Why It Does Not Matter What Matters: Relation R, Personal Identity, and Moral Theory

Bastian Steuwer

Preprint, please cite published version:
Steuwer, Bastian (2020). Why It Does Not Matter What Matters: Relation R, Personal Identity, and Moral Theory. *Philosophical Quarterly* 70 (278):178-198.
https://doi.org/10.1093/pq/pqz064

Abstract:

Derek Parfit famously argued that personal identity is not what matters for prudential concern about the future. Instead, he argues what matters is Relation R, a combination of psychological connectedness and continuity with any cause. This revisionary conclusion, Parfit argued, has profound implications for moral theory. It should lead us, among other things, to deny the importance of the separateness of persons as an important fact of morality. Instead, we should adopt impersonal consequentialism. In this paper, I argue that Parfit is mistaken about this last step. His revisionary arguments about personal identity and rationality have no implications for moral theory. We need not decide whether Relation R or personal identity contain what matters if we want to retain the importance of the separateness of persons.

Keywords: personal identity, separateness of persons, what matters in survival, Derek Parfit.

I. Introduction

Derek Parfit famously argued that personal identity is not what matters for prudential concern about the future. Instead, he argues what matters is Relation R, a combination of psychological connectedness and continuity with any cause. This revisionary conclusion, Parfit argued, has profound implications for moral theory. It should lead us, among other things, to deny the importance of the separateness of persons to morality. Instead, we should adopt impersonal consequentialism. In this paper, I argue that Parfit is

mistaken about this last step. His revisionary arguments about personal identity and rationality have no implications for moral theory. The importance the separateness of persons has for morality does not turn on whether personal identity rather than Relation R is what matters for prudential concern.

When spelling out the moral implications of his view on personal identity and what matters, Parfit mentions a variety of examples. The examples range from revising our views on paternalism and autonomy, abortion, promises and commitments, retribution and desert, and the importance of equality to the separateness of persons objection to utilitarianism. My discussion will be focused on the importance of the separateness of persons objection. The reason for this is that the separateness of persons objection occupies a central place in non-consequentialist moral thinking (Gauthier 1963: 125-6; Nagel 1970: 133-40; Rawls 1999: 23-26; Nozick 1974: 32-3). Utilitarianism ignores the separateness of persons, the argument holds, because it aggregates all benefits and burdens across different persons. Sometimes, however, we are allowed to aggregate different benefits and burdens. In particular, we are allowed to aggregate when these benefits and burdens fall within one life. This is explained by the unity of the individual. Together the separateness of persons and the unity of the individual demand that we should treat inter-personal trade-offs differently from intra-personal trade-offs. Utilitarianism cannot do this.

Parfit's revisionary arguments concerning morality can be reconstructed as attacking both components of the separateness of persons objection to utilitarianism. One argument holds that Parfit's views on personal identity and what matters undermine the unity of the individual. I examine and reject this argument in Section II. Another argument holds that Parfit's views on personal identity and what matters undermine the separateness of persons. I examine and reject this argument in Section III. A last argument holds that his views render the unity of the individual and the separateness of persons less relevant. I examine and reject this argument in Section IV.

Throughout this paper my strategy is to accept Parfit's arguments concerning personal identity and rationality, and to reject the link he draws from metaphysics and rationality to morality. My strategy thereby differs from what Mark Johnston has called 'minimalism'. Johnston remarks that many of our practices, like those of morality and

rationality, lend themselves to certain metaphysical views. Minimalism then holds that the justification of these practices is independent from the truth of the metaphysical position. Metaphysical positions, like those about personal identity, are epiphenomenal to practices like rationality and morality (Johnston 1992, 1997).[1] Unlike Johnston, I believe that the truth of metaphysical positions can have an impact on normative practices like morality. Indeed, I accept, at least for the sake of argument, that Parfit is correct about his link between metaphysics and rationality. I only deny that Parfit's revisionary argument for morality stands.

II. First Argument: Less United Individuals

So why should Parfit's conclusion about the metaphysics of personal identity have any impact on morality? The first suggestion is that Parfit's claim that what matters is Relation R rather than personal identity demonstrates that the unity of the individual is unimportant. When discussing the diminished importance of the separateness of persons, Parfit writes: 'If the unity of a life is less deep, it is more plausible to claim that this unity is not what justifies maximization' (1984: 334-5). In rational decision-making we are allowed to pursue what will bring about the highest sum-total of well-being. If Parfit is right in holding that the unity of the individual is less important, then this cannot be justified by appealing to the unity of the individual. Maximizing the sum-total of well-being would then seem to be justified differently and apply also in inter-personal trade-off, in line with what utilitarianism demands.

One way to explain Parfit's claim that the unity of the individuals in undermined, is by appealing to what we may call the relevant units of moral and prudential concern. Some nationalists claim that nations have moral importance over and above individuals. Nations matter for their own sake. A nation is, under such a view, a unit of moral and prudential concern. For 'moral individualists', the unit of moral concern is a person's entire life. But other proposals are possible. We could focus on parts of lives (I shall call these 'person

[1] Similar arguments are made by Wolf (1986), Adams (1989), and Christie (2009).

stages'), or we could focus on time-slice persons, instances in an individual's life without much temporal extension.[2] Parfit at some point writes that following his view we should 'regard the rough subdivisions within lives as, in certain ways, like the divisions between lives' (1984: 333-4). Elsewhere he speaks of the 'partial disintegration' of persons (1984: 335-336). This seems to suggest that Parfit regards parts of lives, in particular those with high degrees of psychological connectedness, as the basic units of moral and prudential concern. The idea is that the psychological connections that contain what matters come in degrees. Some of these connections wither away over time. We are more strongly connected to our past and future person stages close in time.

Moral theories should therefore take person stages, rather than full lives as their objects of principal concern. For example, questions of distributive justice would then arise between person stages rather than between persons. This means that principles of distributive justice would need to be given a greater scope. They would also extend to trade-offs within a person's life, namely to those between one person stage and its future successive person stage (cf. Parfit 1984: 332-4). But since principles of distributive justice would then apply to such a variety of cases, we might think that we have less reason to care about distributive justice. Our intuition that distribution matters is less strong for intra-personal, inter-stage trade-offs. Yet it is unclear why this is the conclusion we should draw from the widening of the scope of principles of justice. Instead of revising our intuition about the importance of distributions, we could revise our intuition about intra-personal, inter-stage trade-offs. Perhaps we trust this intuition less since it might derive from traditional views about personal identity that Parfit rejects. If we take this answer to the problem, then we would revise our view on individual rationality. We would no longer be justified to pursue the maximum benefit when facing trade-offs that only affect our life. Principles of distributive justice would be extended to all trade-offs involving different person stages (cf. Nagel 1979: 124-5 fn. 16).

This looks like a stand-off between two different ways to adjust our intuitions. However, the defender of Parfit's view has another argument in hand. Talking about person stages is only a useful heuristic. Person stages are united by greater psychological

[2] See: Shoemaker (1999: 391-2) and Brink (1997a: 110-6) for similar distinctions. Shoemaker, following Parfit, calls these persons in parts of their lives 'selves', Brink calls them 'person segments'.

connections and what matters are these connections (Relation R). Introducing person stages can help us to avoid talking about Relation R directly, but it is an imperfect heuristic. The boundaries between different stages are fuzzy and different stages overlap. Once we see this, it is less plausible to just apply our moral principles to different units. To substitute one unit of moral concern for another overlooks the fuzziness around the borders of these moral units. There does not exist a unity of a person stage that is comparable to the unity a defender of the unity of the individual has in mind. We can then rightly ask why we should attach such great importance to the difference between different units when the units are only useful heuristics to refer to persons in different stages of their lives.

But why should we follow Parfit in believing that Relation R leads us to accept person stages as the unit of moral concern? Why should we believe that entire lives are not strongly integrated? Broadly speaking there are two possible arguments, contra Parfit, that lives are strongly integrated. One argument is that something other than Relation R unifies persons. Consider, for example, Kantian replies to Parfit's claims (Korsgaard 1989; Blackburn 1997). These replies can admit that persons are neither metaphysically united nor united via Relation R. Instead, there is something else that unites persons. The Kantian response to Parfit highlights that persons are united by the practical perspective and the necessity to act as agents. In a similar vein, David Brink argues that agency is best ascribed to persons rather than person stages. Considering person stages to be agents would lead to an undue proliferation of various, overlapping agents (1997a: 110-6, 121-3). The second argument, on the contrary, does not introduce any further considerations over and above Relation R that could explain why individuals are unified. Instead, this argument rejects the claim that Parfit's arguments have established that Relation R fades out over time. In line with my general strategy of granting Parfit his claims about metaphysics and rationality, I pursue an argument of the second sort.

Relation R is the relation of psychological connectedness and/or continuity with any cause. Psychological connectedness refers to the *degree* to which the same psychological features are present in two different persons at different times. The psychological connectedness between me, now and me, two seconds ago, is very high. The psychological connectedness between me, now and me, two years ago is lower. I have forgotten some

experiences, do not share all of my beliefs, adjusted my plans of life, and so on. Psychological continuity requires a series of overlapping bonds of strong psychological connectedness. Continuity does not require, however, that connectedness is given between earlier and later stages in the series. As such, psychological continuity, unlike psychological connectedness, is a transitive relation.

The idea that Relation R weakens over time requires an interpretation of Relation R in which Relation R comes in degrees. Only then will person stages show a significantly greater extent of R-relatedness than entire lives. I already mentioned that psychological connectedness is a matter of degree. But is psychological continuity as well? Parfit contrasts connectedness as a relation that comes in degrees with continuity indicating that he does not believe that continuity is a relation that comes in degrees (1984: 206). Nevertheless, Brink offers a construal of continuity in which continuity is a matter of degree. According to Brink, two persons are more strongly continuous with one another if the individual connections in the chain of psychological connectedness that constitutes continuity are stronger (1997a: 132, fn. 31, 1997b: 138, 141-3). An immediate problem for such a view is that continuity is transitive. Parfit defines continuity as a transitive relation in order for continuity to be a possible criterion of personal identity. Since personal identity is transitive, continuity must be as well (1984: 206-7). Continuity is thus defined as transitive precisely to express a form of connection that the non-transitive relation of connectedness does not express. The problem for Brink's view is now that transitivity is defined only as a property of binary relations and not defined for relations that come in degrees.

We can make sense of the suggestion that continuity comes in degrees in another way. We can imagine a family of continuity relations which each specify a different threshold of connectedness that is needed to ensure continuity. A person stage is then more continuous with a past or future person stage if a higher threshold of connectedness is met. For example, continuity$_{\text{STRONG}}$ requires that all overlapping chains consist of strong connectedness, continuity$_{\text{VERY-STRONG}}$ requires chains of very strong connectedness, continuity$_{\text{EXTREMELY-STRONG}}$ requires extremely strong connectedness. Two person stages might then be more continuous if continuity$_{\text{EXTREMELY-STRONG}}$ rather than continuity$_{\text{STRONG}}$ holds between them.

This construal of continuity is a threshold view. According to this view, the weakest link determines the strength of the degree of continuity of the entire chain. The degree of continuity for an entire life is therefore determined exclusively by the amount of connectedness in the moment where the greatest change occurred. This does not cohere well with the reason for which continuity was introduced. Continuity is distinguished from connectedness to explain the psychological connection between persons over a long period of time. Continuity can explain how an old person is psychologically connected to her childhood person stage. But then it does not appear that it should matter very much how these changes occurred.

Take the example of St Paul who according to the biblical story fell on the road to Damascus, heard the voice of Jesus, and decided to stop his prosecution of Christians and convert. St Paul's story is one of a single sharp change. Contrast this with a person who lives an erratic life and changes her life's narrative multiple times. Finally, towards the end of her life she, like St Paul, arrives at a point that is very different from her early person stages. As long as none of the individual changes in her life were as drastic as St Paul's conversion, she would, following the current proposal, be more continuous than St Paul. While it may make sense to think that St Paul's life has not been fully continuous, it makes little sense to think that the erratic life has been more continuous than St Paul's. St Paul's life has a clear narrative that is only disrupted by a single incident. No clear narrative exists for the person with the erratic life. Given that continuity is supposed to account for long-term relations, it seems hardly plausible that degrees of continuity should be so sensitive to single points in time. The reason why continuity is distinguished from connectedness as a separate relation is better accounted for by understanding continuity as an on-off relation.

The argument for person stages as the unit of moral concern therefore cannot rely on an analysis of psychological continuity. But I have admitted that psychological connectedness comes in degrees. If we attach primary importance to psychological connectedness, then we can argue that person stages are the relevant unit of moral concern. If, on the other hand, we attach little significance to psychological connectedness, then we have no grounds for believing Parfit's argument that person stages are the relevant unit of

moral concern. In such a case my previous argument has shown that psychological continuity would ensure that we regard entire lives as the proper unit of moral concern.

While for most parts of his argument, Parfit does not distinguish between the two components of Relation R, we can now see that the different is important. So what is Parfit's argument that psychological connectedness is an important part of what matters? His argument here is very brief. He analyses three components of psychological connectedness/continuity to see whether we care about being connected instead of merely continuous (1984: 301). The first component is memory. If only continuity mattered, then '[i]t should not matter to me that I shall soon have lost all of my memories of my past life.' But this is implausible. Indeed, we care heavily about retaining our memories. Also in terms of desires and intentions Parfit claims that we want more than continuity. Our lives should have an overall unity and should not be episodic with continued fluctuations. Thirdly, Parfit claims that there are parts of our character that we do not want to change. Here again, he claims, connectedness matters.

To assess Parfit's argument, it will be helpful to make the case a bit more concrete. We can take a case where psychological continuity is given but psychological connectedness is low. Alzheimer's is such a case.[3] A person before the development of Alzheimer's is psychologically continuous with the person having developed Alzheimer's. But their psychological connectedness is limited. The person has forgotten many of the memories she once had. It is also likely that many of the intentions or long-term plans that the person had will have changed or she simply will have forgotten them. Maybe there will be further changes due to Alzheimer's that reduce psychological connectedness. If the person used to be very engaged with intellectual activities, her character will inevitably change when the illness leads to a decline of her reasoning skills. As noted earlier, Parfit has argued that in these cases we do seem to care about our connectedness with these persons. I agree with this to some extent. But I think Parfit relies here on an ambiguity in the locution 'what matters'.[4]

[3] We can leave aside complications of late stage Alzheimer's where all psychological connections to one's previous life are cut and so there is no continuity either. Some authors, notably Jeff McMahan, have held that we also have grounds to be rationally concerned with a future Alzheimer's-Self who is not even psychologically continuous with us (cf. McMahan 2002: 65).

[4] A similar observation about the ambiguity of 'what matters' is made by Peter Unger. Unger uses the terms 'desirability use' and 'prudential use' for the contrast (cf. Unger 1990: 92-7).

It matters to us *that* or *whether* these changes happen. The thought of Alzheimer's is truly frightening to many, including me, and we would strongly want to avoid it.

Yet Parfit needs another claim to support his argument. He needs to say that connectedness matters *once* or *when* these changes happen. In other words, he needs the claim that connectedness constitutes the basis for rational self-concern. But here it does not seem plausible to me that we would lose the special bond with the resulting person once we develop Alzheimer's. If someone told us that the person with whom we will be continuously connected will be tortured in the future, we would rightfully be horrified. It would concern and involve us deeply. If we hear that a stranger that is qualitatively similar to us will be tortured, we may have sympathy but will not be as involved as in the previous case. Now how should we react if we hear that a person with whom we will be continuously connected but who will develop Alzheimer's will be tortured? Parfit's claim that connectedness matters should make us be less worried or concerned about this news. We should treat it more like the news of the stranger. Yet I cannot see why we should not react with the same horror and concern to the news as in the case of our continuous self who will not develop Alzheimer's.

Let me illustrate my distinction further with an analogy. Parents often want children to turn out a specific way. At the very least they would like their children to be successful and happy. This matters tremendously to them. But parental love does not relinquish when children do not meet this standard. It does not matter to parents that their child is not successful once this is the way things are. They do not lose the special bond of concern with their children if these happen to be unsuccessful and depressed. Similar things hold in the self-regarding Alzheimer's case. Of course we would prefer a future without Alzheimer's. But this does not mean that we give less importance to our bond with our psychologically continuous Alzheimer's-self.

A second reason to think that psychological connectedness is not what primarily matters is the following. Psychological connectedness will be very high when there is a great overlap in our psychology between past and future selves. But we certainly do not want our life to be static. Even if we are content with ourselves and cannot identify specific parts of ourselves to be changed, we still would want to develop and grow as persons. Parfit to some extent agrees with this general observation, but he remarks that we want our life to have a

certain overall unity. The life should not be episodic (1984: 301). But here similar arguments like the ones I raised before apply. I can concede that we do not want that our life will be episodic. Such a life would not have the requisite unity or narrative that we strive for. We might even think that such a life could not be a good life. In short, we do not want *that* this happens. But does this also mean we shall lose all special bond or interest in the person who is at the end of our episodic journey through time? I doubt that. Our intuitions about the unity of life are intuitions about what makes a life good, but we will still be concerned with our path through life even if our life is deficient in some sense.

We can make the remarks about change more precise. Some decisions are very likely going to result in psychological changes in the personality of the person making the decision. Take the decision of a young adult from a working-class background whether or not to go to university. If she goes, the would-be student will experience a new social setting very different from the one she is used to. She will be exposed to ideas and avenues radically different from those she would have encountered otherwise. This is confirmed by reports of a culture shock for students from working-class backgrounds in higher education. She can foresee that the university experience will change her. It is foreseeable for the decision-maker that one option will lead to significant psychological changes. Psychological connectedness will hold only to a reduced degree between the decision-maker and her future self. This change will only happen, however, if one of the two options is chosen. Assuming that psychological connectedness is the primary part of what matters, this influences the rational assessment of these decisions. The decision to engage in the transformative experience will be less appealing. Any potential gains of higher education will have to be discounted by the fact that we should have less prudential concern for the resulting person. The expected benefits of going to university would need to be very substantial to counteract the lessened concern. This does not strike me as a plausible model for thinking about these kinds of decisions.[5]

The argument becomes even more pressing in the special case when we regard the change positively as an improvement.[6] We do want to change some of our psychological features and would not regard their disappearance as a loss in any way. However,

[5] See: Wolf (1986: 712-3) for a similar observation.

[6] This possibility is also discussed by Brink (1997a: 119-21) and Korsgaard (1989: 120-3).

successful improvement of our psychological features would render us less psychologically connected with our past self. If psychological connectedness expresses what matters, then we should have less prudential or anticipatory concern for our successive improved self. In a way this even undermines the rationality of efforts made in order to improve one's character. These efforts are borne out of a concern for an eventually resulting person that will be psychologically less connected with the person having made the sacrifices. If we should have less rational concern for this resulting person, these efforts may not be worthwhile after all.

There is one feature about the improvement argument that might seem problematic. David Shoemaker objects that, contrary to what I have been assuming, cases of improvements do contain a significant degree of connectedness. Most importantly, there is a shared intention of wanting to improve one's character and life. This intention connects these parts of one's life strongly together. The strong connection is evidenced by the fact that we can identify with our past self in a way that we cannot with an even more remote self, like our past self before we made the decision to change our life (1999: 406-9).[7]

Shoemaker's point is apt for deliberate decisions to improve one's character. But not all improvements need to involve an intention. Earlier, I described decisions which can have a transformative impact on the decision-maker. It seems possible that there are decisions where the decision-maker can foresee that the decision will have a positive transformative impact yet does not choose the option because they intend the improvement. Take the example of parenting. Let us assume that a person foresees that being a parent will induce positive character changes, for example by becoming a more responsible person. But the decision to become a parent may have been made on grounds entirely independent of these changes. In this case, the improvement of the future parent's character is not intended and therefore there is no intention that connects the self of the future parent with the later improved self. Here, too, the fact that the future parent will be less connected to her later self should not make undergoing the improving experience any less rational.

[7] For a more extensive discussion of Shoemaker's point on identification, see: Shoemaker (1996: 328-31). He extends on a point made earlier by Parfit (1971).

What is present, however, is a second-order desire by the future parent to be a more responsible person. This second-order desire is fulfilled in the case of first-order psychological change while it is frustrated in the case of first-order psychological stagnation. While the second-order connection does hold over time in cases of improvement, many other first-order psychological connections will be weakened. Psychological connectedness can accept cases of improvement only if there is a reason why we should privilege second-order psychological connections over first-order psychological connections.

One suggestion here is related to the idea of self-identification. The idea is that there is a sense of alienation towards those first-order desires that we rather not have while there is a sense of self-identification towards those first-order desires that we wish to retain. Alienation and self-identification do provide us with good reasons for regarding some desires as more properly our own than others. Parfit, when he discusses self-identification, draws a contrast between self-identification and non-identification. Non-identification is marked by an attitude of indifference towards a past self. Indifference in turn is marked by the absence of feelings of pride, shame, regret and the like (1971). This analysis of self-identification does not privilege desires that we approve of over those we disapprove of. Shame and regret for having certain desires can just as well provide for self-identification. I think Parfit is right in this construal of self-identification. We talk about people owning up to one's mistakes. A person repentant of a former self that did wrong is not regarding this former self as alien to herself. On the contrary, it would be difficult to understand the intensity of feelings of remorse and guilt if the person would not identify the former self as genuinely herself. Of course, sometimes there is a feeling of alienation from our first-order desires. But alienation is not the same as disapproval, the two can diverge. Since this is the case, the importance of self-identification cannot give us a reason why second-order desires are more important psychological connections than first-order desires. This in turn means that improvements do not necessarily ensure psychological connectedness. The reply to the improvement argument fails. It seems then that Parfit's case for psychological connectedness as a central part of the relation of what matters does not stand.

This concludes my discussion of connectedness and continuity. We should interpret Relation R as giving primary weight to psychological continuity as opposed to

connectedness. We can retain Parfit's central claim that personal identity is not what matters. What matters instead is Relation R. Psychological continuity can, unlike personal identity, be one to many, as illustrated by fission cases where one brain is divided and inserted into two different brainless bodies. Both resulting persons will then be psychologically continuous with the original person whose brain was divided (Parfit 1984: 254-60). But, as it turns out, in our world this difference is not relevant. We do not divide or branch in our real world. For us, personal identity perfectly coincides with psychological continuity. Unlike psychological connectedness this does not come in degree but is an on-off relation. The appropriate unit of moral and prudential concern therefore remains an entire life. The unity of the individual is safeguarded by the unity of psychological continuity.

III. Second Argument: Less Separate Persons

The arguments canvassed so far have sought to undermine the unity of the individual. But instead of focusing on the unity of the individual, we could focus on the separateness of persons. Consider the following famous passage in which Parfit describes his own attitude after coming to believe the reductionist view.

> 'There is still a difference between my life and the lives of other people. *But the difference is less. Other people are closer.* I am less concerned about the rest of my own life, and more concerned about the lives of others' (1984: 281, my emphasis).

One way to interpret this passage is that Parfit is suggesting that we can have similar relations to other contemporaneous persons as we have to our future selves. This includes Relation R which contains what matters. If the way we are related to our future selves is similar to the way we are related to other distinct people, then this reduces the extent to which we are distinct from other persons. Jennifer Whiting and David Brink have suggested, in a similar vein, that our relation to our future person stages is like the relation to close friends or family (Whiting 1986; Brink 1997b: 136-43). If this is so, then we would no longer be justified in putting such great weight on the separateness of persons as a bar to inter-personal aggregation.

There are many different ways individuals can be psychologically related to us. These correspond to the different important features of our mental lives. Sharing memories, intentions, beliefs, or dispositions are ways to be psychologically related. To be one of the psychological relations included in Relation R, the relation has to have a causal component. It is not sufficient that two persons are very much alike in terms of their psychological characteristics. The causal component in Relation R is important to distinguish numerical identity from qualitative identity. Sometimes older people say things like 'you remind me of myself when I was young'. This is a statement about qualitative identity. The older person sees many of the features of her own psychology when she was young in the other person. But this psychological resemblance is clearly not sufficient to establish numerical identity.

With regard to causal psychological connections, we should draw a distinction between those connections that are first-personal and those connections which are not first-personal. By first-personal I do not mean that the connections have to be had by the same person. Rather, I understand a first-personal connection as a non-deviant causal connection between first-personal mental states.

The contrast here is between mental states that are 'from the inside' and mental states 'from the outside'. It means that connections are presented in the first-personal mode of presentation (cf. Parfit 1984: 220-2). The connections must be part of a self-centred scheme of one particular point of view. The distinction is best explained with regard to memory. I might have the memory of hearing Parfit speak. The memory is detached from the person Parfit, just as in a dream we sometimes see ourselves from a third-person perspective. This memory is markedly distinct from a memory in which I seem to recall having Parfit's body and voice and am speaking at All Souls College. This second memory is had 'from the inside'. The memory is one in which I occupy Parfit's self-centred perspective of the world. It is not just that I am imagining how All Souls looked and Parfit's voice sounded. Rather, it is, in Williams's words, participation imagery from the perspective that Parfit occupied (Williams 1973: 43-4; see also Velleman 1996: 48-50). A second example is the link between an intention and a subsequently carried out action. Intentions entail a first-person perspective, they are intentions that the agent performs an action. Intentions are 'inside' of a particular self-centered scheme (cf. Velleman 1996: 70).

Memories or intentions of this sort need not presuppose personal identity. Parfit introduces a revision of the concept of memory, originally proposed by Sydney Shoemaker, that he calls quasi-memories (Parfit 1984: 220-2; Shoemaker: 1970). In quasi-memories the subject seems to remember an experience from the first-personal point of view, someone had this experience, and there is a non-deviant causal connection between the experience and the memory. Similarly, for someone to have a quasi-intention, one has to have an intention, a subsequent action has to be performed, and the intention must cause the action in the right way.

To see the importance of causation in the case of memory, consider a case in which a person who has been in an accident forgets about her experience. At a later point in time a skilled hypnotist implants imagery of an accident into the minds of her audience. By sheer coincidence the hypnotist's imagery of the accident corresponds perfectly to the imagery of the actual accident. Such a case is clearly not one of remembering or quasi-remembering the accident (Parfit 1984: 207; the example is due to Martin and Deutscher 1966: 174-5). The causal connection is even clearer to see in the case of intentions. What is special about intentions is that intentions can lead directly to actions without agential interference. Intentions are causes of actions (Parfit 1984: 261).[8]

With these clarifications in mind, the question arises whether psychological continuity and therefore strong psychological connectedness, requires first-personal connections. The first thing to note is that the examples that Parfit gives as elements of Relation R tend to be first-personal connections. For example, when introducing the relations of psychological connectedness and continuity Parfit introduces them after a discussion of quasi-memories and quasi-intentions (1984: 204-5). This gives an indication that Relation R appears to be a plausible criterion for what matters in large part because it contains first-personal connections.

While this is indicative, there are other arguments which strengthen the case for the centrality of first-personal connections. Consider the relation between you, now, and a future person who happens, by fortuitous coincidence, to have the same character, habits

[8] In both cases we need further conditions that rule out deviant causal chains. For the case of intentions, see: Davidson 2001: 74-82. For the case of memory, see: Martin and Deutscher 1966: 178-91; Sidelle 2011: 744-8.

and other psychological features as you. In short, the person is qualitatively very similar to you. This relation is not Relation R and does not contain what matters. There are two elements missing in this case. One is the absence of a causal link between your psychology and the future person's psychology. The other element is the absence of first-personal connections. Which one of these two elements explains more satisfactorily why qualitative similarity is insufficient for Relation R?

It is hard to see why causation by itself should make such a big difference. There is no obvious reason why causally sustained psychological connections should be particularly important. There is nothing intrinsic to causation that suitably connects with our concerns of what matters. It is difficult to see how the mere fact that some connections are causally sustained could explain what distinguishes fortuitous psychological connections by accident from ordinary cases of personal identity.

Why causation is important is even more puzzling given that Parfit thinks that any causal link would be sufficient to satisfy psychological connectedness (1984: 282-287). If what matters is the effect and not how it was caused, why does it matter that is was caused in the first place? One possibility is that something associated with the causal requirement explains why the relation has to be a causal one. In this case then, it would be this extra factor rather than the causal link as such which explains why the relation between the two persons contains what matters. Ernest Sosa and Jeff McMahan provide arguments of this kind (Sosa 1990: 309-13; McMahan 2002: 60-6). Sosa argues that what explains why causal connections are important is that one important causal connection is non-branching survival. Survival for Sosa refers to being the unique closest continuer of a person. McMahan argues that causal connections are important if they have a physical realiser: the continued existence of enough of one's brain. I agree with the spirit of these arguments that something associated with causation explains the causal requirement. However, unlike Sosa and McMahan, I seek to provide an answer that is consistent with Parfit's own view on what matters.

When I introduced quasi-memories and quasi-intentions as examples of first-personal connections, I highlighted that both are defined as causal notions. Quasi-memories and quasi-intentions must both stand in the right kind of causal relation to previous

experiences or subsequent actions. Without any causal relations then, there cannot be quasi-memories or quasi-intentions.[9] Crucially, in contrast to the generic causal link, it is easy to see why first-personal connections add something significant to mere qualitative similarity. First-personal connections express what distinguishes one's psychology from the psychological make-up of others. We can call this a person's *distinguishing psychology*. Distinguishing psychology is opposed to both *generic psychology* and *core psychology*. Generic psychology refers to the parts of one's psychology that are instantiations of generic psychological features which one shares with others, like character traits or habits. Core psychology refers to the psychological capacities that persons have.[10]

By conveying one's distinguishing psychology, first-personal connections contain what sets oneself apart from others. They express a non-generic sense of 'you' and demark what is special about you. This links well with what matters. The relation of what matters captures a special bond that we have to persons precisely because of what sets them apart; what makes them different and special. Our distinguishing psychology is thereby closely connected to a sense of self. It provides for the possibility of self-identification. As I discussed earlier, when we self-identify, we acknowledge events or persons in time to be of special importance to us. In the example of mere qualitative similarity, it is this basis for self-identification that is missing. The absence of first-personal connections is the more plausible explanation why the relation with a qualitatively similar person fails to contain what matters.

There is another reason in favour of the view that the absence of first-personal connections satisfactorily explains why the relation of mere qualitative similarity does not contain what matters. When introducing first-personal connections, I highlighted that first-personal connections are connections that are 'from the inside' and which provide us with a self-centred perspective on the world. This self-centred perspective is closely related to what matters. It provides us with a perspective from which our projects and ambitions are carried

[9] Alan Sidelle makes a related point about Parfit's discussion of whether what matters is Relation R with any cause, a reliable cause, or its normal cause. Sidelle argues that Parfit's discussion is best understood as rejecting the view that there are any further causal requirements over and above those inherent in the causal psychological connections that constitute Relation R (2011).

[10] The distinction refines the contrast Unger draws between core psychology and distinctive psychology (1990: 67-71). Unger somewhat misleadingly uses the term distinctive psychology for both distinguishing and generic psychology.

out. The continuation of this perspective provides us with the basis for our special concern with our projects and ambitions. The first-personal perspective also explains why we are rightly involved and anticipate experience of future person stages. We can anticipate from the first-person perspective (cf. Velleman 1996: 67-76).

We should conclude that first-personal connections are a central component of Relation R. They explain why Relation R requires causal connections between psychological features, provide for a sense of self-identification and provide us with a self-centred scheme from which we experience the world. Given the centrality of first-personal connections for psychological connectedness, we should further conclude that strong connectedness requires at least some first-personal connections. Strong connectedness is in turn needed for psychological continuity. If, following my argument in Section II, psychological continuity is primarily what matters, then non-trivial R-relatedness requires first-personal connections.

We might imagine two persons regularly exchanging quasi-memories, quasi-intentions, and other first-personal connections via telepathy. Similarly, in cases of fission the two resulting persons would share many quasi-memories, quasi-intentions, and other first-personal connections. These two persons would exhibit strong psychological connectedness and continuity. But aside from these science fiction examples, it is hard to see how in our world quasi-memories, or other first-personal connections, could be shared between separate persons. I know of no mechanism in our world that ensures that first-personal memories or intentions can be shared. And I certainly know of no mechanism by which we can share first-personal memories, intentions and so on, over a prolonged period of time. In our world then, strong connectedness, a requirement for continuity, cannot plausible be met between separate individuals.

IV. Third Argument: Less Importance to Persons

In the previous two sections, I examined and rejected arguments that respectively sought to undermine the unity of the individual and the separateness of persons. We can defend the unity of the individual and the separateness of persons. I will now examine a

third argument. Rather than disputing the unity of the individual or the separateness of persons, it disputes that the separateness of persons or the unity of the individual have moral importance. Parfit's reductionist views on personal identity should give us reason to attach less significance to persons. Parfit argues that a person's existence can be reduced to facts about mental and physical events. Over and above these facts, there does not exist an entity like a Cartesian Ego or a soul (cf. 1984: 219-28, 236-8, 245-52). Because a person's existence just consists in facts about mental and physical events, there is less that is involved in the fact of personal identity. This should give us grounds to care less about the fact of personal identity (cf. 1984: 337-8, 340-1, 1995: 28-41). The argument relies exclusively on Parfit's reductionist answer to the question of what a person is and does not rely on his more specific claims about what matters for prudential and anticipatory concern. We can still believe that reductionism about persons should lead us to adopt an impersonal morality, even if we think identity is what matters prudentially. This line of argument, while often overlooked, deserves scrutiny.

Let me now turn to the argument. What are the reasons for believing that persons matter less under the reductionist view? Parfit describes facts about personal identity as being a 'deeper truth' under the non-reductionist view. He points out that many of us would attach great significance to a separate existence over and above our body and related mental and physical events. Since personal identity is less important, we should also attach less significance to the separateness of our respective existences. This consideration does not seem decisive. Various authors have pointed out that their belief in the moral (and prudential) importance of persons has not diminished even as they have become convinced of a reductionist picture of the person (Wolf 1986: esp. 705-8; Adams 1989: 454-60; Johnston 1997: 159). Their reason for assigning importance to persons depends on the centrality of persons for our projects and social surroundings. It depends on how persons relate to their future and to others. It never depended on there being a separate entity who is this person.

There is a stronger argument for the reductionist critique. If we are reductionist about personal identity, then we can express every fact about personal identity in another way. We can re-describe these facts as impersonal facts about mental and physical events. But if these facts are just equivalent to the more ordinary, impersonal facts, then it is unclear

why we are justified in ascribing greater significance to facts about personal identity than we do to the equivalent impersonal facts. Following this argument it is not so much the absence of a Cartesian Ego that makes the difference, but rather the availability of an alternative, impersonal description of one's life. If these two ways of describing one's life are indeed equivalent, then we should be suspicious whether the added significance we attach to persons is indeed justified.[11]

Mark Johnston provides an objection to this kind of argument. He dubs this line of reasoning by Parfit the 'argument from below'. The argument from below seems to hold that facts about higher level entities are less important as long as they do not involve any superlative, non-reducible entities. It would seem that we can only reason bottom up, from lower level entities and descriptions, and cannot invoke higher level entities in our arguments. The argument from below denies that the value of the whole can be greater than the sum of its parts. If the lower level entities do not carry value, then the higher level entities cannot either. The composition of these entities cannot 'create' value. Johnston provides a reductio against this argument. Together with physicalism the argument from below implies that the only thing that could matter in our world would be microphysical particles. But evidently these are not, by themselves, of any value. Johnston points out how this is not a proof of moral nihilism but rather a reductio against Parfit's argument from below (1997: 154-6, 167-8). We can add that under a dualist view, the argument from below would only count mental events or experiences as having importance to us. But very few of us think that the only thing that has value to us are mental states. To make Johnston's point clearer, we can give examples where Parfit's reductionist deconstruction seems implausible. We can be reductionist about art and say that the Mona Lisa just consists in a poplar panel and various coloured pigments bound together by oil. Presented this way, it is hard to see why we should attach any significance to the Mona Lisa at all.

Parfit replies to Johnston's criticism with some examples of his own (1995: 29-31). In Parfit's examples the reductionist strategy seems more plausible. One example is related to

[11] Parfit makes this argument in *Reasons and Persons* (cf. 1984: 340-1), later however Parfit writes that it was misleading to claim that a person's life could be re-described impersonally. Nonetheless, he insists that an impersonal conceptual scheme would be neither scientifically nor metaphysically worse than our current conceptual scheme (1999).

the definition of death. Plausibly we are reductionist about death insofar as death just means the cessation of certain functions necessary for our continued existence. According to one view it matters how we define and use the word 'death'. But we may plausibly think that what should matter to us morally speaking is which morally important functions cease to exist rather than whether a specific definition is met. Being alive is important only insofar the functions that constitute 'being alive' are important or valuable. We then need a way to distinguish Parfit's more plausible examples, like the definition of death, from other examples, like my own about the Mona Lisa. In other words, we need to show that reductionism about death and persons is a different sort of reductionism from the one involved in art.

I have already alluded to one possible answer, the one Parfit wants to defend. In the case of defining death (and personal identity Parfit may add), we are dealing with merely verbal disputes. In the case of art, on the other hand, this is not the case. Parfit writes:

> 'When I claim that personal identity just consists in certain other facts, I have in mind a closer and partly conceptual relation. … But, if we knew the facts about these [psychological] continuities, and understood the concept of a person, we would thereby know, or would be able to work out, the facts about persons. Hence my claim that, if we know the other facts, questions about personal identity should be taken to be questions, not about reality, but only about our language' (1995: 33).[12]

The most straightforward way to interpret this response is to understand it as analytical reductionism. Analytical reductionism would mean that we could reduce in principle statements involving persons to statements that do not involve persons just in virtue of the meaning of the word 'person'. This form of reductionism seems plausible in Parfit's cases that concern the definitions of words. Analytical reductionism would, however, also mean that the statement about persons and the impersonal statement to which it can be reduced necessarily have the same truth-value. If the difference is merely about our

[12] Elsewhere Parfit writes that under his view the existence of persons is only 'a fact of grammar' (1973: 158), he also writes that most facts about persons only exist because of *the way we talk* (1984: 223, 226, 341). Parfit also defends more explicitly the view that an impersonal conceptual scheme would be no worse than our current conceptual scheme (1999).

language and not about reality, then the relation of equivalence between a statement about persons and an impersonal statement should hold necessarily. But here Parfit provides the best counterexamples against himself. Reductionism about persons does not hold necessarily, non-reductionism may well have been true. If we had evidence of persons remembering events from distant times and were these events confirmed, this would support the case for an immortal soul that can be reincarnated (cf. 1984: 227-8). Reductionism does then not hold as an analytical necessity.

Rather than analytical reductionism, Parfit ought to hold that reductionism about persons is ontological reductionism. Here the idea is that we can translate facts about a specific entity into facts that do not presuppose this entity. Instead of persons we can talk of mental and physical events and their relations. Instead of the Mona Lisa we can talk of colour pigments and their spatial relations, and so on. One specific form of ontological reductionism is constitutive reductionism. Under constitutive reductionism some entities constitute others. A common example for constitutive reductionism are clay statues. The statue does not exist independently from the lump of clay, but neither is it identical to it. Rather the lump of clay constitutes the statue. In cases of constitutive reductionism, we would still say that there is an additional entity in the world. The statue does exist in the world and has an existence separate but not independent from the lump of clay. The existence of the statue will always be parasitic on the existence of the clay. But we can destroy the statue without destroying the lump of clay. This gives us a strong sense how the statue is a separate entity. Facts that hold about the statue are therefore not merely conceptual facts about how to use words, they are facts about a really existing entity. Parfit claims to be a realist about importance by which he means that he attaches importance only to those facts that are ontologically real. But if Parfit is a realist about importance in this way, then he should attach significance to constituted entities. Constituted entities are ontologically real after all. Given that he does not attach significance to persons, his reductionism is most plausibly not a constitutive one.[13]

[13] Here things are getting confused since Parfit does expressly claim to be a constitutive reductionist about persons (1995: 16-7, 1999: 218). However, he describes facts about persons as merely conceptual facts. It might be that I have overlooked something in my argument and that some forms of constitutive reductionism give rise to genuine entities with facts about reality (like statues, art works

Instead of constitutive reductionism, Parfit needs to invoke eliminativist reductionism. According to eliminativist reductionism, the reduced entity does not really exist. It is not part of one's privileged ontology. Instead, we only have terms of convenience that do not refer to any real entity at all. This interpretation gives a strong sense that we would be treating language as more important than reality if we attached significance to persons. The problem with this reading is that Parfit does not give any argument for eliminative reductionism about persons. His reductionist arguments seek to establish that it is possible to give an impersonal description of one's life and that no appeal to a higher entity is needed. But of course the ability to use a different vocabulary does not establish the need to use it. We can similarly give a description of an artwork without mentioning its existence, but this does not mean that we should not include the artwork in our ontology.

We need a different way to distinguish between cases where reductionism does disenchant our ordinary concepts and those cases where it does not. One possible explanation is that in some cases the relations between constituent entities have significance over and above the entities while in other cases they do not. In the example of reductionism about art, it is the special way how the colour pigments of the Mona Lisa stand to one another that makes the Mona Lisa important over and above its individual elements. If we could show that the relations of individual events are not significant in the case of persons, then we would have achieved a reductionist debunking of our concept of a person.

John Broome provides such an argument (1991: 87-90). Broome wants to argue that the relations between the different stages of a person are axiologically insignificant. There would be just as much value in the world regardless of the relation between person stages. Broome draws a comparison between a world with one person and a world with two persons that correspond to the two halves of the first person's life. We can imagine that the two persons are living two different causally isolated lives that correspond to all of the person stages that form part of the first half or the second half of the first person's life

---

and so on) while other forms of constitutive reductionism give rise to merely conceptual facts. Parfit in any case fails to make this argument and I do not know of any good argument to this effect. See also David Shoemaker's post (2006) and ensuing discussion on the PEA Soup blog for more detail on this discussion.

respectively. In this situation Broome says that it is unclear why the world with one person is any different in terms of value from the world with two persons.

Broome's argument should equally hold if we decompose the person's life further into different time-slices rather than comparably big units. If we decompose a life to this extent however it is difficult to see how the importance of fulfilling desires or achieving projects can be captured. Time-slice persons are not extended beyond an ephemeral moment. The satisfaction of desires and the accomplishment of projects however extends in time. Parfit's proposal of the success theory of well-being makes this particularly clear. A success requires some extension in time and does not refer merely to someone being a specific way at a given time (1984: 494-9). On most accounts of a person's well-being, we consider projects or desires to be an important component. Since projects and desires require continued existence over at least some time, the relation between the individual constitutive parts of a person's life does have axiological significance, contrary to what Broome argues.

We should conclude that Broome's argument fails as well. Neither Parfit's claims about 'less deep' truths of personal identity, nor Parfit's appeal to 'merely conceptual truth', nor Broome's argument about the axiological insignificance of the relations that unite a life have succeeded. None of the three arguments has established sufficient ground to reject the importance of persons based on a reductionist metaphysics of persons.

V. Conclusion

I have argued that Parfit's step from the questions of personal identity and of what matters for self-interest to the question of what matters for our moral theorising is not warranted. We can grant Parfit's answer to the question of what matters without having to adjust our moral theories. There is no need to engage in the complex discussion over whether or not identity does or does not matter, if we simply want to defend a person-based form of morality. I have defended the unity of the individual and the separateness of persons against Parfit's challenge. Once we see that psychological continuity and not connectedness is the primary part of Relation R, his challenge fails. And once we understand

that Parfit's reductionism about persons is best understood as a constitutive reductionism, we realize that the unity of the individual and separateness of persons have the same significance as they had before Parfit's challenge.

Parfit's contributions to the metaphysics of personal identity and its implications for rationality and self-interest are truly outstanding. For a while I feared that Parfit provided a strong challenge to my moral beliefs as well. The arguments in this paper have convinced me otherwise. It seems to me that Parfit's argument for a revisionary understanding of morality fails. I can be reassured. It does not matter what matters.[14]

References:

Adams, R. M. (1989) 'Should Ethics be More Impersonal? a Critical Note of Derek Parfit, Reasons and Persons', *Philosophical Review*, 98/4: 439-84.

Blackburn, S. (1997) 'Has Kant Refuted Parfit?' in J. Dancy (ed.) *Reading Parfit*, 180-201. Oxford: Blackwell.

Brink, D. O. (1997a) 'Rational Egoism and the Separateness of Persons', in J. Dancy (ed.) *Reading Parfit*, 96-134. Oxford: Blackwell.

--- (1997b) 'Self-Love and Altruism', *Social Philosophy and Policy*, 14/1: 122-57.

Broome, J. (1991) 'Utilitarian metaphysics?', in J. Elster and J. E. Roemer (eds.) *Interpersonal Comparisons of Well-Being*, 70-97. Cambridge: Cambridge University Press.

Christie, T. W. (2009) 'Natural Separateness: Why Parfit's Reductionist Account of Persons Fails to Support Consequentialism', *Journal of Moral Philosophy*, 6/2: 178-95.

Davidson, D. (2001) *Essays on Actions and Events*. Oxford: Oxford University Press.

[14] Versions of this paper were presented at the Lake Geneva Graduate Conference at the University of Fribourg and at Understanding Value at the University of Sheffield. I would like to thank the audiences for valuable feedback. I would also like to thank David Coombs, Peter Dennis, Todd Karhu, Michael Otsuka, Thulasi K. Raj, and Joe Roussos for comments, discussion, and feedback.

Gauthier, D. (1963) *Practical Reasoning*. Oxford: Clarendon Press.

Johnston, M. (1992) 'Reasons and Reductionism', *Philosophical Review*, 101/3: 589-618.

--- (1997) 'Human Concerns without Superlative Selves', in J. Dancy (ed.) *Reading Parfit*, 149-79. Oxford: Blackwell.

Korsgaard, C. M. (1989) 'Personal Identity and the Unity of Agency: A Kantian Response to Parfit', *Philosophy & Public Affairs*, 18/2: 101-32.

Martin, C. B. and M. Deutscher (1966) 'Remembering', *Philosophical Review*, 75/2: 161-96.

McMahan, J. (2002) *The Ethics of Killing*. Oxford: Oxford University Press.

Nagel, T. (1970) *The Possibility of Altruism*. Princeton: Princeton University Press.

--- (1979) *Mortal Questions*. Cambridge: Cambridge University Press.

Nozick, R. (1974) *Anarchy, State, and Utopia*. New York: Basic Books.

Parfit, D. (1971) 'On "The Importance of Self-Identity"', *Journal of Philosophy*, 68/20: 683-90.

--- (1973) 'Later selves and moral principles', in A. Montefiore (ed.): *Philosophy and Personal Relations. An Anglo-French Study*, 137-69. London: Routledge.

--- (1984) *Reasons and Persons*. Oxford: Clarendon Press.

--- (1995) 'The Unimportance of Identity', in H. Harris (ed.): *Identity*, 13-45. Oxford: Oxford University Press.

--- (1999) 'Experiences, Subjects, and Conceptual Schemes', *Philosophical Topics*, 26/1-2: 217-70.

Rawls, J. (1999) *A Theory of Justice*, rev. edn. Oxford: Oxford University Press.

Shoemaker, D. W. (1996) 'Theoretical Persons and Practical Agents', *Philosophy & Public Affairs*, 25/4: 318-32.

--- (1999) 'Selves and Moral Units', *Pacific Philosophical Quarterly*, 80/4: 391-419.

--- (2006) *Parfit's "Argument from Below" vs. Johnston's "Argument from Above* (PEA Soup Blog) <http://peasoup.typepad.com/peasoup/2006/04/parfits_argumen.html> accessed 16 August 2019.

Shoemaker, S. (1970) 'Persons and Their Pasts', *American Philosophical Quarterly*, 7/4: 269-85.

Sidelle, A. (2011) 'Parfit on 'the Normal/a Reliable/any Cause' of Relation R', *Mind*, 120/479: 735-60.

Sosa, E. (1990) 'Surviving Matters', *Noûs*, 24/2: 297-322.

Unger, P. (1990) *Identity, Consciousness and Value*. New York: Oxford University Press.

Velleman, J. D. (1996) 'Self to Self', *Philosophical Review*, 105/1: 39-76.

Whiting, J. (1986) 'Friends and Future Selves', *Philosophical Review*, 95/4: 547-80.

Williams, B. (1972) *Problems of the Self*. Cambridge: Cambridge University Press.

Wolf, S. (1986) 'Self-Interest and Interest in Selves', *Ethics*, 96/4: 704-20.