledge is a condition of first person thought. The relation by which I refer to myself in the first person is not simply that of being the thinker of these thoughts, but being the object of immediate knowledge. The first person concept refers to the one whose thoughts can be known in this way.[28]

Some philosophers will find this view extravagant. If that is your reaction, do not fear: the argument of section 3 goes through if the demands on first person thought are weaker than I take them to be. Others hold that it is not extravagant enough. On the account sketched so far, first person thought involves a non-descriptive concept, *I*, that refers to whoever is thinking a given thought, and requires a capacity to know that I believe that *p*, when I do, without relying on testimony or inference. The capacity for such knowledge is, in terminology due to Sydney Shoemaker, "immune to error through misidentification": there is no possibility of being justified by the exercise of this capacity in the belief that someone is F but mistaken in believing that it is me.[29] In principle, there could be other sources of knowledge that are equally immediate: non-inferential, non-testimonial, and immune to error through misidentification relative to the concept of myself that figures in the self-ascription of beliefs. Capacities of this kind interact with the capacity for psychological self-knowledge in that they issue in thoughts that refer to me, as their thinker, in the same way. According to Evans, we must have such capacities in order to think of ourselves in the first person. In particular, we must have a capacity for immediate knowledge of our bodily location in space. Without this, as in "the perennial nightmare [that] a human brain might exist, from birth, in a vat, subjected by clever scientists to a complex series of hallucinations," there is no prospect of first person thought (Evans 1982: 250).

I am not convinced that Evans is right about this, though it is worth distinguishing two claims. The first is that we have capacities that yield immediate knowledge of our spatial properties. One is proprioception, which yields first-hand knowledge of one's bodily orientation that is both non-inferential and immune to

28. Does it follow that non-human animals, at least those who fail to self-ascribe beliefs, cannot think of themselves in the first person? That depends on subtle questions about the attribution of capacities. Can one have the capacity for immediate knowledge of one's beliefs, but fail to manifest this capacity because one lacks the concept of belief? If so, it is possible, in principle, for animals without the concept of belief to engage in first person thought. Even animals that lack this capacity can engage in purposive action, perhaps on the model of Cappelen and Dever (2013: Ch. 3).

29. Shoemaker (1968), drawing on Wittgenstein (1958: 66–7).

error through misidentification relative to the concept of oneself that figures in the self-ascription of beliefs. Another is perceptual perspective, as when I know that I am facing a desk by looking towards it. These capacities are Evans' principal focus, and a subject of later controversy.[30] What if my nervous system is hooked up to someone else's body? Could I be justified proprioceptively in the belief that someone is standing but mistaken in thinking that it is me? Or justified in thinking that someone is in front of a table, but not know who it is? A final capacity is practical: knowledge of my intentional actions that is not perceptual or inferential.[31] This too is contentious: not everyone accepts that we have such knowledge.

I am sympathetic to Evans on the existence of these capacities, though I will not defend this sympathy here. Knowledge of agency has special significance for Self-Concern: we will come back it below.[32] But Evans makes a second claim. He holds that, without the capacity for immediate self-location, it becomes "problematic how [a] subject could ever make sense of the thought that he is located somewhere" (Evans 1982: 224). The argument for this claim has two steps. First, that if I lack immediate knowledge of my spatial properties, it is impossible to know them at all. The thought is that, if knowledge of my spatial properties is not immediate, it will depend on inference, or something like inference, from the properties of a given human being. This transition assumes an identity: that I am the relevant human being. But this identity cannot be verified without appeal to spatial coincidence, which initiates a vicious regress. Second, that if I cannot know my spatial properties, I cannot make sense of the thought that I am located in space.

Both steps are disputable, and they are stronger than we need. For Evans, first person thought requires a capacity for immediate knowledge of one's spatial as well as one's psychological properties. The referent of first person thought is the object of this knowledge. Where knowledge of either kind is lacking, such thought becomes impossible. I accept, and will defend, a weaker claim. It may be sufficient for first person thought that one have the capacity for psychological self-knowledge. But if one has immediate knowledge of other kinds, as in proprioception, perceptual perspective, or intentional action, the corresponding capacities play a similar role in fixing the reference of first person thought. Thus, if I have a capacity to form beliefs about my spatial orientation not by testimony or inference but by a direct information-link, and this capacity is immune to error through misidentification relative to the concept of myself that figures in the self-ascription of beliefs, the object of first person thought must be the source of information. (Where my nervous system is hooked up to another body and the source of information is not the thinker of these thoughts, I can no longer think of myself in the first person.)

30. Evans (1982: 220–4); for objections, see O'Brien (2007: 38–42, 202–11).

31. See Anscombe (1963), Evans (1982: 224 n. 34).

32. I have explored such knowledge in a series of essays, from Setiya (2008) to Setiya (2012).

Likewise, if I have the capacity to act for reasons, and thus for practical knowledge of action, the object of first person thought must be the agent. Psychological self-knowledge is distinctive in being essential to first person thought. But it does not play an exclusive role in fixing its reference. The referent of first person thought is the object of immediate knowledge in all its forms.

The argument for this conclusion turns on the conditions of knowledge. Go back to the capacity above: a capacity to form beliefs about my spatial orientation through a direct information-link that is immune to error through misidentification relative to the concept of myself that figures in the self-ascription of beliefs. And suppose that this capacity does not play a reference-fixing role. It could still be reliable, delivering the belief that I am sitting or standing just when I am. But its reliability would be accidental in a way that prevents it from being a source of knowledge. On our supposition, it just happens that the origin of the information to which this mechanism responds is the referent of the concept it exploits in reporting that information. It happens to derive information from me, the object of psychological self-knowledge, and to report that information using a concept, *I*, whose reference is fixed by the capacity for such knowledge. But there is no connection between these facts. It could have derived information from someone else and used a concept that refers to me, or derived information from me and used a concept that refers to someone else! If the capacity in question is to be a source of knowledge, it cannot be in this way accidental that the body whose states it reliably tracks is mine. Instead, that body counts as mine in part because it is the body of which I have such knowledge. It counts as mine, too, because it is the body of the thinker who is the object of psychological self-knowledge. It is not an accident that the same concept, with the same referent, is employed in bodily and psychological self-ascription: it is a concept whose reference is fixed by both capacities. And so it is not an accident that proprioception is reliable about *me*. The upshot is that, in order to be a capacity for knowledge, a capacity to form immediate beliefs about myself must help to fix the reference of the first person.

I have argued that first person thought involves a non-descriptive concept, *I*, that stands for the object of immediate knowledge. This includes knowledge of what I believe, but also knowledge through proprioception, perceptual perspective, and intentional action. The capacity for such knowledge is the basis of first person thought. Hence:

> IMMEDIATE KNOWLEDGE: When I think of myself in the first person, I do so in virtue of standing to myself as the object of immediate knowledge, knowledge that is non-inferential, non-testimonial, and immune to error through misidentification relative to the concept of myself that figures in the self-ascription of beliefs.

This explains the distinctive content of first person thought: the difference between thinking that *I* will suffer and thinking that Setiya will.

I end with a proviso. In developing this view, I assumed, with Evans, that we have immediate knowledge of our spatial properties. In fact, it is knowledge of agency that matters most, since we are interested in the practical significance of the first person. Like Anscombe, I think we have a capacity for knowledge of what we are doing that is not perceptual or inferential; in acting intentionally, we exercise this capacity.[33] I know that I am writing an essay on self-concern, though not by inference or observation of myself. Those who doubt the possibility of practical knowledge, understood in this way, should revise the formula above, giving agency a separate place in first person thought. If I am capable of acting intentionally, first person thought refers to the one who executes my intentions. When I think of myself in the first person, I do so in virtue of standing to myself as the object of both agency and immediate knowledge. This view is oddly disunified. But it is compatible with the argument below. Still, for simplicity, and out of conviction, I will adopt the Anscombean line, on which agency falls under Immediate Knowledge itself.

We are at last in a position to answer our guiding question, about the truth of Self-Concern. If Immediate Knowledge tells us what it is to think of someone as myself, does it make sense to take a non-instrumental interest in my own well-being just because it is mine? Do facts about what will happen to me provide me with reasons for preference whose force turns on their first person character? We will approach these issues indirectly, by considering a picture that is absent from the argument above. This will close a gap in the argument, and it will frame the question of Self-Concern in a more perspicuous way.

3. Self-Interest as Self-Love

In *Surviving Death*, Mark Johnston contrasts "two uses of first-person pronouns: a straightforward indexical use that refers simply to the human being that one is"—the speaker of an utterance or thinker of a thought—"and a truly subjective use where an interesting subjective property, the property of being at the center of a given arena, is in play" (2010: 192). He introduces the subjective use of "I" phenomenologically. Begin with the idea of the visual field as an apparent object: a perceptible array of colours and shapes. Now add the idea of a bodily field, a three-dimensional volume of bodily sensation, then the idea of a tactile field, an auditory field, an olfactory field, and so on. Add, too, "all the items that are in principle open to introspection" including "the deliverances of proprioception and

33. It would be more accurate to say that we have many such capacities, corresponding to things that we know how to do. I pursue this issue in Setiya (2012).

the immediate knowledge of which intentional acts you are currently performing or trying to perform":

> The whole centered pattern, existing at a particular time, and perhaps over time, I call *an arena of presence and action*. There is one such arena here, and I assume you can truly make a corresponding remark about your own case. . . . Think of the arena as a sort of virtual frame or "container" that includes all this; it is if you like the mind considered as a sort of place, the mental "bed" in which the stream of consciousness flows. (Johnston 2010: 139–40)

The phenomenological claim is that, when we introspect, it is as if we are presented with a unified object, an *arena* that includes the contents of consciousness. It seems to us that we can use a demonstrative expression, "this arena" to pick out this object, and a corresponding description, "the one at the center of this arena," which guides the subjective use of "I" (Johnston 2010: 156). For Johnston, the subjective "I" does not refer to a human being. Instead, it purports to refer to a *self* that is essentially at the center of this arena: its persistence conditions are given by the persistence conditions of the arena of presence and action.[34] The problem is that, as Johnston goes on to argue, we cannot really make sense of the future persistence of this arena, and so we cannot make sense of the future persistence of the self.[35]

Johnston's argument is of interest to us in part because he takes it to have practical implications. According to Johnston, it is the subjective use of "I" that figures in Self-Concern (2010: 161–4, 204–6). Since it cannot be true that an event will harm or benefit my future self, there are no forward-looking egoistic reasons. But the idea of the subjective "I" is in several ways peculiar.

Suppose the reference of "I" is fixed by the description, "the self at the center of this arena." Since this description includes a demonstrative, we can ask what happens when I lose track of its object, the arena of presence and action, as when I go to sleep. Waking up the next day, I say to myself, "I was born in Hull," taking for granted that the arena I am now attending to is the one I was presented with yesterday and the self at its center has survived. What is the content of my thought? If it is fixed by an earlier act of demonstration, it should strike me as a substantive claim that I am at the center of *this* arena, the one I am presented with now: that does not follow from the content of what I am thinking. That seems wrong.

34. For arguments to this effect, see Johnston (2010: 192–9).

35. The arena cannot persist as an enduring mental substance, since there is no such thing (Johnston 2010: 168–76); nor can its persistence be understood through the application of explicit criteria, since that would place it beyond the reach of younger children (Johnston 2010: 209–11); instead, the arena is a merely intentional object, an illusion, and there is no basis for prospective identification of the same illusory object over time (Johnston 2010: 222–33).

On the other hand, if the proposition I express when I say "I was born in Hull" is anchored by an attempt to demonstrate *this* arena of presence and action, so that it follows from the content of my thought that I am at its center, the proposition I express is different from the one I expressed when I said to myself yesterday, "I was born in Hull." That these thoughts are about the same thing does not follow from their content. In order to sustain a single thought with the subjective "I" one must continuously introspect the same arena, just as one sustains a thought with a perceptual demonstrative by training one's attention on a given object. Since such attention is disrupted by sleep, the propositions expressed with the subjective "I" will vary from day to day.

Either way, we are faced with something odd. On one interpretation, it is a substantive claim that I am at the center of the arena with which I am presented right now. On the other, there is temporal instability in what is expressed by the subjective "I." Equally strange, I think, is that, on both interpretations, it is not a condition of the subjective use of "I" that one be able to self-ascribe beliefs. "I" picks out the self at the center of this arena of presence and action, which includes everything that is open to introspection. But there is no requirement that one's beliefs, in particular, be objects of introspective knowledge. For all we have been told, when I believe that *p*, I may regard it as an open question whether I have that belief, a question to be settled by further evidence: "Yes, it is true that *p*; but do I—the self at the center of this arena—believe that *p*?" This cannot hold for anything well-conceived as first person thought.

We could ask how Johnston would reply to this objection. But given the peculiarity of the subjective "I" it is more fruitful to ask why he is led to invoke it at all. What motivates the distinction between two uses of the first person, one subjective, the other a mere indexical? Part of the answer may be implicit in Section 2. That first person thought is about the thinker is not sufficient to explain its content. If it *is* sufficient to define the mere indexical use of "I" then Johnston is right to look for an alternative. He is wrong to conclude that "I" can be used in two ways, and that, in its most interesting use, it does not refer to a human being. A better response to the defects of the mere indexical view is to appeal to Immediate Knowledge. Once we do this, the idea that we are presented with an arena of presence and action looks like a misreading of the phenomenology. Johnston is right to suggest that introspection or immediate knowledge fixes the reference of first person thought. The mistake is to confuse immediate knowledge of facts about the world—that I am sitting, facing a desk, believing that I am slow, typing at the computer, and so on—with seeming awareness of an object, an arena, that contains their mental correlates. Speaking for myself, when I introspect, all I find are facts about me in relation to the world, not a virtual frame or container for mental stuff, a bed in which the stream of consciousness flows.

A similar point applies to Johnston's reading of the Fregean claim that "every-

one is presented to himself in a special and primitive way, in which he is presented to no one else," so that first person thoughts cannot be shared (Frege 1918/1997b: 333): "How could we possibly satisfy Frege's constraint if we are restricted to reference to the mutually available stock of real items? Frege's constraint can be satisfied only if one can pick oneself out by way of the merely intentional objects that are available only to oneself" (Johnston 2010: 225). But the privacy of first person thought does not turn on reference to a private intentional object, but on a way of knowing about myself that is not available to others: a capacity for immediate self-knowledge. It is this relation that figures in the constitution of my thoughts. Since others cannot relate to me in the same way, they cannot think about me in the first person. Again, Johnston misconceives introspection as access to special objects, not as special access to an ordinary thing.

Johnston gives two arguments for his position, the first of which exploits our theme: the practical significance of first person thought.[36] According to Johnston, the view that I am identical with Setiya, not merely constituted by him, "makes *superficial* nonsense of our special concern for ourselves, the concern that manifests itself in one's everyday egocentrism and in one's unique fear of death" (Johnston 2010: 146). On this view, "I have available a mode of presentation of the identity fact that [Setiya is Setiya] which others do not have available to them" (Johnston 2010: 147). It is when I apprehend this fact in the first person, when I believe that I am Setiya, that I take a special interest in his well-being. But the fact I apprehend is trivial. How could the fact that Setiya is Setiya justify my response? "The whole structure of my intelligibly egocentric self-concern now looks like it depends on my

36. The second argument turns on "the paradox of auto-alienation" (Johnston 2010: 197). When I introspect, it can seem to me that I could have been someone else: Mother Teresa, Locke's Rational Parrot, or the Prince of Darkness (Johnston 2010: 144). As Johnston points out, the mere indexical view can explain this appearance: "Because "I" is governed by [the semantic rule that it denotes the speaker of an utterance], one can use it to pick out a human being, even while one is ignorant of who that human being is" (Johnston 2010: 145). The identity statement, that I am Setiya, is knowable only *a posteriori*. This creates the illusion that its truth is contingent, and so makes room for "feats of auto-alienation, which falsely separate our supposed selves from the human beings with which we are *necessarily* identical" (Johnston 2010: 145). The same point holds for the view that I accept. I could have immediate knowledge of a human being, and so think of him as myself, even when I am ignorant of which human being he is. It can thus seem contingent that I am Setiya, not someone else, even though it is not. Johnston's objection is that, if this were the whole truth, it would be semantically incoherent—"incoherent given a full understanding of the semantics of the expressions"—to suppose that I might not have been Setiya, knowing that I am (Johnston 2010: 195). According to Johnston, however, this is not the case: even when I know that I am Setiya, there is no semantic incoherence in supposing that I might have been someone else. My response is to dispute this claim. When I assume that Setiya is the object of immediate knowledge, it seems impossible that I be anyone else. What remains is the fact that I might be wrong, that I might have immediate knowledge of Mother Teresa, Locke's Rational Parrot, or the Prince of Darkness, though I do not realize it. But that is consistent with my view. That I might be someone else is an epistemic, not metaphysical, possibility.

confusing a difference in the mode of presentation of a fact for a difference in the fact presented"—in making up a further fact that justifies special concern (Johnston 2010: 148). Johnston's complaint is that we are not subject to such confusion. Self-Concern is not so obviously false.

The problem with this argument is that, as we saw in Section 1, reasons for preference are facts in that they are true propositions. When Johnston talks about two modes of presentation of the same identity fact, he is using "fact" in a different way, for the "ingredients of reality that make propositions true or false" (Johnston 2010: 147). But we do not need a further fact, in this sense, to justify Self-Concern. All we need is a difference in thought. It is one thing to believe that I am Setiya, another to believe that he is identical with himself. The object of the first belief may be a reason where the object of the second is not.

At the same time, Johnston has a point. Those who advocate Self-Concern need not confuse a difference in mode of presentation, and so in the proposition believed, with a difference in truthmaker. They can be clear that what justifies self-concern is how one apprehends an ingredient of reality, not a difference in the ingredient apprehended. But they had better explain why. Why does the mode of presentation involved in first person thought have the rational impact it is given by Self-Concern? This question reflects a general constraint on reasons:

> HARMONY: Where a subject thinks about an object in virtue of standing in a certain relation to it, that relation must accord with the rational significance of the corresponding thoughts.

The idea behind Harmony is that, in describing the relation by which a subject thinks about an object, we aim to characterize not just conditions of reference, but the content of the corresponding thoughts and thus their role in the subject's psychology. We can illustrate this point by applying it to other cases, though the details are controversial. Thus, according to Evans (1982: §6.4), when we use perceptual demonstratives—"that glass", "that building"—we think about an object in virtue of a perceptual relation or information-link that enables us to locate and re-identify that object in egocentric space. This fact accords with the rational significance of the corresponding thoughts, which play a role in guiding bodily movement through that space. It makes sense to reach in a certain direction for that glass when you need a drink because your demonstrative thoughts about the glass contain information about its relation to your body, in non-descriptive form. Likewise, for Evans (1982: §6.3), to think about a location as "here" is to think about it in virtue of more general dispositions to acquire information about it by perceptual means and to engage with it through bodily movement. This accords with the role of such thoughts in intentional action. When I think that something is here, it makes sense to attempt, or to avoid, perceptual or bodily interaction with

that object. Evans may be wrong about the specifics of "that" or "here" but he is right to focus on relations that play these roles.

In light of Harmony, we can reiterate Johnston's complaint. On the mere indexical view, the relation by which I think about myself in the first person is that of being the speaker of this utterance or the thinker of this thought. But it is a mystery why relating to someone in this way would justify concern for his well-being and therefore validate Self-Concern. Why care so much about the one who spoke these words or who instantiates this mental property? At best, these features correlate with ones that matter: they do not constitute a reason for concern. As Johnston writes, "I am able to represent the identity fact that Johnston is Johnston by way of the sentence 'I am Johnston.' For in my mouth 'I' denotes Johnston. But so what?" (Johnston 2010: 147)

If this is right, the force of Johnston's challenge does not turn on thinking of reasons as facts in the sense of truthmakers, not true propositions, or on conflating the two. It rests instead on the demand for Harmony. The question is how this plays out for Immediate Knowledge, the theory of first person thought proposed in Section 2:

> IMMEDIATE KNOWLEDGE: When I think of myself in the first person, I do so in virtue of standing to myself as the object of immediate knowledge, knowledge that is non-inferential, non-testimonial, and immune to error through misidentification relative to the concept of myself that figures in the self-ascription of beliefs.

The answer, I believe, is that Immediate Knowledge does not accord with Self-Concern. Having the capacity for immediate knowledge of someone does not justify non-instrumental interest in his well-being. Why care so much about the one you know first-hand, without the need for inference, whose beliefs you can access in a special way? The epistemic relation that secures first person thought is not a basis for special concern any more than the relation of speaking this utterance or thinking this thought. Since there is reason to care about everyone, the fact that an event will benefit or harm me is a reason for me to want, or not to want, that event to happen. This reason derives from its effects on my well-being, not from its effects on anything else. But its force does not turn on its first person character. Self-Concern is false.

There is a natural rejoinder. To emphasize the epistemic dimension of first person thought, as in Immediate Knowledge, is to risk obscuring the role of agency in fixing its object. It is to emphasize theoretical reason over practical. No wonder we lose the basis of self-concern! On my view, the distortion here is merely apparent. Agency falls under Immediate Knowledge: the capacity for intentional action is a capacity for practical knowledge of what one is doing and why. Those who doubt

this conception may give agency a separate role, along with immediate knowledge, in first person thought. Either way, to think of myself in the first person is to think of myself as agent—though not under that description. Does this make a difference to the argument above? I do not think it does. When we ask, without prejudice, what sort of attitude makes sense in light of this relation, we find no support for Self-Concern. The relation of agency has practical significance, but the value it confers on the agent is instrumental. I should matter to myself as the basic means to efficacy in the world, the source of intentional action. It follows that I have good reason to care about my own well-being. But against Self-Concern, this reason derives from the role of my well-being as a means, not as an end. My well-being affects my capacity to act, to do what I have reason to do. It is important to me in a way that yours is not because my agency depends on it.

This conclusion prompts resistance. You will likely agree that suffering, especially when it is more intense, interferes with agency, and that this is a reason to protest against it. You may be less convinced that this exhausts its significance for me. On the face of it, to be in pain is to be in a state that provides a reason for preference, a reason to want pain to cease that is independent of its effects. The converse holds for pleasure. Maybe this is where Immediate Knowledge comes in: it explains the consciousness of pleasure and pain that make them so compelling. But I don't think this is right. There are two points to make here. The first is that, despite appearances, I have not denied the rational import of pleasure and pain. My topic has been the significance of *thoughts* about oneself. In asking whether the fact that I will suffer is a reason for preference, we are asking whether the belief that I will suffer plays a distinctive role in practical reasoning.[37] According to Self-Concern, it is rational to respond aversively to the belief that I will suffer not just because it represents someone as suffering, but because it represents him as me. It is consistent with this being false that one should respond aversively to the experience of *being* in pain, and that this involves a relation to oneself one cannot have to anyone else. When I am in pain, I have non-instrumental reason to change my condition. What is involved here is a rational response to pain itself, not to beliefs about that state. Immediate Knowledge is irrelevant, as is Self-Concern.

The second point is that I do not reject self-interest altogether. I have argued that we should deny Self-Concern, according to which non-instrumental interest in my own well-being is justified by the fact that it is mine. I have not argued that self-interest is irrational. (By "self-interest" I mean a concern for my own well-being as an end, not just a means, that goes beyond the concern I am required to have for everyone.) In "Love and the Value of a Life," I urge a conception of love on which it is rational to love any other human being.[38] You need not have particular merits

37. See, again, Setiya (2014a).
38. Setiya (2014b: §§1–2).

for me to be justified in loving you. And while relationships give reasons for love, it can be justified without them, as in love at first sight. On the view that I defend, the fact of another's humanity is sufficient to justify love, though not to require it. Love comes in many forms, which differ in various ways. But a common element of most is disproportionate concern for the interests of the beloved. It is rational for me to fall in love with you, whatever your merits, without a past relationship, and so to give your interests extra weight. What goes for you, as another human being, also goes for me. It is not irrational for me to love myself, whatever I am like, and so to take a disproportionate interest in my own well-being. The justification for doing so is not that I am me, but the fact of our shared humanity. Self-love is the primordial case of love at first sight. Or better, since I am available to myself not just perceptually but through immediate knowledge, in both agency and introspection, it is love at first act, or first thought. I am presented to myself in a special and primitive way in which I am presented to no-one else: as the agent of my actions and the thinker of my thoughts. What could be more natural than to love the person who is given to me this way?

The theory of self-interest as self-love is less surprising than it seems. As many agree, it is not a requirement of practical reason to be more concerned with one's own well-being than that of other people. It is not irrational to be selflessly or impartially altruistic. At most, it is rationally permissible to give priority to oneself, an attitude for which we have sufficient but not decisive reason. I depart from Self-Concern on the more elusive question of *why* it is justified: what is the reason for self-interest? According to Self-Concern, first person thought plays an essential role in the justification of disproportionate interest in oneself. On my view, it does not: the justification of self-interest is impersonal, though what is justified is a personal investment in one's own well-being, not as a means but as an end. How odd this verdict is I leave to the reader's judgement. In my view, it is the right conclusion to draw from the interaction of Harmony with Immediate Knowledge. And it resonates with a moral idea I find compelling and have begun to explore elsewhere: the commandment to love one's neighbour as oneself.[39] Instead of being read as a severe, almost inhuman demand for complete impartiality, this formula may point to the fact that what justifies self-love is equally a ground for love of anyone else. Love of neighbour involves a prior self-alienation: "To love a stranger as oneself implies the reverse: to love oneself as a stranger" (Weil 1947/2002: 62).

These thoughts need further exploration, and I will not pursue them here. Nor have I done more than sketch an alternative to Self-Concern. What we can say now is that there is a conception of love that allows for self-interest without appeal to the special significance of the first person. On this conception, just being human is sufficient reason for love, for oneself as for anyone else. The argument of this

39. In Setiya (2014b) and Setiya (2014c).

paper supports this conception indirectly: it explains the rationality of self-interest, despite the failure of Self-Concern.

Having said this, we should not overstate the importance of self-interest in our lives. Only some of my reasons explicitly concern my own well-being. I write an essay in order to do philosophy, play a game with my son because he enjoys it, prepare a handout in order to teach my class. Doing these things may benefit me incidentally, but that is not my reason for doing them, or for wanting to.

A full account of the practical reasoning behind these actions would likely appeal to beliefs about me in which I am represented as myself. Such beliefs may register my instrumental role in producing outcomes that do not involve me. But that does not exhaust the practical role of first person thought—nor have I argued otherwise. I have focused on the intersection of well-being and the first person, rejecting Self-Concern. It is an open question how far non-welfarist reasons are, or might be, essentially first personal. One thinks here not only of personal projects but of relationships with others: reasons to benefit my friends and family, or to save them from harm, that turn on how they relate to me. And one thinks of agent-centred restrictions reasons not to act in certain ways even if the result is more actions of the very same kind.[40] It may be wrong to kill one innocent person, even to prevent more people from being killed. Does the idea that I have such reasons, and that their force is explained by the fact that they involve the first person, conflict with Immediate Knowledge? Not in any obvious way. If first person thought relates me to myself as agent, through immediate knowledge of action, it makes sense for me to take special responsibility for what I do. My agency is involved in personal projects and meaningful relationships: it partly constitutes them.[41] And it may explain the special objection to doing or intending harm.

I do not say that any of this is clear: it is a matter of dispute whether we are subject to agent-centred restrictions, how we are obligated to friends and family, and why, what justifies our individual pursuits. What I claim is that our account of these phenomena should conform to Harmony. If the fact that I can achieve some end is a reason for me to act, where the fact that you can achieve it is not, the contrast must accord with Immediate Knowledge. Likewise, if there is more reason to care about my child than yours, or if the fact that I will cause harm is a reason to avoid a certain option that goes beyond the fact that harm is caused. Self-Concern is just part of a larger puzzle about the practical significance of first person thought. I have not attempted to solve this puzzle here, but in addressing one part, I have tried to show how the puzzle should be solved.

40. A classic treatment is Scheffler (1982: Ch. 4).

41. You may now object: if we can make sense of first-person reasons in relationships with others, why not extend this account to self-interest and my relationship with myself? Am I not my own friend? But the puzzle here is not specific to my approach: it applies to any view on which reasons to benefit friends and family are insistent, while reasons of self-interest are not.

Acknowledgements

For reactions to this material in earlier forms, I am grateful to three anonymous readers, to Stephanie Beardman, Alex Byrne, David Chalmers, Cian Dorr, Kit Fine, Johann Frick, Caspar Hare, Anja Jauernig, Mark Johnston, Tom Kelly, Jed Lewinsohn, Lisa Miracchi, Jessica Moss, Thomas Nagel, Gideon Rosen, Daniel Star, Sharon Street, Crispin Wright, to participants in my spring 2014 seminar at the University of Pittsburgh, and to audiences at MIT, NYU, Boston University, and Princeton.

References

Anscombe, Elizabeth (1963). *Intention* (2nd ed.). Blackwell.

Burgess, John (1983). Why I Am Not a Nominalist. *Notre Dame Journal of Formal Logic*, *24*(1): 93–105. http://dx.doi.org/10.1305/ndjfl/1093870223

Cappelen, Herman and Joshua Dever (2013). *The Inessential Indexical*. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199686742.001.0001

Evans, Gareth (1982). *The Varieties of Reference*. Oxford University Press.

Fine, Kit (2005). Tense and Reality. In *Modality and Tense* (261–320). Oxford University Press. http://dx.doi.org/10.1093/0199278709.003.0009

Frege, Gottlob (1997a). On Sense and Reference (Max Black, Trans.). In Michael Beaney (Ed.), *The Frege Reader* (151–171). Blackwell. (Originally published in German as "Über Sinn und Bedeutung" (1892). *Zeitschrift für Philosophie und philosophische Kritik*, *100*: 25–50).

Frege, Gottlob (1997b). Thought (Peter Geach and Robert H. Stoothoff, Trans.). In Michael Beaney (Ed.), *The Frege Reader* (325–345). Blackwell. (Originally published in German as "Der Gedanke. Eine Logische Untersuchung" (1918). *Beiträge sur Philosophie des deutschen Idealismus*, *I*: 58–77.)

Gendler, Tamar (2002). Personal Identity and Thought-Experiments. *Philosophical Quarterly*, *52*(206): 34–54. http://dx.doi.org/10.1111/1467–9213.00251

Hare, Caspar (2009). *On Myself, and Other, Less Important Subjects*. Princeton University Press. http://dx.doi.org/10.1515/9781400830909

Jeske, Diane (1993). Persons, Compensation, and Utilitarianism. *Philosophical Review*, *102*(4): 541–575. http://dx.doi.org/10.2307/2185683

Johnston, Mark (1989). Fission and the Facts. *Philosophical Perspectives*, *3*: 369–397. http://dx.doi.org/10.2307/2214274

Johnston, Mark (1997). Human Concerns Without Superlative Selves. In Jonathan Dancy (Ed.), *Reading Parfit* (149–179). Blackwell.

Johnston, Mark (2010). *Surviving Death*. Princeton University Press. http://dx.doi.org/10.1515/9781400834600

Kaplan, David (1989). Demonstratives. In Joseph Almog, John Perry, and Howard Wettstein (Eds.), *Themes from Kaplan* (481–563). Oxford University Press.

Kripke, Saul (1980). *Naming and Necessity*. Blackwell.

Lewis, David (1976). Survival and Identity. In Amelie Rorty (Ed.), *The Identities of Persons* (17–40). University of California Press.

Lewis, David (1979). Attitudes de dicto and de se. *Philosophical Review*, *88*(4): 513–543. http://dx.doi.org/10.2307/2184843
Marcus, Ruth Barcan (1961). Modalities and Intensional Languages. *Synthese*, *13*(4): 303–322. http://dx.doi.org/10.1007/BF00486629
McDowell, John (1979). Virtue and Reason. *The Monist*, *62*(3): 331–350. http://dx.doi.org/10.5840/monist197962319
Noonan, Harold (1998). Animalism versus Lockeanism: a Current Controversy. *Philosophical Quarterly*, *48*(192): 302–318. http://dx.doi.org/10.1111/1467-9213.00102
O'Brien, Lucy (2007). *Self-Knowing Agents*. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199261482.001.0001
Olson, Eric (1997). *The Human Animal*. Oxford University Press.
Parfit, Derek (1984). *Reasons and Persons*. Oxford University Press.
Parfit, Derek (1995). The Unimportance of Identity. In Henry Harris (Ed.), *Identity* (13–45). Oxford University Press.
Salmon, Nathan (1986). *Frege's Puzzle*. MIT Press.
Scheffler, Samuel (1982). *The Rejection of Consequentialism*. Oxford University Press.
Setiya, Kieran (2008). Practical Knowledge. *Ethics*, *118*(3): 388–409. http://dx.doi.org/10.1086/528781
Setiya, Kieran (2012). Knowing How. *Proceedings of the Aristotelian Society*, *112*(3): 285–307. http://dx.doi.org/10.1111/j.1467-9264.2012.00336.x
Setiya, Kieran (2014a). What Is a Reason to Act? *Philosophical Studies*, *167*(2): 221–235. http://dx.doi.org/10.1007/s11098-012-0086-2
Setiya, Kieran (2014b). Love and the Value of a Life. *Philosophical Review*, *123*(3): 251–280. http://dx.doi.org/10.1215/00318108-2683522
Setiya, Kieran (2014c). The Ethics of Existence. *Philosophical Perspectives*, *28*: 291–301. http://dx.doi.org/10.1111/phpe.12045
Shoemaker, Sydney (1968). Self-Reference and Self-Awareness. *Journal of Philosophy*, *65*(19): 555–567. http://dx.doi.org/10.2307/2024121
Shoemaker, Sydney (1999). Self, Body, and Coincidence. *Proceedings of the Aristotelian Society, Supplementary Volume*, *73*: 287–306.
Sider, Ted (1996). All the World's a Stage. *Australasian Journal of Philosophy*, *74*(3): 433–453. http://dx.doi.org/10.1080/00048409612347421
Stanley, Jason (2001). Hermeneutic Fictionalism. *Midwest Studies in Philosophy*, *25*: 36–71. http://dx.doi.org/10.1111/1475-4975.00039
Weil, Simone (2002). *Gravity and Grace* (Emma Crawford and Mario von der Ruhr, Trans.). Routledge. (Originally published in French as *Le Pesanteur at la Grâce* (1947). Librairie Plon.)
Whiting, Jennifer (1986). Friends and Future Selves. *Philosophical Review*, *95*(4): 547–580. http://dx.doi.org/10.2307/2185050
Williams, Bernard (1970). The Self and the Future. *Philosophical Review*, *79*(2): 161–180. http://dx.doi.org/10.2307/2183946
Williams, Bernard (1979). Internal and External Reasons. In R. Harrison (Ed.), *Rational Action* (17–28). Cambridge University Press.
Wittgenstein, Ludwig (1958). *The Blue and Brown Books*. Blackwell.
Wolf, Susan (1986). Self-Interest and Interest in Selves. *Ethics*, *96*(4): 704–720. http://dx.doi.org/10.1086/292796