12. Agency and Reductionism about the Self

MARKO JURJAKO

1. Introduction

When thinking about the identity of the self, we are usually thinking about issues related to the problems of personal identity. Eric Olson (2002) distinguishes between several questions that relate to the problem of personal identity. For instance, one of the questions concerns the problem of identification and can be expressed as: who are we? Here an answer could be that one part of my identity is that I am politically a leftist and I identify myself with a group of people who hold liberal political views.

Other important questions related to personal identity include problems of defining the persistence conditions of a person or a self and of deciding what it is that we are or what we are identical with. The former refers to the conditions that, for example, determine when one person at time t is identical with a person at some time before t. The latter problem pertains to the issue what we are made of. For example, some people argue that we are essentially human animals or organisms. (DeGrazia 2005, Olson 1997, Snowdon 2014) Others argue that we are material beings that are constituted by an organism, but are not identical with it, rather our identity is determined by appropriate psychological connections. (Baker 2000, Johnston 1987) Some authors argue that we should be identified with parts of our bodies, such as our brain (McMahan 2002), while still others argue that we are immaterial souls (Swinburne 1984).

Besides the metaphysical aspects, problems of personal identity are important for us because of their practical implications. (Shoemaker 2007) For example, if a person *X* stole something from a person *Y*, then it is important for us to know that a person *Z* at time *t* is the same person as *X* at the time of the stealing, so that we can inflict on *X* a just punishment. Similarly, it seems especially important for anyone to know whether she will be the person who will experience pain tomorrow when visiting a dentist or whether somebody else will experience similar pain.

The contemporary discussion on personal identity seems to be focusing on two broad issues. (Bělohrad 2014a, Schechtman 2014) On the one hand, philosophers try to give an account of personal identity that can vindicate

our practical concerns. These investigations, thus, concern issues that are related to responsibility, blame, prudential concern, moral rights, and so forth. (Shoemaker 2007) On the other hand, some philosophers hold that strictly metaphysical issues, such as the problem of persistence conditions, should be dealt with in isolation from its possible relevance to practical concerns. (see, e.g., Olson 1997: 42, 69)

In this article, without attempting to resolve these perennial issues, I will discuss the relevance of agency for personal identity. Although this article has an introductory form, I offer an opinionated overview of the psychological approach to personal identity, most famously expounded by Derek Parfit (1984, 1995), and the role that agency might play in it.

In the next section, I will provide the background relevant for discerning the importance of agency for personal identity. This will involve introducing the psychologically based criteria of persistence conditions. In the third section, I will introduce and discuss Parfit's Reductionist View of personal identity. In the following section, I will discuss the implications of the Reductionist View for our practical concerns. The last section discusses an agency-based view of personal identity and its prospects for vindicating practical concerns that we relate to the notion of personal identity.

2. Psychological Accounts of Personal Identity and What Matters

There seems to be two dominant positions on the persistence conditions of a person. (Schechtman 2005: 1) One involves the biological view of personal identity that states that persons are essentially human animals or organisms. According to this view, a person at t is the same being at t_1 *iff* the being at t_1 and the person at t are biologically continuous. On this view, a person cannot survive the death of her biological body. According to the so-called neo-Lockean psychological continuity theories, persons are not identical to the biological organisms that may or may not constitute them. Rather, the identity of a person is determined by the psychological relations that comprise one's mental life. On this view, a person could survive the death of her biological body, by, for instance, transferring her memories and experiences into some other functioning body.

If agency has any role in thinking about personal identity, it is likely that this role will be more prominent in psychological accounts of the personal identity. This connection has an intuitive support. To be an agent is to have goals and beliefs about how to satisfy those goals. Furthermore, to be a human agent is to be active with respect to those mental states, to evaluate them, make decisions, and act upon those decisions. Hence, it seems that

agency essentially involves certain psychological features, which enable one to be active with respect to oneself and the world.

In addition, psychological accounts are usually supported by noting their connections with practical concerns that often motivate our interest in problems of personal identity in the first place. (cf. Schechtman 2005) For example, if we imagine that my consciousness and experience will tomorrow be transferred to some other body that will be tortured, then it seems I should be scared and anxious about this future event. This seems to be the case because our intuitions seem to follow the psychological criteria of persistence, and not the fact that we are constituted by some particular organism. (For an alternative view, see Shoemaker 2016, Williams 1970.) Similarly, to use Locke's classical example, if a cobbler and a prince switch bodies, so that prince's consciousness is transferred to the cobbler's body and *vice versa*, it seems that our intuitive response is that now the prince inhabits the cobbler's body and *vice versa*. And for whatever actions we used to hold the prince responsible now we will hold responsible the person who currently inhabits the cobbler's body. (Locke 1690/1998, II.15: 440-441)

In what follows, I will introduce the basic features of the psychological accounts and on this background, in the following sections, explain how agency might be relevant for our conceptions of personal identity.

Contemporary psychological accounts of personal identity draw from ideas developed by John Locke. (1690/1998, §27) Locke famously argued that the idea of a man has different identity conditions from the idea of a person.[1] This can be seen by checking our intuitive responses to hypothetical cases. For instance, let us imagine that some criminal mastermind, doctor *X*, has the technology to change bodies so that he escapes capture after he commits a crime. After performing some evil deed, doctor *X* captures an innocent person *Y* and by using his technology switches their bodies. After every switch, the criminal kills the innocent victim who is now in his former body. The question is, who would we hold responsible for doctor *X*'s misdeeds and would we want to punish the person who is now in *Y*'s body? It seems that we would. After all, the person in *Y*'s body is someone who has consciousness of doctor *X*, identifies with his history and mental states, and continues to carry out his plans. Thus, given our practical concerns we can discern what type of identity our judgments about persons follow and see that what defines personal identity is to a significant extent shaped by our value judgments.

[1] In the contemporary terminology, instead of using the word "idea" we would be talkk ing about the concept of a person or a man.

Derek Parfit, in his seminal book *Reasons and Persons* (1984), developed a psychological account of personal identity and argued that what actually matters for our practical concerns is not personal identity but something close to it. To understand his argument, we first need to lay down the basic structure of the psychological view of personal identity.

Parfit develops a neo-Lockean account of personal identity. According to him, the psychological criterion of personal identity over time includes the following:

> (1) There is psychological continuity if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if
> (2) X is psychologically continuous with Y,
> (3) this continuity has the right kind of cause, and
> (4) there does not exist a different person who is also psychologically continuous with Y.
> (5) Personal identity over time just consists in the holding of facts like (2) to (4). (Parfit 1984: 207)

To understand what this account amounts to, we need to explain points (1)-(5).

Points (1) and (2) can be explained together. Parfit defines psychological continuity in terms of overlapping chains of strong connectedness. He defines psychological connectedness as "the holding of particular direct psychological connections." (Parfit 1984: 206) For example, one of the important psychological relations for personal identity is the continuity of memory. Intuitively, we think that to be the same person from t to t_1 is to be aware or conscious of yourself as the same person. One of the capacities that enables us to be self-aware is memory. However, we can forget things, even important things about ourselves, without ceasing to be the same person. For instance, I may forget what I was doing and experiencing on my third birthday. Nevertheless, it is plausible to think that I am the same person as my three-year-old self.

In order to remedy this problem for the memory criterion, Parfit distinguishes between *direct memory connections* and *the continuity of memory.* A person has direct memory connections with her past self *iff* she can now remember that she had particular experiences in the past. For instance, I remember that this morning I was the person who drank coffee in my kitchen, thus I have a direct memory connection with that person. Furthermore, a person does not have to have direct memory connections with somebody for her to be that person. For *X* to be the same person as *Y* there needs to be *continuity of memory* between X and *Y*. According to Parfit, continuity of memory consists in having an overlapping chain of direct

memories. (Parfit 1984: 205) *X* does not have to directly remember the experiences *Y* had twenty years ago in order for *X* to be *Y*. It is enough that there is an overlapping chain of direct memories.

Parfit generalizes the concepts of direct connections and continuity from the memory condition and introduces other psychological relations that might be relevant for determining the persistence conditions of a person. Importantly, Parfit, at least implicitly, recognizes some basic contours of *agency* as being relevant for personal identity. For instance, he mentions beliefs and desires that one can have over time. He also mentions intentions and later acts that serve as executions of those intentions. (Parfit 1984: 205) Intentions are very important for psychological continuity because their typical structure involves diachronic aspects.[2] At time *t* we may adopt an intention, which structures our plans, beliefs, desires, and characters, and thus governs our behavior until the execution of an action at some later time *t1*.

Including agency as one component of Parfit's psychological criterion might be controversial. Most authors do not regard Parfit as including agency into his psychological criterion of personal identity. For instance, Korsgaard seems to criticize Parfit for not giving an appropriate role to agency in his arguments. (Korsgaard 1989) This might be because Parfit does not really put emphasis on agency. In addition, he discusses the issue of personal identity by putting emphasis on the third personal point of view of agents and their mental states, which might seem to downplay the importance of the active role these mental states might play in personal identity. Nevertheless, as already noted, Parfit's view seems to encompass the agential elements as well. To see this, consider the standard causal theory of action. According to this view

> the agent performs an action only if an appropriate internal state of the agent causes a particular result in a certain way. (…) You turned on the light only if the light came on as a result of some neural and/or mental state you were in. (…) Your action was intentional only if the initiating cause was the desire or intention to turn on the light. If you turned on the light unintentionally, then the light came on because you wanted to do something else instead, such as turn on the fan. So the causal theory says that whether an action was intentional depends on whether it was caused by a particular internal state, a desire or intention to perform that action. (Davis 2010: 32-33, italics in the original)

Thus, according to the standard theory of action, it is necessary to have specific mental states, such as intentions that cause actions in the appropri-

[2] For a discussion of the diachronic aspects of intentions and their relevance for personal identity, see (Bratman 2007, essay 2).

ate way, for an action to count as intentional.[3] Different types of intentional actions might require sufficient conditions that go beyond the causal conditions. For instance, to capture full-blown autonomous agency, we might need to refer to additional higher order mental states, such as second order desires or planning attitudes that regulate first order attitudes. (see Davis 2010: 34) Given the standard theory of action, Parfit's psychological criterion seems to satisfy the necessary condition for including agency into considerations that play a role in determining personal identity. In addition, it seems plausible that Parfit's account, with some caveats, could be extended to include also more sophisticated forms of agency. I discuss this option in section 5.

To recap, psychological continuity consists in these overlapping chains of strongly direct psychological connections, including agential components – which Parfit calls strong psychological connectedness.

According to Parfit (1984: 206), psychological connectedness comes in degrees. For instance, *X* and *Y* may have many thousands of psychological connections or even just a single connection. *X* may share with *Y* all beliefs, desires, and memories and the same intentions and long term plans. Plausibly, in this case *X* would be the same person as *Y*. However, if *Y* had an intention to do something and *X* remembers having the same intention as *Y*, and that is the only psychological similarity between *X* and *Y*, then presumably they are different people. Thus, Parfit maintains that strong psychological connectedness involves having *enough* of direct psychological connections.

> For X and Y to be the same person, there must be over every day enough direct psychological connections. Since connectedness is a matter of degree, we cannot plausibly define precisely what counts as enough. But we can claim that there is enough connectedness if the number of connections, over any day, is at least half the number of direct connections that hold, over every day, in the lives of nearly every actual person. (Parfit 1984: 206)

This inability to provide a non-arbitrary criterion for strong psychological connectedness will have important consequences for Parfit's claim that personal identity is not what practically matters to us. But before we get to that, we need to further explain Parfit's account of personal identity.

Point (3) states that the psychological continuity needs to be caused in the right way. Causal condition in psychological accounts is usually introduced to handle some problematic cases. (cf. Shoemaker 1984) For instance, one of the earliest objections against using memory as a condition for personal identity is the problem of circularity. For a memory to be part

[3] The casual theory of action is not without its critics. For a discussion of some of the objections and a defense of the causal theory of action, see Schlosser (2011).

of the ground of one's persistence through time, the memory needs to be of something that really happened to the person. If I only have memories of events that did not happen, then those memories cannot constitute *my* persistence conditions. Since those things did not happen, they cannot constitute me as the person that I am. Thus, any account of personal identity that uses a memory criterion needs to be able to distinguish between real memories and apparent memories or delusions. However, this seems to involve a circularity. Schechtman gives a succinct exposition of the problem:

> the difference is just that genuine memories are of an experience the rememberer actually had, while delusions are apparent memories of an experience that was not had by the person seeming to remember. Since the memory criterion must define identity in terms of real memories, and real memories are defined in terms of personal identity, the criterion ultimately defines identity in terms of itself. (Schechtman 2014: 22)

It seems, thus, that to determine whether a memory is veridical, we need to know whether the event in question happened to the person. But if this is the case, then we already need to know the identity of the person in order to determine whether she had the relevant experience or not.

The standard solution to this problem involves replacing the concept of memory with a more general concept of quasi-memory. (Olson 2002, §4) Quasi-memory is a memory type experience that is caused in a right way and does not presuppose personal identity. Parfit defines the notion of having an accurate quasi-memory as follows:

> A. I seem to remember having an experience,
> B. someone did have this experience, and
> C. my apparent memory is causally dependent, in the right kind of way, on that past experience. (Parfit 1984: 220)

From having a quasi-memory of cooking lunch today, I cannot conclude that it was *me* who cooked the lunch, rather I can conclude that *somebody* cooked the lunch today. (For further discussion of quasy-memory, see Roache 2006.)

In principle, I can have quasi-memories of somebody else's experience. For instance, we can imagine that a genius neuroscientist can reconfigure my brain states in such a way that they become isomorphic to the relevant brain states of my wife. Now I seem to remember going on a business trip to Brazil and having an awful time there, when in fact it was my wife who went to Brazil and had unpleasant experiences. This possibility aside, quasi-memory, with its causal condition enables us to avoid the circularity objection. The important thing for psychological accounts is that ordinary memory can be seen as a subset of quasi-memories, namely as "quasi-memories of our own past experiences." (Parfit 1984: 220)

It has to be remembered that in psychological accounts of personal identity quasi-memory only partly constitutes personal identity. Other criteria include the continuity of beliefs, desires, values, intentions, etc., which also should be caused in the right way. In particular, Parfit (1984: 207) includes aspects of agency as important elements that need to be properly causally integrated. At one point, he mentions that changes in character need to be caused in the right way if they are going to count as relevant for determining one's personal identity. (Parfit 1984: 207) Continuity of character is preserved if the changes in the character are caused by deliberate decisions, growing older or responses to particular life experiences. However, the continuity of character is hampered if the changes are "produced by abnormal interference, such as direct tampering with the brain." (Parfit 1984: 207) These remarks are congenial to Christine Korsgaard's idea that the right kind of cause is the one that stems from the decisions and commitments endorsed by an agent. (Korsgaard 1989) According to Korsgaard, the psychological connections and their continuity are identity preserving if they are based on *authorial* causes, that is, causes that are based on agent's deliberations and decisions.

Parfit (1984: 207-209) offers different views on how (3) could be understood. On the narrow reading, the *right* kind of cause is the *normal* cause. The normal cause of our psychological continuity could be the brain that implements those connections. In this case, the sameness of the brain would guarantee that the psychological continuity is caused in the right way.

In addition, Parfit (1984: 207-208) distinguishes between the wide and the widest reading of (3). According to the former, any *reliable* cause of psychological continuity is sufficient to be the right cause. On the latter reading, *any* cause can be the right kind of cause for psychological continuity. He does not say much about the difference between the wide and the widest reading of the "right cause," nevertheless he argues that they have better theoretical standing than the narrow reading.

This can be shown by an analogy. Suppose that our brain cells are degenerating and that we have the technology to replace them with some synthetic material. Once all of the organic brain cells die, we still have a functioning brain that causes the psychological continuity since now the synthetic material has taken over the function of the brain cells. If the normal cause were needed for psychological continuity, then in this case after the replacement of brain cells, we would not have the same person, since supposedly this synthetic material is not what *normally* causes the psychological continuity in a person. Based on a similar case, Parfit (1984:

208-209) contends that it is better for a psychological account of personal identity to allow for a wider construal of the notion of the *right cause*.[4]

Clause (4) states that uniqueness is necessary for identity through time. Consider an example. In the so-called simple teletransportation case, we are imagining a person who decides to travel to Mars by means of teletransportation. This is a type of teleportation familiar from TV shows, such as *Star Trek*. The important difference is that in the simple teletransportation case, when a person enters the machine, her brain gets scanned and the psychological states are copied while her body is destroyed on Earth and rebuilt from organic materials on Mars. Also, the important things that get copied are the brain patterns and psychological connections. Intuitively, it seems that if X decides to travel by teletransportation then the person that is psychologically continuous with *X* will really be *X* and not just her replica.

Now consider a case of teletransportation in which *X* enters the machine, his brain gets scanned and the psychological information is sent to Mars where a new body that is psychologically continuous with *X* is formed. But in this case the machine does not destroy *X*'s body on Earth. It seems obvious that the person on Mars is not identical with *X*, even though she is a complete psychological replica of *X*. It is unclear what this imaginary case shows. It could be taken to show that not even in the simple teletransportation the person on Mars is identical with *X*. On the other hand, if we follow the intuition supporting the psychological accounts, we can say that in the first case, X is the same person who travels from Earth to Mars *via* teletransportation, while in the second, this is not the case. What explains the difference is the fact that the uniqueness condition is not satisfied in the second case, while it is in the first one.

Clause (5) expresses what Parfit calls reductionism about personal identity. According to him, the only plausible account of personal identity, whether of physical or psychological kind, entails reductionism, in the sense that what personal identity consists in is exhausted in conditions (1)-(4). Since Parfit grounds his conclusion that personal identity is not what is practically important on the idea of reductionism, in the next section I will explain how reductionism is supposed to support this implication.

[4] Given Parfit's claim that even the widest reading of the notion of *right cause* can be legitimately adopted, he seems to be trivializing the requirement of having the right cause. (Schechtman 2014: 25, fn. 25) Claiming that *any* cause can be the right cause seems to strip the requirement of any importance for determining personal identity. As already indicated, agency-based accounts limit the right causes to the deliberative capacities of an agent, and to capacities that play a role in unifying and integrating different temporal phases of an agent. See also the discussion in section 5 below.

3. Reductionism and the Non-Importance of Personal Identity

According to Parfit, the presented account of identity is reductionist because it claims that the identity of a person consists in obtaining of conditions (1) to (4). In fact, there is no need to invoke any *further* facts, beyond those that underlie psychological continuity that is caused in the right way. According to Parfit, the Reductionist View includes the following two conditions:

a. that the fact of a person's identity over time just consists in the holding of certain more particular facts, and
b. that these facts can be described without either presupposing the identity of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists. These facts can be described in an impersonal way. (Parfit 1984: 210)

In the case of psychological theories of identity, condition (a) states that identity consists in and is exhausted by the psychological continuity that is caused in the right way. Condition (b) states that facts about personal identity can be described in such a way that does not presuppose the identity of a person and by using a language that is impersonal, or we might say, based on a third-person perspective.

Parfit opposes reductionism to what he calls the *Further Fact View* (Parfit 1984: 210) Non-reductionists claim that psychological continuity does not exhaust facts about personal identity. Reductionists deny this, and claim that there is no further fact that determines personal identity. According to Parfit, a good test for whether someone's view of personal identity is reductionist is the following:

> If we accept a Reductionist View, there may be cases where we believe the identity of such a thing to be, in a quite unpuzzling way, indeterminate. We would not believe this if we reject the Reductionist View about this kind of thing. Consider, for example, clubs. Suppose that a certain club exists for several years, holding regular meetings. The meetings then cease. Some years later, some of the members of this club form a club with the same name, and the same rules. We ask: "Have these people reconvened the very same club? Or have they merely started up another club, which is exactly similar?" There might be an answer to this question. The original club might have had a rule explaining how, after such a period of non-existence, it could be reconvened. Or it might have had a rule preventing this. But suppose that there is no such rule, and no legal facts, supporting either answer to our question. And suppose that the people involved, if they asked our question, would not give it an answer. There would then be no answer to our question. The claim "This is the same club" would be neither true nor false. (Parfit 1984: 213)

According to Parfit, in the case of a club, there is a possibility of indeterminacy of identity because the identity of a club does not consist in anything beyond the existence of its members and their behavior and attitudes. Thus, when we know everything about club members and their relations, there is *no further fact* that might determine whether one club is the same club at time *t* and *t1*. If facts about people that comprise a club cannot decide the issue, then the identity of the club is indeterminate.

When the question whether a thing at *t* is the same as the thing at *t1* is indeterminate in the above sense, then we can conventionally decide to say that at *t* and *t1* we have the same thing or we can say that we have a different, albeit, very similar thing. For example, we can either say that the club at t is identical to the club at *t1* or we can say that the two are similar in all respects but are not actually identical. According to Parfit, when the answer is conventional in this sense, then, although the question whether *X* is identical to *Y* is sensible, it is actually *empty*. (Parfit 1984: 213)

Once we accept that "person" does not refer to a separately existing entity and does not involve further facts that go beyond those involved in psychological continuity and/or psychological connectedness and their physical implementation, according to Parfit,[5] we get to two different, albeit, related conclusions. One is the previously mentioned claim that reductionism is committed to the possibility of indeterminism about identity. The second is that identity is not what really matters in survival.

To see why reductionism is committed to indeterminism we have to remember that, according to the psychological criterion, personal identity reduces to a *unique* psychological continuity, which consists in overlapping chains of psychological connections with the right kind of cause. Since there is no non-arbitrary way to say to what degree the overlapping chains of psychological connections need to obtain for numerical identity through time, we can imagine that degrees of connectedness fall on a spectrum. (see Parfit 1995) On the one side of the spectrum there will be a complete psychological continuity between *X* and *Y* from *t* to *t1*. On the other side of the spectrum there will be no psychological continuity between *X* and *Y*. In the first case *X* and *Y* would be the same person, while in the second they would be different persons. Now suppose that there is a

[5] Parfit extensively argues for the Reductionist View and the view according to which a person is not some separately existing entity (Parfit 1984, sec. 88; see, also, Parfit 1995). Without going into details, here I just note that Parfit persuasively argues that the only viable alternative to reductionism is some form of mind/body dualism. As is well known, there are great difficulties for defenders of dualism to explain the identity conditions of a separately existing self, and in what way persons could exist as non-physical substances (see, e.g., Maslin 2001: §1).

neurosurgeon who can change the degree of psychological connectedness (memories, plans, character, etc.) between *X* now and *X* at some later time by changing the neural activation patterns in X's brain. Small changes in psychological connectedness between *X* at *t* and *t1* will not affect his personal identity. However, if little changes are progressively made, at some point the degree of connectedness will fall somewhere in the middle of the spectrum. When that happens, we will not be able to say whether *X* will continue to be the same person or an additional small change will make *X* disappear.

The case is similar to the *Sorites paradox*. Adding one more grain of sand will not make a heap. If we add enough grains to one pile, it will make a heap, but there is no non-arbitrary way of telling when a non-heap becomes a heap. A similar thing seems to follow about personal identity if reductionism is true. It is possible that we know all the identity relevant facts, but still we cannot say whether some person continues to exist or ceases to exist. That is why Parfit says that there are *empty* questions about personal identity. Even if we are looking at the middle of the spectrum we can ask whether *X* still exists or ceases to exist. We can even decide to say that if there is a 60% of overlapping chains of psychological connections between *X* at *t* and *Y* at *t1* then *X* is identical with *Y*. Nevertheless, since, by the supposition of reductionism, we already know everything that is to know about psychological facts and degrees of connectedness, this cut off line would be stipulated and not imposed on the psychological criterion by the facts that determine personal identity. If we care to provide an answer to questions of identity in these borderline cases, the answer could only be given by *fiat* and not determined by metaphysical facts alone.

The second implication of reductionism is that personal identity is not what matters in survival, rather, what is important is psychological continuity and/or connectedness. Parfit refers to psychological continuity and/or connectedness with the right kind of cause as the Relation *R*. (Parfit 1984: 215) To see why according to reductionism Relation *R* is more important than personal identity let us consider another example.[6] Normal human brain has two hemispheres that are connected with the brain area called *corpus callosum*. Let us imagine that both hemispheres could implement in its entirety Relation *R*. In this imaginary case, if *X*'s left hemisphere were to be damaged, the right hemisphere could take over and *X*'s personal identity would be preserved. The same thing would happen if the right hemisphere were to be damaged. Now suppose that *X*'s body and the

[6] We have to bear in mind that according to the psychological criterion personal identity consists in *R* and the uniqueness condition.

right brain hemisphere are destroyed and a doctor decides to transplant his left-brain hemisphere to another person's body that is currently brain dead. In this case, we are inclined to think that *X* would continue to exist in another person's body. Now imagine that for some reason *X*'s both brain hemispheres are not damaged but they are separately transplanted into two different bodies, *Y* and *Z*. Since, by hypothesis, the two brain hemispheres are both fully functional and can fully implement and preserve *X*'s psychological connections, the bodies *Y* and *Z* both have identical *R*'s to *X*. Furthermore, since identity is a transitive relation and *Y* and *Z* are not numerically identical persons, they cannot both be identical to *X*. In fact, according to the Reductionist View, neither is identical to *X* since they both stand in the same relation *R* to *X*, and thus the uniqueness condition is not satisfied.

Nevertheless, it would be inappropriate to say that *X* dies or ceases to exist once his brain hemispheres are divided. As Parfit writes,

> We might say: "You will lose your identity. But there are at least two ways of doing this. Dying is one, dividing is another. To regard these as the same is to confuse two with zero. Double survival is not the same as ordinary survival. But this does not make it death. It is further away from death than ordinary survival." (Parfit 1984: 262)

It seems that this example shows that what really matters in continued existence is not personal identity, rather it is Relation *R*. In cases of division, where Relation *R* takes a branching form, even though we cannot say that literal personal identity is preserved, people do not just cease existing. As Parfit writes, two is different from zero. In ordinary life, Relation *R* tends to coincide with personal identity, but in these extreme cases we see that they can come apart. And when they do, it seems that we are inclined to believe that what really matters to us is that at least one person will bear Relation *R* to us.

To support the latter claim consider the following example. At time *t* I find out that at some future *tn* I will divide into two people who will have identical Relations R to me. In that case, would I stop thinking about the future and decide to live like *tn* will be my last day alive or would I be preparing for the future and devise plans that could be more easily executed by my branching counterparts? For example, I might decide to write a book and to ease the coordination problems, I might indicate which of the two of my continuers should devote his time to that activity. The other one could then devote his time to accomplish some of my other deeply held desires. Even though there might not be anything essentially different between holding an attitude as in the first (where a person thinks his life

is finished) and the second example, it is plausible that we would be more inclined to have the latter attitude towards our knowledge that in the future we will be divided.

4. Problems with the Reductionist View: The Extreme Claim

The importance of agency and related considerations for thinking about personal identity can be discerned once we take into consideration a common objection against the Reductionist View. I will turn to the agency based solution in the next section. In what follows I will present the problem and define the notions that underlie it.

According to what Parfit calls the Extreme Claim, if the Reductionist View is true then it could be argued that we would not have any special reason to care about our own future. We would have equally valid reason to care about ourselves as much as anybody else, we should be indifferent to whether we live or die, etc. (Parfit 1984: 306-307) Many authors find these ideas problematic because they think that many of the practical concerns that we have and care about could not be explained or justified without some reference to the notion of personal identity.

The objection is that if reductionism is true then we could not make sense of our prudential, ethical, and "forensic" practices. Intuitively we think that we bear special relations to our past and future selves. As Korsgaard writes, "I am responsible for my past self, and I bear the guilt for her crimes and the obligations created by her promises. I am responsible to my future self, for whose happiness, since it will one day be mine, it is rational for me to provide." (Korsgaard 1989: 108)

One version of an objection based on the Extreme Claim can be formulated as follows:

I. If reductionism about personal identity is true, then it would not make sense to hold people responsible for their past actions. (One version of the Extreme Claim)

II. Holding people responsible for their past actions is an important part of our social and ethical practices. Thus, it makes sense to hold people responsible for their past actions.

Therefore,

III. Reductionism about personal identity is incompatible with our social and ethical practices.[7] (see Schechtman 2014: 34-35)

[7] Some people would not have problems with accepting the first premise of this argument and therefore with accepting the conclusion. These people would be characterized as revisionist with respect to our common social and ethical practices. Parfit himself seems to be inclined to accept many of the consequences of the Extreme Claim. (see Parfit 1984: ch.14)

The problem with (III), for instance, is that, if true, nobody could deserve to be punished for their past crimes.

Support for (I) comes from the so-called fission cases. Let us suppose that I am a hardened criminal whose two brain hemispheres were for some reason transplanted into two different bodies, *Y* and *Z*. Intuitively, a person should be punished for a crime only if she committed that crime. According to the Reductionist View, there is no non-arbitrary reason to say that I am identical with either *Y* or *Z*. Since *Y* and *Z* are different people, it is plausible to say that I am neither *Y* nor *Z*. Therefore, since there is no *me* after the transplantation, there is no person that should be punished for my crimes. According to this argument, accepting the Reductionist View would commit one to a revisionist perspective on our social and ethical practices.[8]

One could object to the argument that the first premise is not true. Indeed, in the fission case we could not blame *X* for his misdeeds, because after the fission there is no *X*. But fission cases represent only abnormal situations. Normally, we will have a person that does not share with other selves, to a significant degree, Relation *R*. Thus, normally, personal identity would be important for ascribing responsibility and other practical notions. However, this objection would arguably miss the point. The proponent of the argument could say that what it really shows is that when we think about responsibility, for instance, what is really important is Relation *R*. If we are willing to punish *Y* and *Z* for *X*'s crimes, then what really matters is not *X*'s identity, but rather the inheritance of his *R*. Similar considerations could then be applied to normal, non-fission cases.

A more promising objection is that the Reductionist View does not entail some unpalatable version of the Extreme Claim. At least, this is not shown by relying on the aforementioned fission case. In fact, it is plausible that our moral practices only require the continuation of Relation *R*, and not the satisfaction of the uniqueness condition. (Bělohrad 2014a: 316-317; see Parfit 1984: ch.13) In the fission case, it does not offend our intuition to say that *Y* and *Z* should both be punished for *X*'s crimes. After all, they directly inherit what our responsibility judgments seem to track, namely her personality and other components of Relation *R*. Thus, we might say that our practical judgments track, what we might call, *moral selves* and not personal identities *per se*.[9] The purported difference between a moral self

[8] Here the supposition is that even after the fission we would still want to punish *X*'s descendants.

[9] For a similar claim that the relation between responsibility and personal identity should be loosened. (Beck 2015: 315-316) The notion of moral self that I use here is similar to the one Marya Schechtman uses in her book *Staying Alive*. However, it should

and the self of personal identity consists in the fact that while the latter is unique to a person, the former comprises set of mental states, personality traits, dispositions, and a history that, in principle, might be shared by different persons.

The idea of a moral self, I believe, captures the intuitions underlying our practical judgments in the fission cases. For instance, let us imagine that a criminal *X* decides to split her two brain hemispheres and implant them into two different bodies (*Y* and *Z*) in order to increase her chances of escaping the police. Let us say that her reasoning is that if she divides herself into two people who will inherit her mental life, memories, plans, projects, etc., there is a greater chance of one of them escaping. Now, neither *Y* nor *Z* is strictly identical with *X*. They are two different spatiotemporally individuated objects. However, they are the same person as *X* in the sense that they are psychologically identical with *X*. They committed the same crimes, they have the same history, memories, plans, dispositions, etc. In other words, we can say they have the same moral character or the self. Because of that moral self, we feel that *Y* and *Z* are responsible for the crimes *X* committed in the past and they should be punished for them. Thus, this example seems to indicate that our responsibility judgments track what we might call our *moral selves*, and not the strict identity of a person.

However, reductionism about personal identity has other seemingly undesirable consequences. If that which really matters in our practical concerns is our moral self, which is underpinned by Relation *R*, and not our uniqueness (strict personal identity), then it seems we would not have a reason to be especially concerned about ourselves. This point is nicely illustrated by Radim Bělohrad:

> Another practical consequence of reductionism results from the fact that part of relation R, namely connectedness, holds in degrees. That is, I may be more or less connected to my future and past selves. Thus, if R justifies attributions of responsibility, I may be less responsible for the actions of my distant past self than for my yesterday's self. Similarly, if R is what justifies the rationality of my concern for the future inhabitant of my body, when

not be confused with it. Schechtman (2014: 14-15) distinguishes *moral selves* from *forensic units*. She seems to think that our moral practices and practical concerns presuppose the existence of forensic units and not just moral selves. However, the distinction between moral selves and forensic units is subtle. (see, also, Bělohrad 2014b: 566-567) Forensic units are entities that exist in virtue of being the proper targets of our practical concerns (such as, blame, self-concern, etc). Moral self, on the other hand, is comprised of contingent properties that give content to a forensic unit. In other words, having a moral self determines what practical judgments *actually* apply to individual forensic units or persons. Thus, forensic units provide a prerequisite for the existence of moral selves.

> R holds to a low degree, so should my concern. This aspect, in turn, leads to an increase in the plausibility of paternalism, because great imprudence with respect to my distant future self is seen as violating my obligations to others, rather than myself, thus becoming immoral, rather than irrational. (Bělohrad 2014a: 317)

In effect, the unimportance of personal identity leads to the unintuitive view that we do not have rationally binding reasons to be concerned about our distant future selves. Since Relation *R* underpins personal concerns, the more different (in terms of *R*) we are from our future self the less reason we have to be rationally concerned about her. In other words, our distant future selves are to us like any other distant person we may or may not know. However, this seems unintuitive. It seems reasonable to say that we have more reason to care about that person who will inhabit our body in the future than somebody who is physically completely distinct from us. Thus, it is not clear how persuasive this example is.

One could argue that the Reductionist View cannot be true since psychological connections do not support the unity that is presupposed in personal identity. The unit to which we apply practical judgments, such as judgments of responsibility and blame, presuppose that there is a deep unity of *consciousness*. This view has roots in Locke who maintained that personal identity pertains to the unity of consciousness and not to the existence of bodily substances. (Locke 1690/1998: II.9) If the Reductionist View were false, then the Extreme Claim would lose its support. However, Parfit has an argument that pertains to show that personal identity does not presuppose the deep unity of consciousness.

To illustrate this claim, we can once again examine one of Parfit's examples. (see Parfit 1984: 246-247) Let us suppose that my two brain hemispheres possess the same abilities and each can function as a separate unit that can implement conscious experiences and cognitive functions. In addition, suppose that I am able to disconnect the communication between the hemispheres, so that they operate as separate cognitive units. Since both hemispheres can support consciousness, when I disconnect them it is as if I have two minds. This ability would allow me to solve some tasks more effectively. For example, I might be taking a physics exam and not be sure how to answer a question. I see at least two ways in which the question could be answered. Since the time is pressing, I decide to disconnect my brain hemispheres and let each try out one possible solution to the problem. During that time, there is no unity of consciousness, each brain hemisphere is conscious of what it is doing and what it has control over.[10]

[10] For instance, left brain hemisphere controls the right arm, has a visual field of the right eye, etc. The opposite could be true of the right brain hemisphere.

After some time, when the two hemispheres do their parts, I reconnect them, that is, reunite my mind (consciousness), and compare the solutions. Furthermore, after the unification, I remember everything that each hemisphere did separately, and in that sense I am psychologically continuous with both consciousnesses that existed for a while.

This example is supposed to show that "a person's mental history need not be like a canal, with only one channel, but could be like a river, occasionally having separate streams." (Parfit 1984: 247) According to Parfit, although portraying a surprising possibility, the coherence of this example shows that I could be the same person from *t* to *t*n, without exhibiting the unity of consciousness. Thus, we could conclude that the unity of consciousness is not necessary for personal identity.

Other authors have tried to resist the Extreme Claim by giving an account of personal identity that can vindicate our intuitive practical concerns. Most notably, Christine Korsgaard (1989) argued that an agency-based view could vindicate the importance of the deep unity that personal identity presupposes in our practical concerns. She seems to agree that if reductionism about personal identity is true then we are committed to the Extreme Claim.[11] Thus, she develops an agency-based view of personal identity that is supposed to show why and in what way the reductionism about personal identity is not true. In the next section, I will present Christine Korsgaard's (1989) and Michael Bratman's (2007) views on the role of agency in personal identity. I will argue that agency-based accounts are in principle compatible with the Reductionist View. Then I will evaluate the prospects of the agency-based view for solving problems related to the Reductionist View.

5. Reductionism and the Agency-Based View of Personal Identity

In her seminal paper, Korsgaard (1989) argued that Parfit's psychological criterion of personal identity is too theoretical, and when we direct our attention to a practical perspective, we will see that there is stronger unity of the self than Parfit envisioned. I will argue that Korsgaard's agency-based approach can be seen as an extension of Parfit's Reductionist View.

What creates the appearance that Reductionist views cannot account for deeper unities that comprise a person's self is Parfit's theoretical perspective. The most important element of the psychological criterion is the psychological continuity and/or connectedness with the right cause. When

[11] This seems to be the standard understanding of Korsgaard's position (see, e.g., Schechtman 2008: 407-408).

Parfit discusses this criterion, one is left with an impression that in a standard case psychological continuity is a collection of very loose relations. Given that the Reductionist View implies the possibility of indeterminacy of identity, this impression might seem to be justified. In addition, this impression is reinforced by Parfit's claim according to which any cause can count as the right cause for maintaining personal identity (see, above, ft. 4).

However, the possibility of indeterminacy does not imply that in a standard situation a person's identity will be indeterminate. Korsgaard (1989) argues that when we turn from Parfit's theoretical or metaphysical perspective to a more practical or decision-making perspective, we will see that a deeper unity underlies psychological connections that define a person's identity. In fact, Korsgaard, in a Kantian fashion, argues that from the first person perspective, that is, from the perspective of agency we are *forced* to postulate a locus that unifies and sustains, in a right way, all the psychological relations that determine one's personal identity. The relation between the first-person perspective and agency might not be straightforward. I will shortly discuss this issue in order to show in what way they are noncontroversially related.

Lynne R. Baker (2011) provides a simple argument for the view that being an agent involves having a first person perspective on things. Let us recall that on a standard theory of action an action is intentional only if it is produced, in an appropriate way, from beliefs, desires, or intentions. Thus, actions are events that can be explained by invoking the relevant mental states. The fact that these mental states can be used for explaining action indicates that they can be used in some forms of instrumental or practical means-end reasoning processes that lead from those mental states to the performance of the action. Thus, if agents perform their actions in virtue of mental states they possess, and those states may figure in practical reasoning, then an agent is someone who has an *ability* to engage in at least some forms of instrumental or practical means-end reasoning.[12] However, explaining someone's action normally involves seeing things from her own perspective. Donald Davidson has made this point especially salient a long time ago:

> A reason rationalizes an action only if it leads us to see something the agent saw, or thought he saw, in his action-some feature, consequence, or aspect of the action the agent wanted, desired, prized, held dear, thought dutiful, beneficial, obligatory, or agreeable. (Davidson 1967: 685)

[12] According to Baker, these considerations indicate that there is a conceptual connecᴇtion between agency and the ability to engage in primitive forms of instrumental reasoning. (Baker 2011: 3)

When we combine this consideration with the idea that intentional action presupposes having an ability to engage in practical reasoning, we get to the view that "the agent reasons about what to do on the basis of her own first-person point of view. It is the agent's first-person point of view that connects her reasoning to what she actually does." (Baker 2011: 3) Thus, we can conclude that the perspective of practical reasoning involves a first-person perspective. Now we can return to the discussion of Korsgaard's agency-based view of personal identity.

Korsgaard (1989: 109) starts by asking why we have an experience of ourselves as being a unified entity in a particular moment? If we do not believe in a Cartesian Ego, what could explain the experience that we are the person that we are, rather than someone else? Korsgaard suggests that the answer lies in our ability to make decisions and act. When we take a practical point of view, we see that we cannot but to think of ourselves as unified active subjects.

Korsgaard claims that two considerations ground this view. The first "is the raw necessity of eliminating conflict among your various motives." (Korsgaard 1989: 110) The second ground for unity can be discerned from "the unity implicit in the *standpoint* from which you deliberate and choose." (Korsgaard 1989: 111) Let us start by considering the first ground.

There are many moments in life in which we are confronted with different desires, considerations, and options that suggest incompatible courses of action. Since we often successfully act, it means that we often manage to resolve these conflicts. When we resolve them, we act as a unified agent. Korsgaard illustrates this point with the split-brain hemispheres example:

> So imagine that the right and left halves of your brain disagree about what to do. Suppose that they do not try to resolve their differences, but each merely sends motor orders, by way of the nervous system, to your limbs. Since the orders are contradictory, the two halves of your body try to do different things. Unless they can come to an agreement, both hemispheres of your brain are ineffectual. Like parties in Rawls's original position, they must come to a unanimous decision somehow. You are a unified person at any given time because you must act, and you have only one body with which to act. (Korsgaard 1989: 110-111)

The first source of unity comes from the *necessity* to act, which includes resolving conflicts and making decisions as a unified subject. Korsgaard mentions that we have one body, which might be taken as implying that this very fact contributes to the unity of a person. However, having one body is not essential for Korsgaard to make her point. We can confirm this contention by imagining a situation in which we can control two bodies in the same way in which we control two hands. So the pressure to act as a unified agent does not originate from the constraint of a particular body,

rather it seems to be consistent with the idea of agency that is distributed across different bodies.

The second ground involves thinking about the first-person perspective from which we deliberate and make decisions. Korsgaard claims that when we reflect on the deliberative standpoint we are compelled to think of ourselves as unified subjects.

> It may be that what actually happens when you make a choice is that the strongest of your conflicting desires wins. But that is not the way you think of it when you deliberate. When you deliberate, it is as if there were something over and above all your desires, something that is you, and that chooses which one to act on. The idea that you choose among your conflicting desires, rather than just waiting to see which one wins, suggests that you have reasons for or against acting on them. And it is these reasons, rather than the desires themselves, which are expressive of your will. The strength of a desire may be counted by you as a reason for acting on it; but this is different from its simply winning. This means that there is some principle or way of choosing that you regard as expressive of yourself, and that provides reasons that regulate your choices among your desires. (…) This does not require that your agency be located in a separately existing entity or involve a deep metaphysical fact. Instead, it is a practical necessity imposed upon you by the nature of the deliberative standpoint. (Korsgaard 1989: 111)

Here Korsgaard points out that our natures as beings who are faced with choices and can reflect on the mental attitudes and situations we find ourselves in, practically force us to conceptualize ourselves as *unified* sources of agency.

Korsgaard emphasizes the distinction between the theoretical or metaphysical and practical or agential (and thus the first-person) perspective, in a way that might be interpreted as being incompatible with the Reductionist View.[13] For example, the insistence that the unity of a person can be discerned only from an agential point of view, which is naturally interpreted as the first-person point of view, might be taken to negate the Reductionist thesis that personal identity can be determined from an impersonal or third-person point of view (see Korsgaard 1989: 193).[14] Without getting

[13] Schechtman (2008) seems to endorse this reading of Korsgaard.

[14] One way in which Korsgaard is opposing Parfit's reductionism is by claiming that a person should be compared to a state and not to a nation or a club. (Korsgaard 1989: 114-115) While nations are just mereological sums of the people that live in a certain territory, states are more than that. On Korsgaard's view, a state "is a moral or formal entity, defined by its constitution and deliberative procedures." (Korsgaard 1989: 114) However, even if the notion of state provides a better analogy, this does not by itself provide a reason to think that reductionism about personal identity is false. To recall, reductionism includes the claim that personal identity consists in holding of some more particular facts (such as, the obtaining of some set of psychological connections) and the possibility of theorizing about personal identity from an impersonal or

into the exegesis of Korsgaard, I will argue that an agency-based view of personal identity might be construed in a way that is compatible with the Reductionist View.

As Korsgaard points out, emphasizing the practical perspective does not involve commitment to some additional metaphysical fact. Thus, taken at face value, it does not contradict the Reductionist View of personal identity. It merely involves a change of perspective, which involves paying closer attention to those aspects that make us agents.

Parfit's discussion of overlapping chains of psychological connections suggests an overly modular view of a person. (see also Bělohrad 2014a: 320-321) However, nothing in his account commits us to this view. Once we direct attention to the structure and organization of psychological connections, a different picture of the Reductionist View will emerge. In fact, when arguing against the narrow construal of "the right kind of cause" Parfit himself recognizes that what is important about psychological continuity is its organization. That is why Korsgaard construes his view as claiming that "persisting identity is simply formal continuity plus uniqueness." (Korsgaard 1989: 106-107)[15] However, I think that already in this formulation there is a gesture towards a psychological criterion that can account for deeper units that underlie personal identity.

David Shoemaker (1996) has argued that Parfit's psychological criterion of personal identity and Korsgaard's agency view of the self are not incompatible, at worst they provide complementary pictures of a person's identity. I concur with this view. What is hidden in Parfit's discussion of the psychological criterion, however, is the emphasis on the structure or the organizational elements that keep together and unify all these different overlapping chains of psychological connections that underpin psychological continuity.

The latter claim can be explained by an analogy with a solution to a general problem from the philosophy of action. Mariam Thalos (2007: 127) indicates that in the Davidsonian tradition in philosophy of action, the relation between mental motivation and action was conceived in quasi-Newtonian terms. The idea is that desires, as paradigmatic pro-attitudes (plus instrumental beliefs), cause action as a function of their strength.

third-person point of view. On the face of it, the idea that persons are more like states than nations does not violate the two conditions. Or if it does, this is not obvious and some argument should be provided for that claim.

[15] Korsgaard uses the term "formal" in the Aristotelian sense, where it is contrasted with material. In her words, psychological properties relevant for identity are determined by "the way the matter is organized, not in the particular matter used." (Korsgaard 1989: 106)

From a third-person perspective, this picture looks very passive. It portrays intentional action as being a linear sequence of events. However, this cannot generally be the case. This is especially clear in cases of expert performance. For instance, when experienced drivers drive a car, they usually do not pay attention to all the details that are involved in driving. Steering the wheel and moving one's leg from the throttle onto a brake pedal involves a sequence of movements that need to be properly coordinated in order to translate into successful action. However, there is no need to posit special desires that govern them. The solution is to think about agency as underpinned by hierarchical systems that top out in higher-order cognitive processes that play a role of a controlling device. If things go as planned the agential system runs on an autopilot, but when something goes awry then the controlling device usually takes over. (see Thalos 2007: 132-133)

Similarly, Parfit's psychological criterion of personal identity seemingly portrays agents as inactive bundles of linearly ordered psychological processes. In contrast, Korsgaard's emphasis on the practical perspective indicates that agents are not only passive bundles of unidirectional psychological processes, rather they are active bundles of processes whose activity or the ability to control that activity forms and maintains the unity of the bundle. Moreover, this dynamic picture of the self can also be discerned in Parfit's writings. In *the Nineteen Century Russian* story, Parfit (1984: 327) describes a young Russian who is about to inherit a large amount of land. Since he is an ardent socialist, he decides that once he inherits the land, he will donate it to peasants. However, he is also aware that once this happens in the future he might become a different person, someone who does not have socialist ideals anymore. Since he sees his current ideals as essential to his identity, he decides to do two things:

> He first signs a legal document, which will automatically give away the land, and which can be revoked only with his wife's consent. He then says to his wife, "Promise me that, if I ever change my mind, and ask you to revoke this document, you will not consent." He adds, "I regard my ideals as essential to me. If I lose these ideals, I want you to think that I cease to exist. I want you to regard your husband then, not as me, the man who asks you for this promise, but only as his corrupted later self. Promise me that you would not do what he asks." (Parfit 1984: 327)

With this example, Parfit seems to admit that current commitments, plans, and ideals structure and give contours to a person's self. In this sense, Parfit may also be interpreted as claiming that organizational aspects of agency play a determining role in delineating a person's identity. Thus, what comprises personal identities on this view are the agential structures, which involve desires, beliefs, intentions, values, and plans, that control and coordinate different sequences that compose actions.

In general, the notion of controlling structures can be used to naturalistically ground Korsgaard's ideas about the need to resolve conflicts among motivations and to regard oneself as a separate source of agency. Wayne Christensen (2007) provides an empirically grounded evolutionary story of how organisms when encountering a complex environments benefit greatly from acquiring more and more centralized higher-order controlling structures that can successfully maximize their fitness. Control structures function on the basis of feedback loops that send and receive signals from different components that they govern and respond to. Once those control structures have evolved they determine which options should be pursued, they resolve conflicts produced by subordinated systems, and coordinate activities of these subsystems to act successfully. Importantly, nothing compels us to think that these feedback structures have to be consciously accessible. Nevertheless, from an internal perspective, we can plausibly expect that these control structures, given that they are centralized in higher-order cognitive systems, ground a sense of unity in human agents. In effect, we can say that control structures ground the unity, which provides the locus that is an appropriate target of our practical concerns.

The important difference between the agency-based account that I sketched, and Parfit's account is that Parfit seems to allow any kind of cause to play a role in determining the psychological relations that are relevant for personal identity. (Parfit 1984: 207-208) In the agency-based accounts, the right kind of cause must stem from the capacities that underlie agency. In commonsensical terms, these are the capacities that enable us to act for reasons. Thus, Korsgaard talks about causes that enable "authorial connectedness." Those are exactly the capacities that enable us to make decisions and resolve internal conflicts. Using the present terminology, we might say that the controlling structures, which form a basis of human agency, provide a hierarchical framework that shapes the loci of potential experiences, values, and decisions, which in effect ground the relevant psychological connections.

The agency-based account of the self can still be regarded as reductionist. It does not postulate primitive metaphysical selves. In addition, it preserves the indeterminacy related to the Reductionist View, given that the control structures and the components it governs can be replicated or obliterated in the same sense and to the same degree in which psychological continuity theorists suppose that psychological processes can be replicated or obliterated. In addition, personal identity in this sense can be described from a third person point of view, for instance, by talking about control functions and its effects on the behavior of an agent (see Thalos 2007). Thus, it seems that if the Reductionist View is committed to some

form of the Extreme Claim, then the agency-based view will inherit its negative aspects.

However, the agency-based view enables us to see that the Reductionist View does not have all the negative consequences associated with the Extreme Claim. For instance, the agency-based view can accommodate and explain the fact that we have a special reason to care about our own future selves. This aspect of the agency-based view is especially salient in Michael Bratman's (1999) planning theory of agency.

What is distinctive about human agency is the ability to reflect upon our mental states, make plans that structure and govern our daily activities, and the fact that we conceive our agency as extended through time. (Bratman 2007: 21) For the present purposes, the latter two features are more important. The fact that we form plans and that we conceive our agency as being extended in time, indicates that we have foresight and an ability to control our actions when considering what will happen in closer or farther future. Furthermore, our ability to make short or long-term plans sets functional constraints on our available options. If I decide to go to work every morning, then I will have to settle on a plan that will enable me to successfully execute that decision. For instance, the ensuing intention to go to work will constrain me to wake up every morning at a particular time, to choose the most suitable route and means of transportation, to prepare lunch for that day, and so on and so forth. In addition, I will have to fill in that plan with further subplans that can respond to contingencies that might interfere with smooth execution of my intention to go to work every morning. These subplans might include the problem of deciding whether to cook at home or to buy lunch at work. Similarly to Korsgaard, what distinguishes us from other possible types of agents, according to Bratman, is our ability to conceive ourselves as beings that have this type of temporally extended agency. (Bratman 2007: 29)

What is important in this picture is that it explains why we have a reason to care about what happens to ourselves in the future. Given that we can form plans and that our agency is typically temporally extended, we have a reason now to care about what will happen to us in the proximal or more distant future. In particular, according to the agency-based view, what upholds and determines the psychological connections and continuities relevant for personal identity is to an important extent "a result of the agent's activity." (Bratman 2007: 30) What gives me a reason to care about someone who will inhabit my body in the future is provided by the fact that that someone is psychologically continuous and connected with me in virtue of being constituted and/or supported by my temporally extended agency.

It could be objected that while the agency-based view can explain why we have a reason to be self-concerned in shorter time spans, it will have a hard time explaining why we should care about the temporally more distant inhabitants of our bodies. Bělohrad advances this objection as follows:

> according to Korsgaard, living a life consists in planning and executing projects. It is the projects that force the person's identification, that is, authorial connectedness with a future self. The problem is that people's projects hardly ever span the extent of whole lives. Korsgaard may have shown that in order to carry out a plan, unity is required. But what she has failed to show is that these plans that people have and derive reasons from span their whole lives. (Bělohrad 2014a: 323-324)

Bělohrad, most notably, substantiates this objection by relying on empirical studies that provide evidence that people are relatively poor at long-term planning, delaying gratification, and in general tend to discount the value of future events or options. (Bělohrad 2014a: 325-326)

If agency-based account is relevant for justifying prudential concern, it must be able to provide some response to this objection. Here I will just sketch a possible route that an agency-based theorist might take. Notwithstanding the empirical facts about discounting, it can be replied that self-concern is grounded in the fact that people have *capacities* or *dispositions* for planning and extended agency. However, capacities or dispositions do not have to be manifested on every occasion. In addition, we have the ability to think of ourselves as having the capacity for extended agency. We project ourselves, grounded on our ability for agency, into the future. If we did not have these capacities, we would not find it rational to care about what happens to our future selves. In fact, on the one hand, a plausible explanation of why we think self-concern is rational is exactly the fact that we have a capacity for agency and planning and that we see ourselves as this type of agents. On the other hand, it is plausible that the same facts also explain why we normally think it is irrational to be poor at planning and to discount the future. We might conclude that having the capacity for temporally extended agency is a prerequisite of that aspect of personal identity that underpins the rationality of self-concern.

In this respect, the agency-based account mitigates at least one aspect of the Extreme Claim that is normally associated with the Reductionist View. In addition, the agency-based view can accommodate the idea that consciousness is not essential for personal identity. If we go back to the physics exam, we can see that the person with a divided consciousness is the same person as the one who decides to solve the exam by dividing the consciousness. The decision to solve the task in this way and her capacity to execute that intention through temporally extended agency grounds her identity through time.

However, it seems that the agency-based view does not mitigate all the misgivings related to the Reductionist View. In particular, adopting the agency-based view does not vindicate the idea that personal identity, as opposed to Relation *R*, underlies our most cherished practical judgments. For instance, in the fission case, it seems that our practical judgments track Relation *R* and not personal identity.

That being said, in the remainder of this section, by utilizing the fission case, I will explore how the agency-based view could provide a clue in what way personal identity might be spatially and temporally extended.

We start with the observation that agency seems to underpin the moral self, which, I maintained, is what our practical judgments actually track in the fission case. In order to possess a moral self (described in commonsensical terms), one has to possess capacities that underlie normal human agency. In addition, an agency-based account might explain in which sense our practical judgments might be, after all, tracking personal identity in the fission case. As an intuition pump, imagine that the criminal *X*, in order to enhance his chances of escaping the law, devises a plan which has the following elements: *X* decides to divide his brain hemispheres into two bodies, *Y* and *Z*. In addition, he devises total life-plans for both, *Y* and *Z*, in a way that will enable both of them to always perfectly coordinate their actions in escaping the hand of the law, and spending the rest of their lives in some place where they could live freely and happily. If it is really possible for *X* to devise a plan that is so specific and life-encompassing for *Y* and *Z*, and if *Y* and *Z* are capable of executing that life-plan, in a totally coordinated and mutually supportive way, that is, in the way *X* envisaged it, then it seems legitimate to say that *Y* and *Z* would be psychologically continuous with *X*. In addition, their psychological continuity would be constituted and supported by *X*'s agency. If this far-fetched case has any plausibility, then we might say that it provides grounds for thinking that *Y* and *Z* would be the same agent, albeit spatially distributed. However, whether this idea can be defended from the perspective of the agency-based view of personal identity, and what are its possible normative and other implications, I must leave open for future discussions.

6. Conclusion

In this article, the goal was to provide an opinionated overview of the psychologically based account of personal identity and the role of agency within such an account. I followed Parfit's (1984) exposition of the psychological criterion of personal identity. Furthermore, I indicated in what way the endorsement of the psychological criterion commits one to the Reductionist View of personal identity. However, I also argued that endorsing

this view does not commit us necessarily to what Parfit calls the Extreme Claim. In this respect, I showed how the agency-based view might be useful in answering the problems posed by the Reductionist View. By relying on Korsgaard's (1989) and Bratman's (2007) views of agency, I examined the possibilities of extending the psychological criterion of personal identity with considerations related to agency, in order to see whether, and in what way, agency could vindicate practical concerns traditionally related to personal identity. I argued that, though agency-based view is promising in accommodating some of the practical concerns we relate to personal identity, it probably leaves out some of the intuitive practical concerns we might also have. I finish by sketching an example, which pertains to show in what way an agency-based view might ground personal identity that is, not only temporally, but also spatially extended.

Acknowledgments

I would like to thank Boran Berčić, Luca Malatesti, and Leonard Pektor for reading and commenting on various versions of this article. Special thanks goes to Radim Bělohrad for giving extensive comments and suggestions on how to improve it. Of course, all the remaining doubts about the content are due to me.

The first draft of this article was written in the summer of 2016. at the BIAS institute (Nerezine). Research and writing was supported by the project Identity: Criteria of Synchronic and Diachronic Identity (University of Rijeka) and project CEASCRO (Croatian Science Foundation, grant number 9522).

REFERENCES

Bělohrad, R. (2014a). "Can We Do Without a Metaphysical Theory of Personal Identity in Practice?" *Prolegomena* 13: 315–334.

Bělohrad, R. (2014b). "On Schechtman's Person Life View." *Ethical Perspectives* 21: 565-579.

Baker, L. R. (2011). "First Personal Aspects of Agency." *Metaphilosophy,* 42: 1-16.

Baker, L. R. (2000). *Persons and Bodies: A Constitution View.* Cambridge University Press.

Beck, S. (2015). "The Extreme Claim, Psychological Continuity and the Person Life View." *South African Journal of Philosophy* 34: 314–322.

Bratman, M. (1999). *Intention, Plans, and Practical Reason.* Stanford: CSLI Publications.

Bratman, M. (2007). *Structures of Agency.* Oxford University Press.
Christensen, W. (2007). "The Evolutionary Origins of Volition." In D. Ross, D. Spurrett, H. Kincaid, & G. L. Stephens (eds.) *Distributed Cognition and the Will,* 255-287. MIT Press, A Bradford Book.
Davidson, D. (1967). "Actions, Reasons, and Causes." *The Journal of Philosophy* 60: 685-700.
Davis, W. A. (2010). "The Causal Theory of Action." In T. O'Connor & C. Sandis (eds.) *A Companion to the Philosophy of Action.* Oxford: Wiley-Blackwell: 32-39.
DeGrazia, D. (2005). *Human Identity and Bioethics.* Cambridge University Press.
Johnston, M. (1987). "Human Beings." *Journal of Philosophy* 84: 59–83.
Korsgaard, C. (1989). "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy and Public Affairs* 18: 101-132.
Locke, J. (1690/1998). *An Essay Concerning Human Understanding.* London, New York: Penguin Classics.
Maslin, K. T. (2001). *An Introduction to the Philosophy of Mind.* Cambridge: Polity Press.
McMahan, J. (2002). *The Ethics of Killing: Problems at the Margins of Life.* Oxford University Press.
Olson, E. T. (2002, August 2016). "Personal Identity." *The Stanford Encyclopedia of Philosophy*. (ed.) E. N. Zalta, Stanford, California, USA.
Olson, E. T. (1997). *The Human Animal: Personal Identity Without Psychology.* Oxford University Press.
Parfit, D. (1984). *Reasons and Persons.* Oxford University Press.
Parfit, D. (1995). "The Unimportance of Identity." In H. Harris (ed.) *Identity*. Oxford University Press: 13-45.
Roache, R. (2006). "A Defense of Quasi-Memory." *Philosophy* 81: 323-355.
Schechtman, M. (2008). "Diversity in Unity: Practical Unity and Personal Boundaries." *Synthese* 162: 405–423.
Schechtman, M. (2005). "Experience, Agency, and Personal Identity." *Social Philosophy & Policy* 22: 1-24.
Schechtman, M. (2014). *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life.* Oxford University Press.
Schlosser, M. E. (2011). "Agency, Ownership, and the Standard Theory." In J. H. Aguilar, A. A. Buckareff & K. Frankish (eds.) *New Waves in Philosophy of Action*: 13-31. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
Shoemaker, D. W. (2007). "Personal Identity and Practical Concerns." *Mind* 116: 317-357.
Shoemaker, D. W. (2016). "The Stony Metaphysical Heart of Animalism." In S. Blatti & P. Snowdon (eds.) *Essays on Animalism: Persons, Animals, and Identity*: 303-328. Oxford University Press.
Shoemaker, D. W. (1996). "Theoretical Persons and Practical Agents." *Philosophy & Public Affairs* 25: 318-332.
Shoemaker, S. (1984). "Personal Identity: A Materialist's Account." In S. Shoemaker & R. Swinburne (eds.) *Personal Identity.* Oxford: Blackwell.

Snowdon, P. F. (2014). *Persons, Animals, Ourselves.* Oxford University Press.
Swinburne, R. (1984). "Personal Identity: The Dualist Theory." In S. Shoemaker & R. Swinburne (eds.) *Personal Identity*: 1-66. Oxford: Blackwell.
Thalos, M. (2007). "The Sources of Behavior: Toward a Naturalistic, Control Account of Agency." In D. Ross, D. Spurrett, H. Kincaid & G. L. Stephens (eds.) *Distributed Cognition and the Will*. MIT Press, A Bradford Book: 123-167.
Williams, B. (1970). "The Self and the Future." *Philosophical Review* 79: 161–180.