Westfälische
Wilhelms-Universität
Münster

› Subjective Prudential Value —What is Left of It?

Annette Dufner

23

Preprints and Working
Papers of the Centre for
Advanced Study in Bioethics
Münster 2011/23

› Subjective Prudential Value —What is Left of It?

Annette Dufner

A classic view of the self-interest theory claims the following. I should prefer benefits for myself over benefits for others whenever I can do so without endangering the future cooperation of these others with myself. However, as the name of the self-interest theory already suggests, and as is often ignored, the actions it recommends rely to a significant extent on assumptions about the nature of the self. In order to know what lies in my self-interest I have to make assumptions about what my self is.

In recent years there have been various attempts to conclude from personal identity, or the nature and separateness of the self, to views about what the self-interest principle will deem to be a rational course of action, or what we have a special self-biased or self-referential reason to care about. Some authors, mostly inspired by Locke and Hume, have argued that the self is psychologically disintegrated across time. According to such views, my relationship with my own distant future Alzheimer self might be just as disconnected as my relationship with separate people. Some authors believe this reduces the plausibility of self-interested behavior as traditionally construed. Various other authors, partly from completely different philosophical traditions, have also presented views on what matters about the self that likewise have the capacity to change our views about practical reasoning. Some of them emphasize the partly social character of the deliberating and developing self—a strategy that will also endanger the idea of a sharp borderline and normative asymmetry between self and others that is assumed by traditional conceptions of self-interest. In this paper, I will try to provide a novel suggestion on these matters.

I would like to start the investigation by looking at a relevant feature of some classic dilemma cases. Such cases usually ask us to imagine what we would do if we could secretly violate moral rules to secure ourselves a benefit. Plato raises the question whether we would really always be inclined to stick to moral rules if we had the ring of Gyges that could make us invisible.[1] A more contemporary case, discussed for example by contractarians such as David Gauthier, asks us whether we would really stick to the promise to help a neighbor with the harvest, if our plan is to leave the community and emigrate right after the harvest.[2] These questions seek to demonstrate that there is a fundamental tension between the actions that the principle of self-interest and the actions that a more impartial moral principle will recommend. Whenever I can be immorally selfish without getting detected, I have prudential reason to do so, one might think.

However, these classic cases tend to ignore contextual facts about the self that can affect the intuitive responses to the cases, for example the idea that it is a natural fact about the self that preferences, aims, and recollections tend to fade and change over time, as well as the idea that people's social setting can shape their values and preferences in a relevant way. To demonstrate the work of these hidden premisses, consider a slightly modified version of the classic cases. The question might not be whether one would betray an aloof king, as in Plato, or a soon-to-be-left-neighbor to secure oneself an immediate benefit, but a more "personal" question:

> Imagine your parents responsibly decide to make their inheritance arrangements early on. As they start having conversations with you and your siblings about the matter and start organizing their affairs, you realize that a deceitful and ruthless plot against your brother—a person you have always been very close with—could secure you a number of extra household items in a couple of decades when your parents might die.

Even though this case is structurally similar to the classic cases, it seems to pump different intuitions. People tend to find it somewhat less plausible that going for the plot would be rational from a self-interested perspective. As I will seek to demonstrate, the reason for this is not merely strategic, but has to do with the fact that the classic cases ignore the potentially social character and sources of the self, as well as the natural disintegration of the self across time. Both of these aspects are relevant to what we are. Moreover, both of these aspects can be accommodated in a unified account of the self.

The Separateness of the Self makes Egoism Rational

According to a widely-held conviction it is the separateness of persons that makes egoism rational. This separateness, and subsequent non-identity, of different persons seems to be a precondition of the belief that it is always better if I myself receive a certain good than if someone else does. When the good ends up in my life there is allegedly something more immediate or direct taking place from my point of view that is unrivalled by cases in which others receive a similar or even a larger good.[3] This immediacy makes it rational to be more concerned about oneself

1 Plato, *Republic*, 359c-360e.

2 Gauthier, "Why Contractarianism?", in Vallentyne, *Contractarianism and Rational Choice*, p. 24.

3 This idea has been formulated in a particularly clear way by Henry Sidgwick. In *The Methods of Ethics* he writes: "It would be contrary to common sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently "I" am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of the

than about anyone else. The argument runs from the nature of the self as a descriptive fact, or the separateness of the self from others, to the plausibility of rational egoism.

There is also a further feature of the standard view about egoism. Since all parts of my existence are equally parts of myself, I have equal reason to be concerned about them all, the thought goes. I should subsequently not only display an entirely self-centred agent-relative attitude toward potential benefits, but also a time-neutral attitude when it comes to their distribution in my life. All that matters is that the benefits show up in some part of my life, whichever one that might be.[4] Both of these views have been formulated by Henry Sidgwick, among others.

There is also a standard view about the kind of psychological relation that makes the self separate from others. It tells us which property it is that applies only among the various stages of a single person's life rather than between separate people. According to this standard view a person at two points in time is the identical person if and only if there is *psychological continuity* between these two beings and no branching will occur. The no-branching requirement is supposed to exclude the notorious thought experiment involving hypothetical medical procedures that could cause a person's psychology to branch and continue in two separate strands in two separate bodies.[5] In a normal causal chain, involving no such hypothetical branching, though, the relevant kind of psychological continuity is generally understood as an uninterrupted chain of diachronic connections such as memories of the past, but also fear of future pain, the holding of intentions and plans for the future, or simply special care for the future self. The occurrence of such psychological traits among the various temporal parts of the self forms the allegedly particularly intimate relationship within which we stand with ourselves and that is different from our relationship with all other selves out there. It is this self-other asymmetry that seems to make egoism rational.

Personal Change Undermines Self-Preference

As indicated before, there are at least two ways of attacking this rationale with arguments from the "personal". First, one can try to argue that the internal cohesion of the self across time is weaker than egoists tend to think and that there should be a subsequent discount rate for self-bias when it comes to the good of one's temporally increasingly remote future self. If a person could either get one unit of happiness now or one unit of happiness in the distant future, it might be rational for this reason to prefer getting the benefit now rather than being time-neutral. If we apply a discount rate of, say, 20 percent to benefits that arrive in the distant future, then the one unit of happiness in the future has to get discounted to 0.8. Second, one can argue that the separateness of the self in relation to others is less complete than the egoist tends to think. Especially close personal relationships might be such an immediate part of one's life

existence of other individuals: and this being so, I do not see how it can be proved that this distinction is not to be taken as fundamental in determining the ultimate end of rational action for an individual." (Sidgwick, *The Methods of Ethics*, p. 498.)

4 This idea has also been formulated by Sidgwick when he insisted "[…] equal and impartial concern for all parts of one's conscious life is perhaps the most prominent element in the common notion of the *rational* […]." (Sidgwick, *The Methods of Ethics*, p. 124.)

5 For a contemporary formulation of the thought experiment see Parfit's case My Division in, *Reasons and Persons*, p. 254,55.

that the usual separateness of persons is diminished and our rational self-bias will consequently have to be less absolute. In other words, one can attack the potentially time-neutral attitude in self-biased reasoning and one can attack its obvious agent-relativity.

According to the first line of reasoning, it is a relevant fact that people tend to change their preferences and plans across time and that their recollection of past attitudes is often vague. In addition, there is of course a diminishing likelihood of anticipating one's future preferences correctly, but this might be a varying epistemic problem rather than a universal fact about the nature of the psychological. Due to the partial psychological disintegration across time, it is one thing to engage in a time-consuming plot to secure oneself a bunch of extra household items from one's parents, if one can get them now that one wants them. But it is another, arguably less plausible thing to make these arrangements, if one will not get them for another couple of decades. It should be rational to care somewhat less about certain kinds of benefits in the distant future than about similar goods in the nearer future. If one could get a particular benefit either now or in a couple of decades, it would consequently be rational to prefer getting it now rather than later.

Buying into this rationale does not imply that it would become rational to prefer the present absolutely and to begin acting in stupidly shortsighted ways. If one can get a small benefit now at the expense of a pretty certain, large disadvantage in the future, it would still be rational to avoid the large disadvantage. The reason is that the size of the disadvantage can outweigh the discount of its value that comes with the diminishing strength of the temporal psychological relations between the present and the future. For example, imagine Jones can get a benefit that will give her one unit of happiness now, but at the nearly certain cost of 100 units of grief in the future. Even if we apply a discount of, say, 20 percent to the future grief, she would still end up with 80 units of grief—a much larger loss than the 1 unit of happiness is a gain. Nonetheless, if she had to choose between receiving some sort of a benefit now or the equal benefit in the remote future, it would be rational for her to prefer getting it sooner rather than later.

Of course this rationale does not work for all kinds of goods. In particular, some holistic or perfectionist goods do not arrive at particular points in time. If a person wants to lead a successful professional life, it would be odd to develop a temporal preference about the exact arrival of this good. Instead, the person is probably hoping that success will characterize his or her life as a whole. To keep the focus of the investigation manageable, I will set this concern aside. Modern perfectionists believing that living a successful professional life is something one should aim for will generally acknowledge that there are other kinds of goods as well and chances are they can be brought to accept a discount rate for *those* kinds of good. If what has been said is true for at least some kinds of goods, then the idea that it is plausible from the point of view of self-interest to be equally concerned about all parts of one's future is wrong. Moreover, the reason for this would not just lie in the lack of epistemic access to one's mindset in the future, but in a general fact about the psychology of the self: the fact that one's self's preferences and other mental states tend to change across time, irrespectively of whether we know and expect this, so that the temporal parts of the self are often only loosely connected.

Consequence: Smaller Scope and Less Force for Self-Preference

In addition to believing that the separateness of persons makes egoism rational, and that rationality requires equal concern for all parts of one's life, classic authors also worried about the relationship between self-interest and morality. For example, Sidgwick worried about a dualism of practical reason, according to which it is always rational to act morally, but unfortunately

also always rational to act in a self-interested way. This is rather bad news for morality, one might think. Cases like the ones mentioned in the introduction push the concern further by insinuating that it might sometimes even be *more* rational to be selfish.

Arguably, though, accepting a discount rate for self-interested concern means that the dualism of practical reason becomes less fundamental: If we can allocate a benefit to ourselves instead of someone else in our remote and only loosely related distant future, then the force of our prudential reason to do so will become weaker. If my distant future self is only loosely related to my current self, this should reduce my self-interested bias in favor of this future self. As a result, the tension between the rational degree of self-bias on the one hand and morality on the other hand will get reduced. Since self-bias is then only plausible to a lower degree, there can also only be a lower tension between self-bias and morality. If the degree of plausible self-preference is subject to a discount rate, so will the tension between self-preference and morality. A tension can of course still arise, but only with reduced strength.

Personal Relations Undermine Self-Preference

The second line of attack against the problem that self-interest and morality will sometimes prescribe different actions pursues an opposite route. It does not deny the unity of individual selves across time, but rather the idea of sharp borderlines between different present selves. In ordinary life, people sometimes say that certain important others turned them into what they are. According to this way of talking we are not entirely self-directed isolated individuals, but rather parts of a more extended psychological network. This way of viewing the self does not focus on the allegedly intimate psychological relation within which we stand with ourselves. Instead it emphasizes the similarity between our own relationship with ourselves and the psychological relations that apply between what we ordinarily think of as separate selves.

Various groups of thinkers have suggested such views about the self. Neo-Aristotelians sometimes literally acknowledge friends as "other selves", while communitarians stress the primacy of a social setting as a prerequisite for the personal preference and belief formation of individuals. Participants in the agency debate and neo-Kantians have claimed that what occurs within an individual agent during a deliberation and decision process can be compared to the relations within a group agent or a political society.

Without wanting to defend such views in detail, here are some examples. Inspired by Aristotle's view about friends as other selves, David Brink insists that psychological features such as continuity of believes, desires and intentions that characterize an individual's mental life across time are relevantly similar to the psychological relations that hold among separate persons. Separate persons can likewise come to share believes, desires and intentions. Their shared mental life is then characterized by the same kinds of psychological relations that apply within individuals. Carol Rovane arrives at a related view when analyzing the metaphysical nature of persons. As she ends up arguing in *The Bounds of Agency*, the agency-regarding relations that she takes to characterize the nature of persons apply both in individual and in social cases. In both cases, the relations are relevantly similar in kind, she argues. Christine Korsgaard also relies on a related assumption when comparing the agency-related deliberative decision process in individuals to political decision procedures in an ancient city. Just like the population of a city-state, she finds, an individual can be of a divided mind and has to make legislative decisions among many conflicting inclinations in order to be able to produce actions. I do not want to base the upcoming discussion on any particular one of these positions, even though I might occasionally use the Aristotelian terminology of other selves, or the communitarian idea of the

social sources of the self. The relevant aspect they all share and that the discussion depends on is the view that first personal and social psychological relations can be similar in kind.[6]

Consequence: Bigger Scope and Less Force for Self-Preference

One may now argue that this conviction is able to reduce the room for tension between the self-interested imperatives of egoism and the other-regarding imperatives of morality. To see how this can work in detail, it is important to remember where philosophers generally construe the borderline between egoism and morality. Egoists allegedly believe that they never have reason to care about others for their own sake. They only care about them for *instrumental* reasons. Their goal is to be perceived as cooperating members of society so that others will help them in furthering their own self-biased interests. This point is uncontroversial. Everybody agrees that even brute egoists generally have such an instrumental or *strategic* reason to act morally. This strategic consideration is counter-factually instable though.[7] Whenever it is possible to maintain the appearance of cooperation without actually cooperating, while continuing to enjoy the benefits, rational agents seem to have a supreme reason to do this rather than to actually cooperate.

The conviction that first person and social psychological relations are similar in kind can lead to a position that is counter-factually more stable than the instrumental consideration. Imagine a group of other selves who are engaged in a shared system of cooperation. One day, Smith discovers that he could maintain the appearance of cooperation and continue sharing the benefits without actually contributing. The strategic perspective cannot prevent that this appears rational. However, when considering the nature of the group members as his other selves, things seem to change. Since the relationship in which he stands with his extended selves is similar in kind to the relationship in which he stands with himself, betraying the others would to a significant extent be similar to betraying himself. If it were not for the shared thoughts and the influence of others, he might not have his current desires in the first place. Acting to the disadvantage of those who have played, and presumably will continue to play an integral role in what he wants in the first place does not appear as rational as at first sight.[8] If this is convincing, Smith has a *self-referential and at the same time non-instrumental reason not to betray others*.[9]

6 Aristotle, *Nicomachean Ethics*, book IX; Brink, "Rational Egoism, Self, and Others"; Brink "Rational Egoism and the Separateness of Persons"; Rovane, *The Bounds of Agency. An Essay in Revisionary Metaphysics*; Korsgaard, *Self-Constitution. Agency, Identity, and Integrity*.
Thinkers like Charles Taylor have in a sense gone even further. For him, modern conceptions of an individual's mind, or the idea of an individual thinker who is extrinsic to and more or less watching his or her train of thoughts, is actually the result of a social process or a history of ideas. On his view this apparently entirely private or "inward" perspective on ourselves might not exist in this way if it were not for a long social history providing the path for it. (Taylor, *Sources of the Self. The Meaning of Modern Identity*.)

7 This way of putting things in the present context has been formulated by Brink, "Rational Egoism, Self, and Others," p. 344–49.

8 Arguably, they can continue to play such a role even if she never sees them again as Gauthier's example mentioned in the introduction stipulates.

9 The idea of self-interested while at the same time non-instrumental reasons to care about others is not news. Broad raised it in the context of his discussion of self-referential altruism. (Broad, "Egoism as a Theory of Human Motives".)

This does not mean that any action by an egoist that benefits another person and that happens for non-instrumental reasons will create the desired overlap with morality. To appreciate the point, it is important to remember that even egoists can sometimes have an accidental preference to give to other people and to see them happy, just like they might have a preference to see the flowers on their window sill flourish. In this case giving to others might be non-instrumental, but it will not happen out of a genuine concern for the other person. Instead it will satisfy the egoist's self-interested preference to see other people happy, to be surrounded by smiling faces so to speak.[10] Moreover, outwardly altruistic behavior can be a great way of achieving a particular kind of narcissistic self-satisfaction and flattering social recognition. When observing a person helping others, it is therefore hardly possible to know for sure whether the motivation and the reasons are genuinely and purely altruistic or mostly self-interested—in fact, the agent him or herself might not even know.

Nonetheless, according to the traditional understanding an egoist never cares about others for their own sake rather than for the sake of satisfying his own accidental desires. If we believe in other selves, though, the matter can appear differently. Since others are our extended selves who play an integral part in our self-development and preference formation, even an egoist has reason to care about them for their own sake. I will say a few more things about this later.

So far, we can say the following. If the argument is correct, the thesis of the social sources of the self can shift the borderline between egoism and morality. The overlap among the recommendations that egoism and morality generate for particular cases will be prone to increase. It will become less plausible to betray other selves in the classic dilemma cases from above. Moreover, the reason for this would not be instrumental, as the uncontroversial strategic consideration suggests. Instead, the reason would be self-referential, while at the same time being non-instrumental.[11] Betraying the others would be like betraying a part of oneself.

To make this more specific, one may want to add that— just like concern for one's own future self—concern for other selves might come with a discount rate depending on social or psychological proximity. The reason for such self-interested and non-instrumental care might be greater the closer people are with each other. It might apply to a particularly strong degree to close family members and friends. It might also still apply in the cases of more distant family members or of people at the other end of town, albeit to a reduced degree. If an egoist can benefit another person, and there is no competing benefit for himself, then he might consequently have a greater reason to give the benefit to a close family member than to somebody at the other end of town.

10 Economists and rational choice theorists often ignore this difference about motivations and reasons. They tend to believe in some form of an unrestricted preference satisfaction theory according to which whatever a person prefers will then be in his or her self-interest. Many philosophers disagree. If a person prefers a benefit to go to someone else and actually cares about that other person *for his or her own sake*, then philosophers are often hesitant to describe the action as self-interested. Moreover, philosophers who believe in preference satisfaction theories tend to favor *restricted* versions over unrestricted versions of the theory: If a person prefers a benefit to end up in another person's life, then they tend to view the benefit in the other person's life rather than the preference satisfaction as the relevant good. Assuming that the preference satisfaction about the allocation rather than the preference satisfaction that comes from the enjoyment of the benefit is the relevant good in such cases is not the only plausible position one can hold. If one believes the latter, then the allocation is of course not self-interested in the restricted sense.

11 One may also stick to Broad's terminology and speak of "self-referential altruism" here. (Broad, "Egoism as a Theory of Human Motives".)

Depending on the moral theory one endorses, morality and this form of self-referential egoism might recommend the exact same when it comes to entirely other-regarding choices and there is no competing benefit for the egoist himself. This would be the case whenever the moral theory one endorses acknowledges the exact same inter-personal discount rate for other-regarding moral obligations as self-biased reasoning under the premise of the thesis of other selves suggests. For example, morality might demand that we care x times more about persons who are close to us than about others. If the degree of self-referential psychological relatedness that we share with our close ones is also x times higher than our relatedness with others, then morality and self-referential egoism would have to demand the very same actions.

Nonetheless, there might still be cases in which morality and self-referential egoism do not demand the exact same degree of concern for someone. For example, morality might demand other-regarding actions that give weight to others over-proportionally to the degree of our psychological relatedness to them. In such a case, an egoist would not necessarily be willing to act according to the morally required degree of concern. This would in particular be the case if the action that morality demands would be in conflict with another action that would bring more benefit to the egoist or his closer extended selves—a problem I will say more about later.

It is possible to employ the notion of compensation to support the points discussed so far. Initially, egoists think they will experience compensation if they deny themselves a present benefit in order to achieve a similar or a larger benefit in the future. At the same time they believe that a benefit to someone else, including close ones, can never compensate them for a personal burden. Their beliefs about compensation initially tend to be time-neutral, but agent-relative in a very narrow sense. If the egoist can get convinced, though, that very distant future benefits might not always compensate them fully, they will have to give up their time-neutral attitude. If they can get convinced that they can experience compensation if a benefit goes to an other self, they will have to adjust their previously too narrow agent-relative conviction.

The first argumentative strategy, according to which our preferences change in relevant and unpredictable ways across time, seeks to shrink the gap between rational egoism and morality by *shrinking* the temporal scope for justifiable self-bias. The second strategy tries to accomplish the same by *increasing* the scope for self-bias to other individuals in our social setting.

Objection: We Really Only Care Instrumentally for Other Selves

Some people will not be inclined to accept this kind of twisting and turning of the borderline between egoism and morality. One central objection one may raise is that benefiting other selves for the sake of them as mere extensions of our own self, is not a particularly moral thing to do. Morality requires that we care about others for the sake of them *as others*. Since this is partly how we usually understand the borderline between egoism and morality, a concern for others on the ground that the relations one shares with them are similar to the relations that apply in one's own life might not create the desired overlap between the two principles of practical reason. Caring for them as extended selves and caring for them as others would be two very different things.

This is a serious concern and it requires some further specifications. One might have to think of the thesis about the social sources of the self as a two step process. In a first step, everybody in a social network is likely to have instrumental reasons for being interested in others. Everybody in these systems can benefit from mutually shared and developed practices, projects, or ways of explaining the world. In a second step, however, these relationships as such may well give everybody an additional *non-instrumental* reason to care. The reason is that the similarities

and causal influences people come to share with their social network are just like the similarities and causal influences they exert on themselves. More importantly, people would never have let others gain this kind of influence over themselves, had they not in the first step also recognized them as helpful *others*. Once they have granted others this influence over themselves, they have selectively endorsed some of their influence as part of their own self-development. The others now represent both a source of external influence and a part of one's own personal development and preference formation.

If one accepts this reasoning, and if such self-interested and non-instrumental reasons to care about others exist, then the tension between egoism and morality that the classic dilemma cases invoke will indeed lose some of their force. People would have a counter-factually stable, self-referential *and non-instrumental* reason to care about and not to betray members of their social setting, even if they could allocate benefits to themselves instead.

The Me-ness Objection and First-Person Memories

There is also a further important objection that threatens both the attack on time-neutrality and on the extreme agent-relativity in self-biased reasoning. This objection claims that there is a more or less metaphysical fact of identity or me-ness that is not susceptible to the temporal vagaries and social aspects of a person's self development. Take the example from above in which a person considers the foolproof betrayal of a close brother to secure a distant and questionable inheritance. The example is supposed to raise the question whether there might sometimes be a peculiar self-related reason to refrain from securing oneself additional possessions rather than granting them to someone else. As I have argued so far, this reason could then be seen as self-referential while at the same time being a non-instrumental. Someone wanting to deny the existence of such peculiar self-referential and non-instrumental reasons might then be tempted to react as follows.

> *The Me-ness Objection.* Of course I might have changed my preferences by the time the inheritance comes and of course the plot might change an idealistic aspect of my relationship with my brother, but this does not change the fact that *I* will get the inheritance instead of my brother if I go ahead.

According to this reaction the fact that a future self will be me is an additional and decisive reason in favor of the plot that goes over and beyond the unpredictable vagaries of my personal psychological development. Even if I do not want the objects anymore by the time I get them and even if my brother is a major force in my personal development, it will nonetheless be the case that *I* will get them, rather than anyone else. From the perspective of rational self-interest it may be claimed that I have supreme reason to engage in the plot for this reason. If this is what we believe, then the inheritance case is merely another version of the classic cases that seek to demonstrate that there is an unbridgeable gap between the principle of self-interest and the demands of morality.

If the me-ness objection is ultimately convincing, then the plan to create more overlap between egoism and morality with the arguments against time-neutrality and agent-relativity might not work out as easily. The reason is that an egoist could always insist that me-ness speaks in favor of individual self-bias as it is ordinarily construed. However, people compelled by this thought will have to tackle the following difficulty: Whenever we want to explicate what this metaphysical me-ness actually is, a straightforward answer will of course point at the kinds of psychological relations again that we have been reducing things to all along: continuity

of psychological experiences, beliefs, desires and intentions, for example. A proponent of the objection will have to give a different or an additional account.

It is well-known that Derek Parfit has tried to present a host of arguments to the effect that a metaphysical fact of me-ness should not matter in our practical deliberations. I am not going to rehearse this debate here.[12] We may or may not agree with his opponents that me-ness has some normative force. As long as we believe that the psychological relations I have been discussing have *more* normative force, the argument can still stand. It is only at risk if we believe that me-ness is the only thing that matters or that the force of me-ness is always overwhelming.

I do not believe that the normative force of the me-ness objection without appeal to the psychological relations that underlie it can be overwhelming. Most of the compelling force of the me-ness objection seems to rest on the experience that—despite all the considerations discussed so far—psychological features appear particularly immediate or vivid in the first person case. This immediacy might constitute something like an irreducible me-ness component. The notion of memory becomes interesting in this context if one assumes that at least some of our memories, or at least one aspect of our memory, has a distinct phenomenological character that is usually only accessible "from within" a particular first person chain of experiences.[13] For example, it is often thought that qualia, the content of our sense impressions, can only be accessed by individuals themselves, not by their friends. The experience of what the difference between red and green is like for Joe is in this sense an entirely subjective fact. Epistemic access to what exactly it is like to remember particular sense impressions "from the inside" or what particular events felt like "from the inside" might be restricted to individuals. If this is the case, then the me-ness objection has some force and individual persons will in a way be rather separate from each other.

Nonetheless, even if we admit that some psychological events have such an irreducible first person character that is only accessible "from within", it is nonetheless the case that these memories tend to fade or even change across time. In fact, since they are entirely personal memories and cannot get verified by consulting others, they are particularly prone to fade or shift as time passes by. This means they can probably not serve the purpose of justifying a time-neutral form of self-bias. Nonetheless, they might still provide egoists with a rationale for some self-bias across certain periods of time. If one wanted to justify a time-neutral view on the basis of this distinct first person aspect of some memories, one might have to argue that there is a continuous subject of experiences that is having those memories across time, even if those memories fade or change. The idea of such a subject of experiences is not exactly new. The suggestion here at hand, however, claims that there might well be a distinct and irreducible first person character of certain psychological events, while at the same time insisting that this first person character is time-relative.[14]

It seems relevant to point out that this position about the time-relativity of a potentially irreducible fact of me-ness in psychological relations is compatible with what has been said earli-

12 For a denial of Parfit's position as developed in *Reasons and Persons* and elsewhere see for example Whiting, "The Non-branching Form of What Matters".

13 I owe this point in part to *-deleted-*.

14 There are large debates in the philosophy of mind regarding the reducibility of the mental. It is a miracle to me why these debates and the debates about the reducibility and normative force of diachronic mental relations are not brought together in some such way more frequently. Despite some influential works, this still seems to be an area in which the specialization and division among philosophers into theoretical and practical camps has gone too far.

er about sharing psychological relations with other selves. Other selves can still share memories with each other to a certain extent, namely whenever we can plausibly construe these memories as not containing an entirely subjective element. For example, two friends might remember having attended a particular wedding together. They might remember having missed the train on the way there and having joked about their traffic problem with uncle Joe after the ceremony. Such recollections would not include memories of what one of the friends thought to herself during the mess-up and did not communicate. They would also not include objects that could only be seen from a particular spatial perspective onto the dinner table.

If this position works, then it is possible to accept the me-ness objection, while nonetheless maintaining that practical self-bias ought to be mostly time-relative. An egoist would have reason for self-bias in the strict sense of the term within that time-frame within which the discussed first person memories of experiences, beliefs, desires and intentions can occur. Outside this time-frame the ground for such self-bias would fall apart.

It is also important to point out another important fact. If the force of the psychological relations is supposed to be more overwhelming than the force of the me-ness relation, then we have a very good reason for combining the two strategies discussed so far: If the me-ness relation plays only a subordinate role when it comes to particular normative judgments, then we do not just have reason to insist that the relevant first-personal relations tend to fade across time, but also that there is now no metaphysical reason anymore why some of those self-constituting psychological relations that matter for practical purposes should not in principle be shareable by separate selves. While Parfit has devoted most of his work on this topic to arguments for the former, it would actually be incoherent to think that it will not also lead to the second effect. He only hints at this extremely briefly in an article entitled "Comments"[15] and in an unpublished manuscript. If the force of the me-ness relation declines, it has to decline in all respect, not just in respect to time. It also has to decline in regard to how sharp the borderline between my *present* self and the *present* self of my closest companions is to be construed for practical purposes. It is therefore by no means arbitrary to combine the approaches discussed. They both seem to be natural consequences of a certain position about the lower practical force of the me-ness relation when it comes to particular normative judgments from a self-referential perspective.

Self-Referential Prudential Value—What is Left of It?

The two strategies of reducing the tension between egoism and morality focus on cross-temporal and cross-personal psychological relations respectively. Each of them can in principle come with a discount rate: cross-temporal relations tend to change or fade, and the strength of our cross-personal relations might depend on how close we are with someone. In addition to discussing the two kinds of relations separately, we can now also ask which *relative* amount of weight they have for our self-understanding and the subsequent plausibility of self-biased actions in comparison with moral actions.

The question we set out with was the following: Is it rational to betray a close family member you're on good terms with in order to secure yourself a questionable benefit in the distant future? The answer to the question will turn on whether the cross-temporal relations between our present self and our future self or the cross-personal relations between ourselves and our

15 Parfit, "Comments," *Ethics*, 1986, 96: 832–872, p. 871.

other self, the family member, are intrinsically more important for us. When deciding about the relative importance of these relations we can consider various possibilities. I will focus on three. According to the first option, one could argue that, as the cross-temporal relations fade, both cross-temporal and cross-personal relations can become equally strong and important. Second, one could argue that cross-temporal relations are always somewhat stronger and more important than any relationship with others. Third, I will discuss the possibility that, as the cross-temporal relations fade, cross-personal relations might sometimes actually become stronger and more important.

Let's look at what the first option would yield, the option that cross-temporal and cross-personal relations can in some situations become equally strong.[16] This would mean that the discount rate for cross-temporal personal relations can bring the ground for special care about one's own future self either all the way down to zero or at least down to the very same level to which we have special reason to care non-instrumentally about other selves. The idea is that cross-temporal relatedness—including phenomenological experience memory—tends to fade over time. The relatedness between myself now and myself in the distant future might in fact be so faint that the self-referential relations between myself now and the self of a separate person are equally strong.

This would imply the following. On a short-term or intermediate basis it can still be possible for self-preference to arise and be justifiable. During periods of strong cross-temporal self-relatedness, the relatedness with other people will have to be seen as less strong and self-preference will consequently appear rational. However, this self-preference would generally only be plausible within a limited period. For example, at the age of 45 this time period might extend 25 years into the past and 25 years into the future. At the age of 50, the time period would then likewise extend 25 years into the past and 25 years into the future. Within this period intentions and experiences that were formed in the past are still very vivid and continue to shape our anticipations and plans for the future.[17] Within this forward moving time frame self-preference will be justifiable due to the particularly strong cross-temporal connections that hold within it.

However, outside this time frame the relationship between one's present self and the future self can be weak enough so that the self-referential relationship between one's present self and the self of a separate person is just as strong. For example, the psychological relationship between oneself at the age of 45 and oneself at the age of 85 might be rather weak, especially if the 85 year old suffers from clinical forgetfulness as in the unfortunate, but common case of dementia. This would have the following further implications. When deciding whether a particular benefit should go to oneself one year from now or to the present self of another person,

16 There is some evidence that Parfit holds this view. For example, in section 114 of *Reasons and Persons* he argues, "[a] reductionist is more likely to regard this child's relation to his adult self as being *like* a relation to a different person" (emphasis added). This passage seems to imply that the psychological relations between temporally remote parts of an individual self and two different selves can be of equal strength and importance. (Parfit, *Reasons and Persons*, p. 335.)

17 It seems important to include a part of the *past* within the current selfhood units of the two persons. If we had to decide at this very moment between benefiting our own future self and the *present moment* of another person, the decision might not turn out the same way, if we do not include a part of the past in our consideration. The reason is that only very few causal psychological relations will hold between one's own *present moment* and the *present moment* of a significant other. The ground for self-referential concern about significant others is not only based on present similarity, but also on causal facts about the past. See Hurka, *Virtue, Vice, and Value*, p. 202: "In all their forms, the virtues of loyalty are a response to facts about the past."

there would be room for rational self-bias. Over the course of a year the first person relations are still very strong, so that self-bias could speak in favor of benefiting oneself. On the other hand, when deciding whether a particular benefit should go to oneself forty years from now or to the present self of another person, things will appear differently. The self-referential relationship with an important other self might be just as strong as the relationship with oneself forty years from now. Consequently, there would be no rational room for self-preference in a traditional sense left.

For example, imagine Mr. X had to decide whether to benefit his own self forty years from now or the current self of his wife. The psychological relationship between X's current self and his own self forty years from now might be so weak that the relationship between his current self and the current self of his wife is just as strong. This means that X has just as much self-referential and non-instrumental reason to benefit the current self of his wife as his own self forty years from now. There is no room for rational self-preference in a traditional sense anymore in the decision X is facing.

Of course the size of the expected benefits would be important in addition to the relevant psychological relations. Imagine X had to decide between benefiting his own self *twenty* years from now or the present self of his wife. Presumably within the time frame of twenty years there would still be some room for ordinary self-preference left, albeit to a reduced degree. But if the benefit that could go to the wife would be a lot larger than the benefit that his own self in twenty years would get, then the balance could actually tip in favor of the wife. The small amount of ordinary self-preference that would still be plausible in this case can arguably get outweighed by the large size of the benefit that could go to his wife with whom he is only slightly less related than with his own self twenty years from now.

There are further questions to be asked. For example, we may have to ask whether there should be a further discount rate for the temporal proximity between X's present self and the increasingly remote future selves of his wife. The question might not be whether X should give the benefit to his wife's *present* self, but rather to her *future* self. If the choice is between X's own self twenty years from now and his wife's self twenty years from now, it seems unlikely that the basis for ordinary self-preference has disappeared. On the other hand, if the choice is between X's own self forty years from now and his wife's self *three* years from now, then there might actually be little room for self-preference in favor of X's own self. The reason to benefit the remote part of himself might be so low that the reason to benefit his wife three years from now is just as high.

I will not go into further detail about these matters, but it is worth spelling out once more what a view according to which cross-temporal and cross-personal relatedness can sometimes become equally strong and important would imply for the tension between egoism and morality, or between self-referential reasoning and the other-regarding demands of morality. It would mean that courses of action that benefit one's own future self would remain prudentially plausible within the time frame within we are particularly strongly psychologically related to ourselves. If the benefit would only arrive after the end of this time frame, a course of action that would benefit someone else could be equally prudential.

Such a position seems to have the advantage of granting to egoists that there is something special about first-person relatedness, especially phenomenological self-relatedness. This is a conviction that egoists feel very strongly about and might not be willing to give up. At the same time, one can hold that there is now a non-instrumental self-referential reason to act in favor of others: The reason is that having done so will be just as good from a self-biased perspective. Apart from first person phenomenological relations, psychological relations to close others are

like the relationship we have with ourselves. And since first person relations tend to fade and change, those cross-personal relations can sometimes be just as important for oneself.

I would now like to turn to the second option. According to this second option, cross-temporal relations in our own lives are always at least somewhat stronger and prudentially more important than cross-personal relations.[18] This view is more intuitive than option one for people who believe that one has always an at least somewhat stronger prudential reason to be concerned about oneself than about others, and it is closer to the traditional conceptions. In terms of easing the tension between egoism and morality, this option does somewhat less work, though. If the cross-temporal relatedness in our lives is always stronger and prudentially more important than any connections across people, then the egoist's rationale and morality remain to a larger extent at odds. The idea that there is always somewhat more cross-temporal psychological relatedness in one's own life will presumably be a sufficient reason for egoists to have a general attitude of self-preference in all cases.

Nonetheless, the strength of this egoist rationale might often be weak. One's cross-temporal relationship to one's own distant self can be so weak that it is *almost* as weak as the relationship toward other people. This means when deciding whether a particular benefit should go to one's own self or the self of another person, the calculation may tip in favor of the other person. This could be the case whenever the benefit the other person could receive is significantly larger than the competing benefit for oneself. However, if the benefits were equal, egoists acting according to this theory would see overriding reason to benefit themselves in all cases.

I am now going to turn to the third option. This third alternative would claim that the cross-personal relations can in some cases actually become stronger and prudentially more important than the cross-temporal relations within one's own life, even though the cross-temporal relations will usually still prevail in our reasoning within a time frame of, say, 25 years into the past and 25 years into the future.[19] This view claims that there is room for ordinary self-preference within this time period. Over larger stretches of time, however, the vividness of our past-directed memories fades and our future-directed intentions change. Outside this forward moving time-frame, the psychological relations we share with others would become *more* important, not just equally important.

The first option claimed that the cross-personal relations could in this case become equally strong and important as the cross-temporal relations in our own life. But one could also argue that the cross-personal relations can sometimes be *stronger* and *more* important than the

18 Both Derek Parfit and David Brink have anticipated this alternative. In section 115 of *Reasons and Persons* Parfit writes: "Those claims [about compensation] treat weakly connected parts of one life as, *in some respects, or to some degree*, like different lives" (Parfit, *Reasons and Persons*, p. 337, addition and emphasis added). This formulation does *not* say that the weak connections between remote parts of one's own life are *as weak as* the connections between one's own life and other lives, although other passages of the book leave this possibility open. Instead, it says that weakly connected parts of one's own life are only *to some degree* like connections to other lives. The idea seems to be that the similarity between intra-personal and cross-personal relations can become close, but intra-personal relations will always be somewhat stronger and more important. Brink seems to have had a similar option in mind. There are passages in which he argues that there is a discount rate for our interest in extended selves, but never for the various temporal parts of our own self. (Brink, "Rational Egoism and the Separateness of Persons," Postscript; section 8; pp. 121, 128, 133–34.)

19 Neither Brink nor Parfit have anticipated this alternative. Brink dismisses relations that can in principle only hold in the life of an individual self altogether and he rejects a discount rate for cross-temporal relations. Parfit raises the possibility that cross-personal relations might also matter, but does not consider the possibility that they can become stronger and prudentially more important than our normal cross-temporal relations.

faint recollections of current times and the significantly changed plans and preferences one will have 25 years from now. Why could my psychological relatedness to a close friend not be much stronger and more important than my relatedness to an entity forty years from now with which I might not even share basic convictions anymore? Moreover, in the unfortunate case of dementia this entity might not even remember what happened five minutes ago, let alone what it was like to be me now.[20]

In the quest for reducing the tension between egoism and morality the proposed view does more work than the other two alternatives discussed. Alternative one suggested that cross-temporal relations can sometimes become as weak as cross-personal relations. In those cases the egoist has equally strong reason to benefit his own future self or the self of another person. In all other cases, the discount rate makes sure that the plausibility of self-preference is somewhat smaller than in a time-neutral view, but it remains nonetheless larger than the self-referential bias in favor of others. Alternative two did worse. It suggested that cross-temporal relations are always more important than cross-personal ones. According to this alternative, the plausibility of self-preference and any subsequent tension with morality would decline due to the discount rate, but apart from this the plausibility of self-preference and the fundamental tension between egoism and morality would prevail.

In both of these alternatives egoists will have *some* reason to benefit their other selves. However, in both views egoists will only have *overriding* reason to benefit others if the benefit that could go to them would be larger than the one that they could secure for themselves. Alternative three, however, sometimes gives egoists *overriding* reason to benefit others, even in cases in which the expected benefits are equal.

For cases with equal benefits, alternative one was able to establish that it is at the most just as rational to benefit others as oneself. In the classic dilemmas cases this would at the most eliminate the impression that it would be *more* rational to benefit oneself than to act morally. It does not offer a decisive reason *in favor* of acting morally, though. In this sense the dualism of practical reason would remain intact. At best, it would appear rational to act morally, but it would likewise remain rational to act selfishly.

Of course alternative one, and even more so alternative two capture modern conceptions and intuitions about self-interest better than the more far reaching alternative three. Nonetheless, it appears to me to be defensible to claim that there are a few cases in which we have a self-referential and non-instrumental reason to benefit others that overrides any reason to act self-interested in the traditional sense of the term, even for cases with equal benefits.

20 One may wonder how much of this the well-known branching case discussed at the beginning of this paper already implies. It may well be said that the pre-branching person has a stronger reason to care about the non-identical post-branching selves than about his own self thirty years from now had he not branched. This means the branching case implies that one might have a stronger reason to benefit the self of "another person" in the near future rather than one's own self in the distant future under normal conditions.
There are two differences between the proposed approach and what the branching case implies, though. First, comparing one's attitude toward a potential branching off-shoot with one's attitude toward a distant future self after an ordinary development amounts to comparing attitudes toward of two *future* selves, the future post-branching beings and the person's own self thirty years from now. I am arguing that a choice between benefiting the *present* self of another person and one's own self thirty year's from now should turn out in favor of the other person. Moreover, and maybe more importantly, the branching incident involves an abnormal cause. The relevant consequence from the proposed view will prevail though if we add a restriction of normal causes. If we insist that both the distinctly intra-personal relations such as first-person memories as well as the sharable relations such as beliefs, desires and intentions are only prudentially relevant if they have a normal cause, the proposed view will remain unaffected—a fact I take to be a desirable feature of a normative view.

The best way of making this view compelling consists of course in picking a case in which the first person psychological relations are very weak and the cross-personal relations are particularly strong. So take the relationship between yourself now and your late stage Alzheimer self. Jeff McMahan appropriately described late stage Alzheimer patients as "isolated subjects."[21] They are conscious and able to function in many ways, but they have lost their sense of being temporally extended beings. They have mostly no memories and no past, and their anticipations and intentions for the future are forgotten the minute they were formed. Moreover, they often display a rather changed character. Nonetheless, there seems to be a continuous chain of at least synchronic, or momentary, conscious experience between your present and the Alzheimer self. While McMahan argues this gives us reason for special first person concern for one's future isolated self[22], it also seems defensible to claim that this kind of concern can get outweighed by a much stronger self-referential and non-instrumental concern for significant others in the present.

While the future Alzheimer subject might strictly speaking be the same person, and while one might have a special responsibility to ensure it will be able to lead a decent life, it is psychologically speaking a stranger. It is even hard to imagine what it will be like to be such a being. In comparison, the closest and most durable personal relationships in people's lives appear a lot more intimate. Close friends can often read each other's minds without speaking. Such friends can quietly look at an object together and both remember an experience they had with it. If one were to present one's own future Alzheimer self with the object, nothing would happen. This asymmetry can even occur if there is no Alzheimer's disease in the picture and we are merely dealing with very strong forgetfulness. The psychological relations between oneself and an intimate friend can consequently be a lot stronger than the psychological relations between oneself in the present and oneself in the distant future. The friends can share and influence each other's memories, beliefs, desires and intentions. These relations are sometimes much diminished between people in the present and the future selves they will come to be.

The basis of self-interested actions are intimate, diachronic psychological relations. Such relations can occur in varying kinds and strengths in contexts that are quite different from what a prima facie understanding of self-interest sometimes suggests. Nonetheless, looking at the psychological basis of self-referential reasoning does not seem to be able to resolve the conflict between egoism and morality entirely. In many situations self-referential reasoning will recommend self-preference as ordinarily assumed, while morality might recommend something else. However, in some situations, especially when significant others and long time periods are involved, self-referential reasoning can actually speak against what a traditional understanding of self-interest would suggest. Egoists will then have a self-referential and non-instrumental reason to benefit others. Even in the moderate first alternative, a larger size of a possible benefit to a close person can tip the scale. If one is willing to accept the third alternative, according to which the relevant first person relations can actually become weaker than the relevant cross-personal relations, the scale can even tip if the benefits are of equal size. In these cases, it would literally be non-instrumentally more rational from the viewpoint of self-referential reasoning to benefit others.

21 McMahan, *The Ethics of Killing*, p. 55, 65.
22 McMahan, *The Ethics of Killing*, p. 65.

Literature

Aristotle. 2000. *Nicomachean Ethics*, translated by Roger Crisp. Cambridge: Cambridge University Press.

Brink, David. 1990. "Rational Egoism, Self, and Others." *Identity, Character, and Morality*, edited by Owen J. Flanagan and Amélie O. Rorty, 339–78. Cambridge: MIT Press.

Brink, David. 1997. "Rational Egoism and the Separateness of Persons." In *Reading Parfit*, edited by Jonathan Dancy, 96–134, Oxford: Blackwell.

Broad, C. D.. 1971. "Egoism as a Theory of Human Motives." In *Broad's Critical Essays in Moral Philosophy*. London: George Allen & Unwin.

Gauthier, David. 1991. "Why Contractarianism?" In *Contractarianism and Rational Choice: Essays on Gauthier's Moral by Agreement*, edited by Peter Vallentyne, 13–30. Cambridge: Cambridge University Press.

Hurka, Thomas. 2001. *Virtue, Vice, and Value*. Oxford: Oxford University Press.

Korsgaard, Christine. 2009. *Self-Constitution. Agency, Identity, and Integrity*. Oxford: Oxford University Press.

McMahan, Jeff. 2002. *The Ethics of Killing*. Oxford: Oxford University Press.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, Derek. 1986. "Comments." *Ethics*, 96:832–72.

Plato. 2006. *The Republic*, translated by R. E. Allen. New Haven: Yale University Press.

Rovane, Carol. 1998. *The Bounds of Agency. An Essay in Revisionary Metaphysics*. Princeton: Princeton University Press.

Sidgwick, Henry. 1981. *The Methods of Ethics*. Seventh Edition. Foreword by John Rawls, Indianapolis: Hackett Publishing.

Whiting, Jennifer. "The Non-Branching Form of What Matters." In *The Blackwell Guide to Metaphysics*, edited by Richard M. Galen, 190–218, Oxford: Blackwell.