

Can We Do Without a Metaphysical Theory of Personal Identity in Practice?

RADIM BĚLOHRAD

Department of Philosophy, Masaryk University in Brno,
Arna Nováka 1, 602 00 Brno, Czech Republic
belohrad@phil.muni.cz

ORIGINAL SCIENTIFIC ARTICLE / RECEIVED: 27-01-14 ACCEPTED: 17-06-14

ABSTRACT: In this paper I defend the idea that we need a metaphysical theory to justify identity-related practical concerns, such as self-concern. I outline D. Parfit's theory, in which the concerns receive a metaphysical justification. Then, I focus on two objections: C. Korsgaard's claim that the concerns are justified by the unity of agency, and M. Johnston's contention that the concerns are *prima facie* justified independently of a metaphysical theory. I argue that even if these theories have a point, they do not cover a range of situations in which justification may be sought. It is in these situations that a metaphysical theory may find its place.

KEYWORDS: Justification, metaphysics, personal identity, practical concerns.

Introduction

There are at least two major trends that can be traced in the metaphysics of personal identity. One trend seeks to develop a theory of personal identity to cohere with the broader picture of the identity and persistence of spatio-temporal objects. The other attempts to develop a theory of personal identity to justify some of the practical concerns and attitudes that have traditionally been taken to presuppose the concept of personal identity, such as self-concern, responsibility, compensation, blame, among others. Ideally, these theories should cohere in one unifying, practically relevant theory of personal identity. However, this ideal has turned out to be very difficult to achieve. The difficulty has been demonstrated clearly in Derek Parfit's theory and conclusions, which stated that identity does not matter for our practical concerns. (Parfit 1984) The number of papers discussing the details of Parfit's theory and the legitimacy of his conclusions is immense. However, less attention has been paid to the very assumption that it rests on: that the practical concerns

derive their justification from a metaphysical theory of personal identity. Some authors have attacked this assumption, claiming that the justification of the concerns is ultimately practical and that any metaphysical theory attempting to provide reasons for these concerns is irrelevant.¹ In this paper I will focus on two of such challenges and discuss whether they pose a threat to the metaphysical justification of the concerns, offer a viable alternative and, if so, leave room for any practical relevance of metaphysical theories.

First, I will focus on Parfit's theory, showing the relevant details of his methodology and setting out the basis for some of the key challenges in respect of this. I will then present Christine Korsgaard's challenge that Parfit's revisionary conclusions only follow because he ignores the view that personal identity is a practical, not theoretical concept, and that persons are agents who determine their own identity. I will then focus on Mark Johnston's challenge that the practical concerns need no metaphysical justification, because they are *prima facie* justified. I will go on to show that, even if Korsgaard and Johnston are right, there remains room for a metaphysical theory of personal identity.

Parfit's theory

Many people believe that the persistence of people is secured by psychological continuity. In part 3 of *Reasons and Persons* (Parfit 1984) Parfit defines psychological continuity as a chain of strong connectedness, where connectedness consists of the holding of particular direct psychological causal connections, such as an experience and a memory of it, an intention and an experience of executing it, and the persistence of a belief or a character trait. The beings that are unified by psychological continuity are termed *persons*. The beings that are unified by strong connectedness are *selves*. There are different theories about what enables connectedness. Non-reductionist theories hold that it is the existence of a continuous psychological subject, such as the soul or the *res cogitans*, that figures as the common subject of the various experiences in time. Parfit shows that non-reductionism is implausible. He believes that connectedness is the result of the existence of the body, brain and the related physical and psychological processes which together constitute the subject. He further argues that the physical aspect of continuity, that is, the continuous existence of the same brain is not important. Thus, the persistence of persons can be explained solely by reference to the existence of a chain of interrelated psychological events and processes. The problem is that psychological continuity as such is purely formal and, thus, replicable. That is, one may be psychologically continuous with a number of numeri-

¹The methodological alternatives are summarized and assessed in Shoemaker (2012).

cally distinct people. This intransitivity makes psychological continuity unsuitable as a criterion of numerical identity. This, however, need not trouble us, according to Parfit, because numerical identity is not important. The fact of numerical identity obtains when there is only one formal continuer of a person. But nothing important rests on there being just one such continuer. All that matters to us is psychological continuity and connectedness – relation *R* – not that it holds uniquely. Thus, it is *R*, not identity, which grounds our practical concerns.

One practical consequence of reductionism could be viewed as the dissipation of the self. If identity does not matter, one should be content with the mere existence of experiences causally connected to one's current experiences. Thus, if I intend to construct a nuclear fusion machine, it does not matter whether I accomplish that myself during my lifetime, or whether my son finishes it according to my instructions. The fact that the causal chain leaves my body is unimportant. What matters is the existence of mental states causally linked to my own mental states. Thus, every time I succeed in making such a link to another person, they carry part of what matters to me. "Me", then, exists scattered in the impressions I make on people. This has further consequences for some of the practical concerns. For instance, it is rational for me to care for others who are linked to me in this way. Also, it may not be immoral to compensate others for burdens imposed on me, as long as there is such a causal link between us.

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 *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.

These revisions of practical concerns follow, according to Parfit, from the acceptance of the correct metaphysical picture of the nature of people. For simplicity, I will mainly focus on one of the concerns – self-concern – in the rest of the paper.

Korsgaard's challenge

According to Korsgaard (Korsgaard 1989), there are two sides to a person. A person as a passive entity is the subject of experiences. As an active entity it is an agent, a doer of deeds. She believes that Parfit wrongly bases his theory on

the passive aspect only, seeking unity of experience and reducing agency to a form of it. She proposes that if we concentrate on the active aspect of persons, then the conclusions Parfit comes to do not follow.

Korsgaard agrees with Parfit that practical concerns are grounded in unity, but she rejects the view that that the relevant unity is a metaphysical unity of experiences. Instead, she suggests it is the *unity of agency* which grounds the concerns. For Korsgaard, metaphysics does not determine whether one should be self-concerned about a future person. In her opinion, even if I am only weakly connected to a future self of mine, I should still feel self-concern for the self, because the self will still be the same agent as me out of *practical necessity*.

This is important, because if it turns out that practical necessity provides a strong unity relation, then the fact that our metaphysical unity is weak does not justify the revision in practical concerns that Parfit proposes, and metaphysical theories of personal identity become irrelevant for the justification of the practical concerns.

On the other hand, if we can show that practical necessity doesn't provide a strong unity relation, the revision of the practical concerns may be in place and we may still need a metaphysical theory of personal identity to guide us in the revision.

Let us see, then, what the practical necessity that Korsgaard refers to consists in. Korsgaard begins her account with an explanation of the necessity of unity *at a time*. A major role in this type of practical unity is played by the fact that even though I might simultaneously have conflicting desires and intentions, giving me the impression that there are more subjects in my body, I am eventually bound to come to a single unanimous decision, because I only have one body to act with. So it is *the fact that I have a single body* which prevents me from carrying out simultaneous incompatible acts.

Another aspect of the necessity of unity *at a time* is that people deliberate from a particular standpoint. For example, I have desires that may conflict, but I also have desires about the desires. These give me reasons for weighing the first-order desires and choosing among them. I do not feel that I passively wait to see which desire happens to win. I feel that I actively choose among them and choose according to a principle that is expressive of my will. The *identification with a principle of choosing*, an identification which I must do to act, is, then, another aspect of the necessity of unity at a time.

Korsgaard admits, though, that in the process of identification with a principle of choosing, one may be guided by metaphysical considerations. That is, I may decide on particular acts affecting the well-being of a future person on the basis of a belief that the person will be me, for instance. She believes, however, that there are more important grounds for such identi-

fication, which are independent of any metaphysical beliefs and which are imposed on us by the necessity to act.

Korsgaard concedes that what has been said about unity *at a time* is fully compatible with the idea that a body is occupied by nothing but a chronology of instantaneous conscious subjects (Korsgaard 1989: 112). That is, it is fully compatible with the idea that there is no unity *over time*. So an argument is needed to show that diachronic unity is practically necessary, too. Korsgaard claims that if the subjects want to have any kind of a life, they have to form a unity. Life does not consist in a series of temporally isolated experiences. Living a life means *having plans and projects* that take time; and having and completing a project requires my cooperation with the future inhabitants of my body and my identification with their expected needs and desires. This identification is confirmed each time we brush our teeth or exercise, as we project ourselves to the future and see our lives from the perspective of the future inhabitants of our bodies. The identification with these future selves is what actually gives us reasons for our current choices and actions. This is why Korsgaard says: "... to the extent that you regulate your choices by identifying yourself as the one who is implementing something like a particular plan of life, you need to identify with your future in order to be *what you are even now*" (Korsgaard 1989: 113–114, original emphasis). Since we only have one body and one life to live, it follows that we must consider ourselves as unified over time: "The unity of our life is forced upon us, although not deeply, by our shared embodiment, together with our desire to carry on long-term plans and relationships" (Korsgaard 1989: 113).

With the concept of unity of agency, Korsgaard explains the unity of consciousness. For her, unity of consciousness is just an instance of the unity of agency. The unity of consciousness is the absence of any difficulty coordinating and integrating conscious activities.² Such coordination is only possible if there is some form of communication between the different conscious parts; and communication is preconditioned by the unity of agency.

Like Parfit, Korsgaard draws an analogy between the unity of persons and the unity of certain groups. Parfit argues that persons are like nations. The existence of nations consists in the existence of people living on a territory. If a nation mingles with a great number of immigrants over the course of several decades while its territory expands, there is no fact of the matter as to whether it is the same nation. Similarly for persons – the existence of a person amounts to the existence of a body, a brain and a series of inter-related physical and mental events. If a substantial amount of these is lost or replaced, the identity of the person becomes indeterminate. Korsgaard believes persons should rather be likened to states. States, unlike nations,

² For a critique of this claim, see Shoemaker (1996: 323–324).

are formal entities defined by their constitutions and deliberative procedures whose citizens have actively constituted themselves into a single agent, have adopted ways of resolving conflicts, making decisions, interacting with other states and planning together for an ongoing future (Korsgaard 1989: 114). Korsgaard believes that the turning of a nation into a state is pragmatically necessary, and so is, analogically, the turning of a bundle of experiences into a person:

When a group of human beings occupy the same territory, for instance, they have an imperative need to form a unified state. And when a group of psychological functions occupy the same human body, they have an even more imperative need to become a unified person. This is why the human body must be conceived as a unified agent. (Korsgaard 1989: 115)

The final sentence of this quotation illustrates a point worth stressing: no matter how many conscious subjects might there be in the history of a single body, they are *all* bound to be unified into a single agent by the desire to live a life and by sharing the same body. Korsgaard re-emphasizes this point at another place: “So long as I occupy this body and live this life, I am this rational agent, the same one” (Korsgaard 1989: 126). That does not mean, however, that an agent cannot form larger unities with other agents based on common projects and plans. Families, companies and political parties may all be unified by the very same relation that unifies conscious subjects within one body, which also justifies the extension of self-concern to these entities (Korsgaard 1989: 127). Korsgaard’s point is, rather, that the *minimal extent* of an agent equals to that of one life lived in one body. Importantly, the claim that *all* of the conscious subjects sharing a body are necessarily unified is needed to block Parfit’s conclusion that self-concern may reasonably fade proportionally to the reduction of psychological unity. If all subjects in a body are unified out of practical necessity, as Korsgaard claims, no reduction of self-concern is warranted.

There is one fundamental difference from Parfit’s theory that emerges in Korsgaard conception of the unity of agency. As we have seen, in Parfit’s theory the metaphysical unity of people is provided by relation R, part of which is the scalar relation of connectedness, and when connectedness wanes, so does unity. Thus, the persistence of a person is inconsistent with sudden drastic psychological changes which disrupt psychological continuity. However, if one introduces the idea of agency into the picture, this need not be so. Agency involves regarding oneself as the originator and author of one’s actions. The relation between an agent and experiences or actions which the agent actively arrives at, is convinced of, decides on or endorses, Korsgaard terms “authorial connectedness”. Authorial connections enjoy a privileged status among other psychological connections – if the psychological changes

are the result of your active decision and identification with the product of the changes, you will remain the same agent even if the changes are drastic. On the other hand, Korsgaard also discusses thought experiments about mad surgeons (Korsgaard 1989: 122–123), involving drastic *external* interference with a person's psychological continuity. In one such experiment, a man is seized by someone who tells him that he is going to be tortured the next day. Before that, however, all of his memories will be wiped out and replaced by a complete set of new memories. Korsgaard agrees that in such cases the changes bring about the loss of the person's identity. But this is not so much due to the intrinsic nature of the changes, but rather to the external and unauthorized origin of the changes.

The implications for self-concern are obvious. Where one actively identifies with the changes to occur, they have reasons for concern even if the changes disrupt psychological continuity. Suppose, hypothetically, that, like the character Joel in the film *Eternal Sunshine of the Spotless Mind*, you decided to have a substantial portion of memories wiped out, which would lead to a substantial disruption of psychological continuity. Nevertheless, because it was your decision to initiate such a profound change, you would have reasons to identify with and feel concern for the person occupying your body after the change.

An assessment of Korsgaard's challenge

Before I raise some objections to Korsgaard's theory, it is important to clarify what exactly it is that I target and want to achieve. Let me, therefore, briefly recapitulate. Parfit argues that with the truth of reductionism come important implications for our identity-related practical concerns. Particularly, where the metaphysical relations that unify people weaken, the practical concerns may legitimately weaken as well. The case which I find most illustrative is the relationship between people and their distant future selves, such as the relationship between a carefree teenager and the grumpy old man who will occupy his body in sixty years. On Parfit's theory, if the teenager is weakly psychologically connected to the old man, his feeling little self-concern for the old man's well-being may not be irrational, because the man, while being the same person, is a different self. For Korsgaard, these theoretical considerations do not determine the rationality of self-concern. Instead, practical considerations do so. And if we focus on these, we must see that the teenager is the same agent as the old man, because he only has one body and one life to live, and in order to live that life he must identify with all the conscious subjects that will occupy his body. In the next few paragraphs I would like to show that although people desire to live and they only have one body, they frequently fail to have projects that span their whole lives, fail to derive rea-

sons for their current decisions from the distant future selves, and, thus, they fail to form agential unity with their distant future selves. As a result, many people's actual lives cast doubt on the thesis that all conscious subjects in one body are necessarily unified into a single agent. But since we have seen that one needs this thesis in order to block Parfit's revisions of self-concern and other practical concerns, it follows that the revisions withstand Korsgaard's challenge.

Before I elaborate on this further, however, I shall first focus on Korsgaard's analogy of persons and states to see whether it provides good reasons to accept the necessity of the unity of agency.

The unity of states

As we have seen, Korsgaard likens the pragmatic unity of agents to the unity of states. Let us see, then, to what extent the unity of states is pragmatically necessary. Korsgaard claims that when a group of human beings occupy the same territory, they have an imperative need to form a state. However, history gives us plenty of examples that refute this alleged necessity of unification. Firstly, there are groups of people or nations that have been unified with a larger group on a territory against their will or who have been hesitant about the need for unification: the Baltic states, which were annexed to the Soviet Union in 1940; the Québécois in Canada who have held two referenda on sovereignty; and the nationalists in Northern Ireland are just a few. Secondly, there are groups who have freely declined to join with larger unions, such as Switzerland or Norway with respect to the EU. Of course, the EU is not a state, but it is a clear example of the unity of agency. Thirdly, there are states that have disunited, such as Czechoslovakia in 1993, in spite of sharing common history on the same territory. The historical experience of these groups suggests that the need for unification is not an imperative.

It may be argued that these are not counterexamples to the necessity thesis: in each case, the reluctance to join, or the desire to part, may be viewed as the result of a stronger sense of unity within a sub-group, and a fear of the loss of identity within the larger political or economic unit. But then the argument rests on what is meant by "occupying the same territory". If the argument is not to turn into a triviality (in which "the same territory" simply means an area inhabited by a group of cooperating agents), the notion of territory must be given an independent demarcation, such as a geographical boundary. Is Europe, then, such a territory? If so, Switzerland refutes the necessity of the unity of agency. If not, is Germany, for instance, such a territory? It is true that the German people were forcibly disunited after WWII and reunited after the fall of the iron curtain. But some current opinions advocate that the reunification has not been very successful and is not consid-

ered a need by many (Fuchs 2008; Zeitchik 2003). These issues suggest that the necessity of unity is conditional upon the need and willingness of a group of people to cooperate. In fact, Korsgaard's notion of the unity of agency is best interpreted as intentional cooperation. But people living in an independently demarcated territory may not always need or be willing to do so. It may be likely that if you put a group of people on such a territory, unities will quickly and naturally emerge. But it is by no means necessary that one encompassing unity within the territory will form. The situation in which the protagonists in *The Lord of the Flies* found themselves called for unity as acutely as any, yet they failed to co-operate as one body.

I conclude, then, that historical perspective shows that the unity of states is not pragmatically necessary and is a weak analogy to use to argue for the necessary unity of individual agency.

It may also be argued, however, that while the analogy is not watertight, the comparison with the unity of states illustrates the basic mechanisms underpinning a need for unification, which, although not quite necessary in the case of states, are always present in the case of persons. The crucial difference between states and people that might support this view lies in the spatial divisibility of territories and bodies respectively. While a group of people who wish to disunite are not bound by essential indivisibility of the territory and can easily form independent functional units in its sub-regions, the human body cannot be functionally disintegrated in a similar manner. It is physically impossible for the human body to separate into multiple independently controlled units. The human body, by its very nature, must exist as a physically integrated whole, and, regardless of the number of potential subjects of experience that might occupy it, they must exhibit unity of action in order to live. If that is so, people might still be unified in virtue of having just one indivisible body.

This idea may be plausible in the case of unity at a time. However, it is far less plausible in the case of temporal unity. In what follows, I will argue even though to carry out a long-term plan people have to identify with the future selves that are part of the plan, this does not mean that people have to identify with *all* the conscious subjects that will occupy their bodies in the future. This is because people, as a matter of fact, may not have plans long-term enough to include distant future selves.

The temporal extent of projects

As we have seen, 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. But that is precisely what needs to hold for Korsgaard to be able to draw the conclusion that the fact that a future person will occupy my body forces me to identify with the person. I only have such an imperative if the future person is involved in my projects. But if my projects do not reach far enough in the future, there will be some future conscious inhabitants of this body that I do not currently identify with and, therefore, do not form unity with in order to carry out the projects. In other words, Korsgaard only shows that there must be unity in one project team, and I fully agree with that. She does not show, however, that the team necessarily consists of all the occupants of one body.

The question whether people's projects span their lives is an empirical one and evidence seems to support the claim that people's lives are not unified in the manner Korsgaard supposes them to be. First of all, there are self-reports of *episodics*. Galen Strawson (Strawson 2004), who claims to be one of them, describes episodics as people who do not figure themselves, considered as selves, as something that was there in the (further) past and will be there in the (further) future (Strawson 2004: 430). Strawson describes a self as "that which I now experience myself to be when I am apprehending myself specifically as an inner mental presence or self" (Strawson 2004: 433). Although he recognizes that as a human being he is one continuous entity, considered as a self, the events in his remoter past and future did not/will not happen to him. At the same time he claims that he does not have any deep interest in his past or concern for his future:

As for my practical concern for my future, which I believe to be within the normal human range (low end), it is biologically – viscerally – grounded and autonomous in such a way that I can experience it as something immediately felt even though I have no significant sense that I* will be there in the future. (Strawson 2004: 434)³

Although Strawson's claims are primarily targeted at narrative theories of personal identity, they apply to Korsgaard's agential theory as well. Strawson's explicit admission of the lack of unity with his distant past and future selves, while admitting that he is one continuous human being, clearly suggests that Korsgaard's claim that the conscious subjects occupying one body must be conceived of as a unified agent does not have universal acceptance. In fact, several empirical experiments show that Strawson's remark about people's practical concern for their future being low end expresses a relatively pervasive feature of human lives.

³ Roughly, Strawson uses "I*" to refer to himself as a self (as defined above), while "I" for himself as a human being.

A number of experiments have focused on the future orientation of people, including issues such as planning, future concern, inter-temporal choices, consequence anticipation, the subjective value of future gains and losses. There is a salient group of people that on the face of it appear to refute Korsgaard's views about the practical necessity of unity – adolescents. There is a popular stereotype that adolescents are notoriously shortsighted, incapable of making long-term plans and commitments and anticipating the consequences of their acts. If living a life amounted to having long-term projects, many adolescents could not be described as living lives in Korsgaardian terms. It may be objected, though, that adolescents cannot constitute a counterexample, as the research that confirms this stereotype also shows that their low rate of future orientation is associated with the underdevelopment of several psychological faculties and brain functions. These may involve the immature formal operational thinking observed in adolescents, along with limitations in working memory, the slow maturation of prefrontal cortex and the increase in reward salience caused by the hormonal changes in puberty. (Steinberg, et al. 2009) It could be claimed that adolescents are not yet mature agents, which society recognizes by denying them certain rights and curbing the demands we place on them.

However, further research shows that adolescents may only be worse at what adults are actually quite poor at, namely long-term future orientation. Economists have studied an aspect of future orientation under the term *temporal discounting* and have confirmed that people show relatively less regard for delayed gains and losses compared to immediate ones. Indeed, the assumption that they do so is a standard assumption in economic research (Löckenhoff, O'Donoghue and Dunning 2011). One study showed that 14 to 15-year-old adolescents discounted the value of \$1,000 to be received in a year to only \$343. Adults aged between 26 and 30 years old were shown to discount the same \$1000 to \$442, within the same time horizon (Steinberg, et al. 2009). The monetary representation of the discount rate illustrates that, while more future-oriented than adolescents, adults still show a notable disinterest in future outcomes (Löckenhoff, O'Donoghue and Dunning 2011). Apart from age, differences in temporal discounting have also been correlated with cultural differences (Du, Green and Myerson 2002) and social roles (Trommsdorff 1986). This suggests that the general human tendency to discount the future has both quantitative and qualitative individual variations.

Explanations for this phenomenon have varied, but a general explanatory pattern may be proposed. Future orientation emerges from the interplay of two brain systems. The socioemotional or "hot system" mediates impulsivity and reward seeking, while the cognitive or "cool system" is responsible for impulse control and deliberative abstract reasoning. Even though with increasing age, the balance between the two systems shifts in favor of

the cool system, enabling adults to down-regulate the decision-related emotional responses which appear to push adolescents to seek immediate rewards (Löckenhoff, O'Donoghue and Dunning 2011), this does not eliminate the general tendency to discount future outcomes.

Some authors suggest that future orientation is a multidimensional construct which includes: rational anticipation of future consequences; premeditation; and the length of time one is able to project one's imagined life into the future (temporal extension). Other factors may include the optimism or pessimism about one's future and the extent to which one thinks one has control over one's future. Steinberg et al (2009) suggest that each of these require a separate line of explanation. What is salient is that the cited components of future orientation behavior can reasonably be supposed to play a key role in the formation and extent of projects and plans which, according to Korsgaard, constitute our lives and determine reasons for our actions. These findings suggest that one's future orientation has limited temporal extent, being rich and vivid in the near future and fading with time.

The temporal extent of the unity of agency

We have seen that Korsgaard, after all, admits that one may fail to connect authorially to a future self. This can happen if one is changed by *external intervention* against one's will. So it must be admitted that the unification of all conscious subjects occupying a body may not always be necessary.

David Shoemaker (Shoemaker 1996) shows, however, that it is not just in cases of external intervention that one may fail to connect to conscious subjects that share one's body. He describes a case, focusing on *past orientation*, in which an eighty-year-old woman examines actions that she performed as a twenty-year-old and realizes that she actually wants to distance herself from the values, beliefs and actions of the twenty-year-old and achieve an attitude of indifference toward her (Shoemaker 1996: 330). Shoemaker argues that this case constitutes an example of *authorial disconnectedness*, that is, a state of disunity resulting not from external intervention, but from wilful and deliberate personal decision to cut oneself off from one's past.

While one may wonder whether it is psychologically possible to become absolutely indifferent to one's past, I believe that the case does show that one can actually become opposed to the principles that once determined one's decisions and actions. The person distancing herself in such a way is explicitly refusing to become a member of the previous self's project team, and refusing to continue living the life that the previous self had planned. So I do agree that this is another case that casts doubt on Korsgaard's claim that as long as one occupies the same body and lives the same life, one is the same rational agent.

The research outlined above, however, shows still another mode of non-identification, and one, which, I believe, is much more widespread. While the case outlined by Korsgaard is one of *unauthorized* external interference, and the case defended by Shoemaker is one of *deliberate* internal change and dissociation, the research suggests a *passive and unintended* failure to identify with one's future selves.

The findings may imply several things relevant for Korsgaard's claims. Firstly, it seems that even if people desire to live and, as things stand, can only live in one body, they cannot always be considered unified with their distant future selves. Secondly, when people derive reasons for their current acts, the strongest reasons usually come from their identification with nearer future selves, not the ones in distant future. Lastly, living a life may consist in having plans and projects, but these are usually shorter-term than Korsgaard supposes them to be.

I shall have more to say about the implications for the role of metaphysics in personal identity once I have assessed the other challenge. Let me, thus, turn to it now.

Johnston's challenge

Let me now turn to Mark Johnston's argument presented in his "Human Concerns Without Superlative Selves" (Johnston 1997). Johnston objects to Parfit's theory that the revision of practical concerns that Parfit suggest does not follow from his rejection of non-reductionism.

First, Johnston shows that Parfit's position is best interpreted as *ontological reductionism*. According to this doctrine, the existence of a person consists in the existence of a body with a brain and the physical and mental processes occurring in them. At the same time, a person is not *identical* to the sum of these entities, because people and bodies have different persistence conditions. Rather, people are *constituted* by bodies. According to Johnston, because people are not identical to their bodies, it follows that the fact that a person exists is different from the fact that a body exists.

Non-reductionism, Parfit's target, embodies the claim that a person is neither identical to, nor constituted by the body. The existence of people, therefore, is completely independent of the existence of bodies.

Johnston claims that although in both conceptions the existence of a person is a fact that is different from the fact that a body exists, there is an important difference. The facts that are based on the relation of constitution, Johnston terms *ordinary further facts*. The facts of non-reductionism are termed *superlative further facts*. Superlative further facts inflate our ontology. The existence of a soul or a *res cogitans*, independent of the underlying physi-

cal and mental phenomena, would further extend the list of the fundamental entities. Ordinary further facts are ontologically innocent – the only entities that exist are bodies, brains and the relevant processes, and the obtaining of the facts about persons adds nothing to this list. In Johnston’s words, making statements in the discourse in question carries no commitment to entities other than those spoken of in some other, philosophically favoured discourse (Johnston 1997: 262).

Parfit’s mistake, Johnston claims, is that showing that there are no superlative further facts does not justify the conclusion that the practical concerns that seem to have tracked the superlative further facts are unjustified. Johnston purports that even in the absence of superlative further facts, the practical concerns are reasonable and justified *prima facie*. If this is so, he suggests that no revision of our practical concerns as defended by Parfit is necessary.

Self-concern and self-referential concern

Johnston defends a position which he terms *Minimalism*. According to minimalism, “metaphysical pictures of the justificatory undergirdings of our practices do not represent the real conditions of justification of those practices” (Johnston 1997: 260–261). The contention is, therefore, that the metaphysical pictures are mere theoretical epiphenomena that spring from the reflection on the practices, but do not have any significant impact on them.

Johnston focuses primarily on self-concern. He believes that self-concern does not require the existence of a continuing soul or a *res cogitans* for its justification. That is, the justification is not grounded in a superlative further fact. Self-concern derives its justification from the fact that it is one of self-referential concerns, “directed outwards from one’s present self to one’s future self, one’s friends, family, acquaintances, neighbourhood, and so on” (Johnston 1997: 268). In each of these concerns we care about the objects in a non-derivative way, for their own sake, and having these concerns is natural. In fact, not caring non-derivatively for a family member or a friend “would be regarded as lacking a kind of attachment which is often a central part of living a significant life” (Johnston 1997: 268). The network of self-referential concerns obtains a holistic *prima facie* justification – these concerns are considered legitimate unless reasons are provided to cast doubt on some of them. However, they cannot all be doubted at once. This indicates why the discovery that there are no superlative further facts of personal identity has no effect on the justification of self-concern – it was never justified by the superlative further fact in the first place.

To clarify the irrelevance of the metaphysical underpinnings of our practical concerns, Johnston invokes a variant of Parfit’s teletransportation

thought experiment. This is framed as follows: imagine a teletransportation device is used to produce perfect replicas of astronauts for a deep space probe who would never return to earth. Suppose further that the machine leaves the original in a slightly worse mental shape than the replica, with mild memory loss and disorientation. If one conceives of this scenario from the first-person perspective, one can see how natural self-concern is. I would strongly prefer that my replica, as opposed to me, be sent on the mission. Why is that, assuming that my replica will be more R-related to me? Shouldn't I strongly prefer that my replica stay and avoid the horrors of the space mission if what really matters is relation R, as Parfit suggests? Johnston claims that we take such a position because each of us has a bias in favor of himself and his own welfare as against that of strangers who happen to be R-related to him (Johnston 1997: 274). Parfit is bound to say that the rational choice is to leave my replica on earth and go myself. The bizarreness of this attitude is highlighted by an adjustment of the scenario. Imagine that the astronaut is not me but a friend of mine. If what matters is relation R, I should antecedently prefer that my friend's replica be left behind with me on earth. Johnston finds this a *reductio* of Parfit's position. For Johnston, "[F]riendship does constitutively involve valuing the friend for his own sake and not being disposed to weaken the bonds of friendship just because of some psychological change in the friend" (Johnston 1997: 276). My friend could legitimately object that I do not care for him for his own sake and only care about his psychological make-up. If what matters is psychological continuity and connectedness, as Parfit claims, friendship has come to an end.

This case is meant to demonstrate that concern for friends does not require an underlying metaphysics of friendship postulating friends as entities existing separately from their bodies, brains and functions. We would naturally find it absurd to demand that. At the same time, it shows that in the absence of such entities, we are not entitled to treat our friends as sums of mental and physical features that we can replace when they change. It is reasonable to be concerned for our friends for a simple reason – because they are our friends.

Analogically, even though the concern for our future selves does not require the existence of a soul or a *res cogitans*, it is not based merely on some notion of qualitative closeness between me and my future self. I am concerned for my future self for a straightforward reason – because it will be me (Johnston 1997: 270).

In sum, self-referential concerns are justified by ordinary further facts (that one will exist, that someone is one's friend, etc.), not superlative further facts, and their justification does not require any sort of metaphysical underpinning.

An assessment of Johnston's challenge

While I believe that Johnston's arguments for the naturalness and reasonableness of our concerns are sound, I think that they are consistent with the idea that in some cases a metaphysical theory may actually be useful in guiding us if the rationality of our actions is questioned.

To illustrate the usefulness of metaphysics, let me first focus on Johnston's contention that friendship constitutively involves valuing the friend for his own sake and not being disposed to weaken the bonds of friendship just because of some psychological change in the friend. I completely agree that we do not replace friends just because they have slightly changed, and that to do so would rightly be regarded as cold and inhuman. On the other hand, our own experience tells us that many friendships develop, and sometimes end for understandable reasons. Friends drift apart, perhaps as a result of other relationships they enter or political engagements they develop or the ideas they adopt as a result of life events. If that is so, the reason some friendships end is the consequence of some sort of psychological incompatibility that develops between the people. This shows that there must be a threshold and that friendship is at least partly based on the psychological make-up of friends.

I completely agree with Johnston that if I recognize someone as a friend, I do not question whether I should care for him for his own sake. The problem is, though, how to structure our concerns for people who have changed so much that we are uncertain about whether they still are friends. If our self-referential concerns are *prima-facie* justified, we may simply trust our gut feelings and rest content with any form of attitude that we naturally adopt toward such people. But this would ignore the fact that in some cases the attitude we naturally adopt may be considered irrational by people around us, and we simply fail to see that. Consider, for instance, the case of a wife whose husband has turned into an abusive alcoholic, while she still believes that he is going to change and keeps forgiving his attacking her. If her attachment to the husband persists, at a certain point it may become unreasonable. However, in interpersonal relations people do not always see the borderlines clearly and may adopt or retain attitudes that are inappropriate.

Respecting the analogy between the concern for friends and family and the concern for one's future self, similar observations apply. My non-derivative concern about my well-being is completely natural and can withstand various changes in my psychology. Sometimes, however, the changes may be more drastic, and what the reasonable attitudes in such cases are may not be clear to anybody involved. A person diagnosed with dementia pondering whether to sign an advance directive may be an illustrative example. Someone might suggest, inspired by the thesis of non-derivative concern: "You should

be concerned for the demented person for its own sake because it will be *you*". However, such claim simply begs the question in this scenario. Informed people reasonably disagree about whether the demented being will be the original person, and, thus, about whether her concern would be reasonable.

Another illustrative example of a case where a natural self-referential concern is actually present, but is very likely misplaced, is the concern that people in some cultures have about the way their bodies are treated after death. Some people prefer that their bodies be buried, rather than cremated, because they believe that God can resurrect the body. While we may say that this concern is quite natural, a careful metaphysical analysis in Shoemaker (2009: 23–57) shows that this practice can hardly be in their self-interest and help secure immortality.

Furthermore, Johnston himself allows for the possibility of a revision of *particular* concerns when he says: "One can fail to identify with one's future self. But this will seem reasonable only if there is some considerable reason to inhibit the natural tendency to so identify, the natural tendency around which is built one's concern that one's own life continue, go well, and be worthwhile" (Johnston 1997: 268). Thus, it seems that even if we recognize that self-referential concerns are natural and generally reasonable, there are still cases in which we may legitimately seek reasons to care. It is the cases in which it is not clear whether the ordinary further facts which ground our concerns obtain that make us wonder what the rational thing is for us to do. This is where a metaphysical analysis of the relations underlying our concerns may play its role.⁴

Conclusion

We are now in the position to put the individual strands of the argument together. Derek Parfit believes that our practical concerns require a metaphysical theory of personal identity to be justified. He has shown that the metaphysical theory that was assumed to justify the concerns is false, and that the true theory of personal identity, reductionism, cannot justify the concerns in the form that we have been used to. Thus, the concerns have to be reformed. However, both Christine Korsgaard and Mark Johnston think that his conclusion is too hasty. Both reject Parfit's assumption that the concerns are justified by means of a metaphysical theory. Korsgaard believes that the concerns are justified by practical necessity, given the fact that we are embod-

⁴ Shoemaker (2007: 346) concludes that metaphysical theory can play a *revelatory* role. It may tell us what it is that enables a person to have a special concern for her future selves as opposed to others and reveal heretofore-unrecognized connections between our attitudes towards ourselves and attitudes towards others.

ied and we must cooperate with whoever it is that has occupied and will occupy the body in order to live a life. The metaphysical fact that these subjects might not be *me* is practically irrelevant. Johnston argues that the practical concerns are justified independently of any metaphysical theory and in spite of the absence of the superlative further facts by the obtaining of ordinary further facts and by being completely natural.

I have attempted to show that the relation of unity of agency defended by Korsgaard is not strong enough to support the practical concerns that are collapsing under the attack of Parfit's arguments. The temporal extent of the unity of an ordinary person is not wide enough to justify that all of the conscious subjects in her body are part of one encompassing project, thus constituting the same agent. As a result, if the unity of agency is the relation underlying our practical concerns, their traditional form is untenable. In fact, Shoemaker (1996), drawing on his conclusions about authorial disconnectness, convincingly argues that agents on the Korsgaardian picture closely resemble selves on the Parfitian picture. Thus, if agents are the objects of our concerns, a revision of the concerns seems inevitable.

This similarity between the metaphysical entity of *the self* and the practical entity of *the agent* has further implications for the central question of this paper – the role of metaphysics. If the above arguments are correct, it would seem that we may simply make a choice. We may either consider ourselves from the theoretical perspective as selves or from the practical perspective as agents. The practical implications for the rationality of practical concerns will be similar.

I have also attempted to show that Johnston does not completely eliminate the need for metaphysical analysis, either. I agree that our self-referential concerns are completely natural if we are certain that the ordinary further facts on which they rest obtain. But in many cases people are at a loss whether they do obtain: Is the man whom his new wife is setting against me still *my friend*? Is the adolescent who has been brainwashed by a religious cult and scorns everything I have taught him still *my son*? Will the person lying on the hospital bed in permanent vegetative state still be *me*? We do not doubt that all of the concerns are perfectly natural if we give a confident yes to all of these questions. But what is the rational stance towards these people if we are in doubt?

Of course, these worries are not worries about the identity of the relevant individuals in terms of bodily or biological continuity. Rather, the worries are whether the individuals are psychologically unified strongly enough for us to consider them to be the same agents (on Korsgaard's theory), or whether they have psychologically changed so dramatically that our lack of attachment would not be regarded as cold and inhuman (on Johnston's theory).

Korsgaard's theory would succeed in defending the traditional form of concerns only if our life-constituting projects extended from the first conscious inhabitants of our bodies to the very last ones. This would show that agents are practically unified from the moment consciousness first appears in a body until its demise, thus making metaphysical observations about unity irrelevant. Johnston's theory would eliminate the need for metaphysical analysis of personal identity if for every person (including our future selves) we could determinately say whether or not they stand to us in the relation that makes the concerns that we naturally have or naturally fail to have reasonable. I have attempted to show that these conditions may be violated in ordinary lives. This leaves us with a limited, but still useful role for metaphysical analysis. There are people with whom I actually fail to identify due to emotional or cognitive limitations. But perhaps it would be in my best interests to do so. There are people for whom I currently feel no non-derivative self-referential concern. But would it be reasonable for me to do so? Where agential unification with and natural concern for people are actually absent, the metaphysical analysis of the underlying relations to them might tell us whether we are being reasonable.⁵

References

- Du, W., L. Green & J. Myerson. 2002. "Cross-cultural comparisons of discounting delayed and probabilistic rewards", *The Psychological Record* 52, 479–492.
- Fuchs, C. 2008. "Das verschmähte Paradies", *Zeit Online*. <http://www.zeit.de/campus/2008/05/ost-west-unis>. Accessed 31 October 2013
- Johnston, M. 1997. "Human Concerns Without Superlative Selves", in R. Martin, J. Barresi (eds.), *Personal Identity* (Oxford: Blackwell Publishing), 260–291.
- Korsgaard, C. M. 1989. "Personal identity and the unity of agency: A Kantian response to Parfit", *Philosophy & Public Affairs* 18, 101–132.
- Löckenhoff, C. E., T. O'Donoghue, D. Dunning. 2011. "Age Differences in Temporal Discounting: The Role of Dispositional Affect and Anticipated Emotions" *Psychology and Aging* 26, 274–284.
- Parfit, D. 1984. *Reasons and Persons* (Oxford: Oxford University Press).
- Shoemaker, D. 1996. "Theoretical Persons and Practical Agents", *Philosophy & Public Affairs* 25, 318–332.
- . 2007. "Personal Identity And Practical Concerns", *Mind* 116, 317–357.

⁵ *Acknowledgments*: The work on this paper has been supported by grant no. 13–00624P of the Czech Science Foundation (GAČR). I am also indebted to David Shoemaker for discussing aspects of C. Korsgaard's theory with me.

———. 2009. *Personal Identity and Ethics: A Brief Introduction* (Ontario: Broadview Press).

———. 2012. “Personal Identity and Ethics”, In: Zalta, E. (ed) *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/spr2012/entries/identity-ethics/>. Accessed 31 October, 2013

Steel, P., C. J. König. 2006. “Integrating Theories of Motivation”, *Academy of Management Review* 31, 889–913.

Steinberg, L., S. Graham, L. O’Brien, J. Woolard, E. Cauffman, M. Banich, M. 2009. “Age Differences in Future Orientation and Delay Discounting”, *Child Development* 80, 28–44.

Strawson, G. 2004. “Against narrativity”, *Ratio* 17, 428–452.

Trommsdorff, G. 1986. “Future time orientation and its relevance for development as action”, in R. K. Silbereisen (ed.), *Development as Action in Context: Problem Behavior and Normal Youth Development* (Berlin: Springer), 121–136.

Zeitchik, S. 2003. “German Ostralgie: Fondly recalling the bad old days”, *The New York Times Opinion*. http://www.nytimes.com/2003/10/07/opinion/07iht-edzeitchik_ed3_.html?pagewanted=. Accessed 31 October 2013