

Diversity in unity: practical unity and personal boundaries

Marya Schechtman

Received: 1 December 2006 / Accepted: 16 August 2007 / Published online: 27 September 2007
© Springer Science+Business Media B.V. 2007

Abstract In the spirit of the discussion in Daniel Kolak's *I Am You: The Metaphysical Foundation for Global Ethics*, I consider the way in which divisions that we usually think of as borders between distinct people occur within a single life. Starting with the dispute between constructionist and non-constructionist views of persons, I argue for a view that places the unity of persons in the dynamic generated by simultaneously taking a constructionist and non-constructionist view of oneself. In order to unify ourselves as agents we need to treat past and future selves as others, but to motivate this endeavor we need to think of ourselves as temporally extended agents, and so identify with past and future selves. Understanding this dynamic illuminates the structure of our agency and the unity of the self.

Keywords Personal identity · Self · Reductionism · Agency · Practical reasoning

In his remarkable and provocative new book *I Am You: The Metaphysical Foundation for Global Ethics*, Daniel Kolak urges us to take a close look at the divisions (psychological, physical, and phenomenological) that we usually take to form the boundaries between persons. If we do, he argues, we will notice that frequently these very same divisions occur within the life of a single person or reach across the lives of many people. In other words, while these divisions may indeed be *borders* they do not set the *boundaries* of a person, and lack the metaphysical and practical significance they are often accorded. Kolak makes this point within the context of arguments against both our ordinary concept of the individual person and some of the deflationary

M. Schechtman (✉)
Department of Philosophy, M/C 267, University of Illinois at Chicago,
1404 University Hall, Chicago IL 60607, USA
e-mail: marya@uic.edu

rejections of that concept that have been popular in the philosophical literature.¹ These arguments, he says, lay the ground for a global ethics. Kolak's analysis is amazingly comprehensive, and I will not engage directly with the details here. In the spirit of his investigation, however, I will look at one debate within personal identity that intersects with the issues he discusses, and demonstrates just how complicated the attempt to distinguish borders from boundaries can become.

A major division within theories of personal identity is the division between constructionist and non-constructionist accounts. Constructionist accounts view persons as constructs out of temporal parts, while non-constructionist accounts see these parts as abstractions from a unified person. On the former view the parts of a person have metaphysical priority over the whole; on the latter view the whole has priority over the parts. Richard Wollheim, a committed non-constructionist, describes the distinction this way:

A constructionist theory...holds that everything that needs to be said about the events that make up the life of a person...can be said without introducing a person who has them. On a constructionist theory, a person arrives on the scene only when there is a set of suitably interrelated events, and then the person is or is identical with that set. ...By contrast, a non-constructionist theory is a theory that maintains that no event in a person's life, even taken singly, can be adequately described without introducing the person who has it. There is always some person who integrally enters into any event that is of the kind that makes up the life as a person. (Wollheim 1984, p. 16)

At issue is the integrity of a person's life taken as an extended whole and the depth of our unity as persons.

The distinction between constructionist and non-constructionist views is important not only metaphysically, but also practically. While there is no logically necessitated connection between particular metaphysical and practical positions, constructionists typically view personal identity as a less practically significant relation than do non-constructionists. On the constructionist view the distinctions between the different temporal phases of our lives are deeper than we usually take them to be, and so our relation (from the standpoint of the present) to our "past and future selves"² is shallower than we usually take it to be. This means that when it comes to questions of prudential concern and distributive justice, the boundaries of an individual person might reasonably be given less weight than we generally give them. Non-constructionists, on the other hand, see the whole person as a deeply significant unit, and tend to hold that it must be viewed as such in our practical dealings.

¹ For instance that offered by Derek Parfit (1984).

² I use quotation marks here because I wish to use these terms—future self and past self—in a way that is neutral with respect to the question of whether they are, in any meaningful sense, different "selves" from the "present self." This is not easy to do, and the absence of language that is non-question-begging on the debate between constructionists and non-constructionists is, I think, philosophically significant in itself. For now, however, I will simply assert that this language is meant to be neutral and will omit quotation marks from now on.

In this paper I will describe a third way of thinking about personal identity that is neither fully constructionist nor fully non-constructionist, but rather incorporates both elements. On this view, we need simultaneously to view persons as deeply unified wholes, irreducible to their parts, and also as constructs out of distinct stages. The dynamic between these two views is, I suggest, part of what gives personhood its unique character. The position I describe takes off from Christine Korsgaard's non-constructionist view, which is offered as a response to Derek Parfit's constructionist psychological continuity theory. I therefore begin with a brief review of the relevant features of Parfit's account, followed by a description of the bare bones of Korsgaard's response and the features I will (and will not) be taking from it for my own view. Next I argue that if we look at how the non-constructionist unity described in that view is accomplished in practice, we will see that constructionist elements must be added without taking the non-constructionist elements away. Finally, I discuss some of the practical implications of my analysis.

1 Parfit's constructionist view

Derek Parfit's account of personal identity is a neo-Lockean psychological continuity theory. Although his version of the view has some peculiarities, for the most part it is fully representative of this approach. Psychological continuity theorists take from Locke the idea that diachronic personal identity should be defined not in terms of the persistence of any substance—material or immaterial—but rather in terms of psychological connections between the different temporal portions of a person's life. This basic insight is then altered and supplemented to avoid standard objections to Locke's original view. The resulting view holds that a person at one time is the same person as a person at another time just in case there exist between them overlapping chains of direct psychological connections (connections such as that between a memory and the experience remembered, between an intention and the action that carries it out, or between the different temporal parts of a persisting belief, desire, or value). Psychological continuity theorists may also specify the *number* of connections that must be present between the overlapping stages,³ or require that these connections be caused in some appropriate way,⁴ and often it is required that these connections hold uniquely—i.e. T' at t_2 can only be the same person as T at t_1 if the right kind of connections hold between T' and T , *and* these connections do not hold between T' and anyone else at t_1 .⁵

One of Parfit's many contributions to the debate on personal identity is to make explicit the constructionist character of these views, and the practical implications of that constructionism. He makes a distinction between reductionist and non-reductionist accounts of personal identity (a distinction which, for our purposes, is the same as the

³ Parfit (1984, p. 206) says it must be “at least half the number of connections that hold, over every day, in the lives of nearly every actual person.”

⁴ Parfit is fairly unique in arguing that any cause is acceptable. See Parfit (1984, pp. 207–208).

⁵ This is to avoid problems raised by cases of people “fissioning” into two continuants, each psychologically continuous with the original person. We need not concern ourselves with these issues for our purposes.

distinction between constructionist and non-constructionist theories), and says that the psychological continuity theory is reductionist because it claims:

- (1) that the fact of a person's identity over time just consists in the holding of certain more particular facts, and
- (2) 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, p. 210)

The fact of personal identity consists in the holding of connections between the contents of consciousness at different times. At each moment we have future selves and past selves, to whom we are related by chains of psychological connections, but from whom we are, in an important sense, distinct.

There are many powerful considerations supporting the kind of psychological account Parfit provides, and Parfit himself offers additional arguments. There are, however, also some serious objections to this view. One perennial complaint is that this (or any) kind of constructionist account cannot make sense of the practical significance of personal identity. We have a host of important practical relations to ourselves (throughout our lives) that we do not have to anyone else (e.g. egoistic and prudential concern are appropriate only for oneself; one is rightly held responsible only for one's own actions). But these practical relations seem to require that a person's life have a deep unity or integrity. They no longer seem rational if we are really constructs out of parts as Parfit describes. The basic sentiment behind this argument is nicely expressed in a line from Sidgwick quoted by Parfit (1984, p. 307): "Why," Sidgwick asks, "...should one part of a series of feelings...be more concerned with another part of the same series, any more than any other series?"

While several psychological continuity theorists have tried to respond to this objection by showing that the practical significance of identity *can* be justified on the psychological approach, Parfit offers a different and more radical response. The fact that personal identity is not a terribly deep or important relation may indeed be a surprising truth, but if the arguments show us that it *is* the truth, says Parfit, then we need to adjust our practical reasoning rather than our theory of personal identity. He thus accepts the deflation of identity's importance as a consequence of his metaphysical view, and adds that this fact can be salutary as well as disappointing. "When I believed that my existence was ... a [deep] further fact," he says, "I seemed imprisoned in myself. My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness" (Parfit 1984, p. 281). His recognition of the truth of reductionism, however, liberates him from the tunnel, making death and future suffering seem much less traumatic and important (Parfit 1984, pp. 281–282).

It is in response to this aspect of Parfit's view that Christine Korsgaard offers the non-constructionist theory from which I will develop my alternative view. Before moving on to Korsgaard's view, however, I want to pause a bit longer over Parfit's. Parfit defends his reductionism as an abstract metaphysical thesis to which he is inexorably led by the arguments, but concedes to his critics that the results are profoundly counterintuitive. I would like to suggest, however, that the constructionist approach is in fact not entirely counterintuitive. In some respects we do experience our lives as

a succession of stages with real divisions between them. Before closing this section, I would thus like to explore the “constructionist intuitions” at work in our thinking about ourselves.

The ways in which we think of our lives as segmented are, of course, complicated and various, and I certainly will not give an exhaustive account of them now. Here I will merely give the general idea by mentioning three significant divisions that we find in our lives—ways in which our relation to other portions of our lives (future and past selves) is relevantly like our relation to other people. The first are *divisions of consciousness*. There are all kinds of divisions of consciousness, and ways in which both neurobiologically and metaphysically we might say that the consciousness that makes up a person’s psychological life is divided into parts. My concern here, however, is with the ways we *experience* ourselves as divided. So the divisions I speak of are not divisions between instantaneous subjects, or 3–6s intervals of uninterrupted information-processing, but rather the divisions of consciousness that we experience as person-segmenting.

In addition to what we hold in immediate consciousness, there is a penumbra of information and experience that sits close to the surface. This includes autobiographical information that can be quickly and easily accessed (e.g. my current phone number and address, or the names of my current colleagues) and also recent experiences that are still very much alive emotionally (e.g. the hurtful remark made by my colleague last week or the unanticipated and much-appreciated surprise party my friends threw me a few months ago, or my impending move to another state). These can be contrasted with experiences from the more remote past that can be recalled only with great effort, or by using external aids such as photographs or journals or the memories of others, and also from experiences anticipated in the remote future, which are uncertain in character. It is not, of course, crystal clear where the line between the “present self” and “past” and “future” selves should be drawn in this regard. Nonetheless, most of us do experience a division between a somewhat extended present, which is readily accessible to consciousness, and more remote parts of our lives which, while perhaps accessible, are distant and misty. At the extreme of this continuum are the actions and experiences that no amount of therapy or effort can recall to consciousness—the actions and experiences that Locke said categorically are not one’s own.

Next there are *divisions of interests, concerns, and projects*. The immediate projects, occupations and priorities at one life phase are frequently quite different from those at another, different enough to make these other life phases seem alien and remote, as if they are the lives of another person. The life of a parent of small children in mid-career, for instance, will differ from her life as a college student in many respects. The mid-career parent will keep different hours, engage in different activities, and have different priorities, preoccupations, and desires. While there will undoubtedly be some continuities between these life phases, such a person may well look back at her college years and think that they seem like the life of another person.

Examples of this kind of division are common in the literature. Thomas Nagel (1970, 74 note), for instance, considers the story of a young man who leads a wild and carefree life but can anticipate that when he reaches middle age he will become conservative and stodgy. He discusses ways in which the young man may think of this middle-aged self as another person. Parfit himself offers the case of a 19th Century

Russian nobleman who is poised to inherit a great deal of land. He is currently sympathetic to the plight of the peasants, and plans to transfer the land to them when it comes into his possession. He fears, however, that he will be corrupted by wealth and change his ideals and hence his mind. He therefore draws up a document automatically diverting his land to the peasants, a document that can be revoked only with his wife's consent. He asks her not to provide that consent, no matter how much he begs, saying, "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" (Parfit 1984, p. 327). A second important way in which we experience ourselves as segmented, then, has to do with these kinds of differences in practical concerns at different times in our lives.

Finally, there are *divisions of will*. Certainly no one's will is entirely effective. Even in the present, the fact that I will something does not mean that I am actually going to do it—no matter how strongly I will myself to get out of bed and get ready to go to work I may find that I just don't move. Nonetheless, there seems to be an important difference between the limitations of will when I am willing to do something right now and when I am willing something for the more distant future. If I will something right now I have made a choice about what to do, a choice that I may or may not be able to execute. If I will something for the future, on the other hand, when the time comes to execute my will I need to once again endorse the choice; in some sense I must make the choice all over again. Will for the future, it seems, can fail in more and different ways than will in the present.

In some respects this point is just an extension of the two divisions already described. If I will some future action it may be that when the time comes to act I no longer remember what I have willed (due to divisions of consciousness), or have altered my values and interests in such a way that I no longer wish to do what I have willed (as Parfit's young Russian fears will happen when he inherits the land). In this way, the sense that willing for the future is a more tenuous matter than willing for the present depends on the kinds of segmentation I have already described. But the more general, formal, asymmetry between willing for the present and willing for the future, (i.e. the fact that what is willed for the present need, in some sense, be chosen only once while what is willed for the future must be chosen both now and then) also suggests that divisions of the will should be considered in their own right.

This sort of segmentation is a preoccupation of the Existentialists. Nietzsche opens the second treatise of *On the Genealogy of Morality* by asking the question: "to breed an animal that is *permitted to promise*—isn't this precisely the paradoxical task nature has set for itself with regard to man? Isn't this the true problem of man? (Nietzsche 1998, p. 35). To be permitted to promise, he says, we need to develop a real "*memory of the will*: so that a world of new strange things, circumstances, even acts of will may be placed without reservation between the original 'I want,' 'I will do,' and the actual discharge of the will, its *act*, without this long chain of the will breaking" (Nietzsche 1998, p. 36). Promising requires our being able to know what we will do in the future, and this is not something we can take for granted. Sartre also makes this kind of division one of the centerpieces of his work. He holds, famously, that we are radically free and that nothing we do now can fix our behavior in the future. Self-consciousness, on his view, inserts a "nothingness" into our being that makes us able at each moment to

break entirely with any past plans, resolutions, or habits. The problem of fixing future behavior for Sartre is not something that comes up just in anomalous circumstances—where there is a problem of memory, a special temptation, or disturbing information to be hidden—it is our very condition. “I await myself in the future,” he says, “where I ‘make an appointment with myself on the other side of that hour, or of that day, or of that month.’ Anguish is the fear of not finding myself at that appointment, of no longer even wishing to bring myself there” (Sartre 1992, p. 73).

There are, then, at least three different (though frequently interrelated) dimensions along which we often seem to ourselves to be segmented or constructed out of parts. I have not, of course, proven that any of these divisions actually exists or that if it does it has deep metaphysical or practical significance. This is not my purpose here. For now all I wish to do is articulate some of the ways in which we do seem to ourselves to be made up of parts, and in this way to show that, counterintuitive as the constructionist view of persons may be, it is not entirely without intuitive support. We will return to these constructionist intuitions later. For now, however, having articulated them, we will put them to one side and consider what a viable non-constructionist view might look like.

2 A non-constructionist alternative

Derek Parfit argues that a constructionist view of persons is true and that this means that there is no deep unity in a person’s life. He challenges those who believe in such a deep unity to explain what kind of relation between the different temporal parts of a person’s life could provide it. In “Personal Identity and Human Agency: A Kantian Response to Parfit” Christine Korsgaard (1989) takes up this challenge. The kind of unity we are looking for may be impossible to find if we look where Parfit looks, she says, but he is looking in the wrong place. In particular, Parfit assumes that the unity we need to define is a phenomenological unity or unity of conscious subject. Korsgaard argues that the unity found in the life of a person is not a unity of subject, but a unity of agent. “Suppose Parfit has established that there is no deep sense in which I am identical to the subject of experiences who will occupy my body in the future,” she says, “I will argue that I nevertheless have reasons for regarding myself as the same rational agent as the one who will occupy my body in the future. These reasons are not metaphysical, but practical” (Korsgaard 1989, p. 109).

A human life involves pursuing projects, plans and goals. To live a life one needs to be engaged in activities that unfold over time, and to undertake these activities one needs to think of oneself as a continuing self. Korsgaard says: “you may think of it this way: suppose that a succession of rational agents *do* occupy your body. I, the one who exists now, need the cooperation of the others, and they need mine, if together we are to have any kind of a *life*. The unity of our life is forced upon us...” (Korsgaard 1989, p. 113). The point is not just that actions that take a long time to complete require coordination, but more fundamentally that actions require reasons and reasons by their very nature reach beyond the present moment. “To ask why the present self should cooperate with the future ones is to assume that the present self has reasons with which it already identifies, and which are independent of those later selves” (p. 113).

But this is not what our reasons are usually like, she says, “and 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*. When the person is viewed as an agent, no clear content can be given to the idea of a merely present self” (pp. 113–114).

Korsgaard’s view is thus non-constructionist. One of the things a person is, she argues, is an agent, and to be an agent one must conceive of oneself as a persisting whole. Action is not possible if one understands oneself as a time-slice connected only tenuously to other time-slices. So to be an agent we must think of ourselves first as unified wholes. Even our current actions, taken singly, cannot be completely described without reference to the persisting person whose actions they are. Of course, this anti-constructionism comes with something of a change of topic. Korsgaard is not providing us with a deep, metaphysical unity that defines a person over time, and she readily admits that she is not. She denies, however, that this means that the unity she describes is not a real or important one. The pragmatic necessity of unifying ourselves as agents is genuine, and the unity that results from our conceiving of ourselves in this way is also genuine. Korsgaard thus offers us a view on which we create ourselves as agents by conceiving of ourselves as such. By thinking of ourselves as unified practical entities we become agents who cannot be understood merely as a construction out of present selves.

The view on which I will focus is not quite the same as Korsgaard’s, but it does incorporate some of its central features. In particular I offer a view according to which the unity of a person is a practical or pragmatic rather than a metaphysical unity, and a person’s conception of herself as a persisting self is capable of generating a genuine unity. The differences between my view and Korsgaard’s can be made clearer by considering an example, which will serve also to help demonstrate the plausibility of the general approach.

Leonard Shelby, the protagonist of the film *Memento*, suffers from a “condition” which leaves him unable to lay down any new long-term memories. This condition, we eventually glean, is the result of a brutal attack in which Leonard was assaulted and his wife raped and murdered (or at least, so Leonard thinks). Leonard retains normal memory of the time before the assault, but afterwards can remember things for only a short while. Every 7 min or so, he reawakens to his surroundings as if he is seeing them for the first time. In order to have any kind of life at all, Leonard must thus use indirect methods to communicate information to his future self. He writes future-Leonard notes, placing them where they are sure to be seen, and collects and labels Polaroid photographs. He tattoos especially crucial information on his skin. His most important tattoo tells him that “John G. killed and raped my wife.” Since the assault, Leonard’s life has focused almost exclusively on hunting down John G. to exact revenge. The tattoo informing him of the attack presents him with the substance of his quest, while his notes and polaroids bring him up to speed on the progress he has made so far, and help him find the resources to continue his pursuit. Leonard’s case is an especially crisp example of the kind of phenomenon Korsgaard is describing because the only kind of unity Leonard has is unity of agency. He is fragmented as an experiencing subject, lacking memory connections of any sort between the 7-min segments that make up his life. But there is a unity in his life nonetheless.

It is more than the fact that there is a single human animal bearing the name “Leonard” that makes us think of him as a single, persisting person. Leonard’s project of avenging his wife’s death pulls his life together; we see it (and hence Leonard) as a unified whole because it is a series of purposive activities that together constitute acting on a long term plan. The non-constructionist aspects of the agential view are also clear in this case. Leonard breaks naturally into 7-min segments during which his consciousness is more-or-less continuous. The goal of pursuing John G. cannot, however, sensibly be thought of as the goal of any one of these Leonard segments. For one thing, if these are really taken to be distinct entities, the assault and loss being avenged did not happen to the present-Leonard, but only to some past-Leonard to whom he is connected in certain ways. And since the project of finding John G. and bringing him to justice is one that will take considerable time, revenge cannot reasonably be the project of present-Leonard understood as present-Leonard. There is no way that present-Leonard could carry out this project. In order for his vendetta as he understands it to make sense, Leonard must be thought of as a unified whole—someone who suffered the violent loss of his memory and his wife, lives with the consequences now, and seeks to bring the assailant to justice sometime in the future. The very possibility of the activity that unifies his life depends upon taking a non-constructionist view of Leonard.

Of course, Parfit would happily acknowledge that this activity is only *reasonable* if we think of Leonard as a unified whole. His point is that Leonard is in fact not a unified whole, no matter how we think of him, and so this activity is in fact not reasonable. Indeed, if any present-Leonard could recognize this it might well be liberating in just the way that certain Buddhist traditions see the recognition of no-self as liberating. Kosgaard, you will recall anticipates this Parfitian response to her view, but argues that the practical unity of an agent is no less real for not being a metaphysical unity. Leonard’s case helps us to understand this claim by giving us a way of thinking about what this kind of unity amounts to. If we look carefully at Leonard, we can see that his quest for justice is just about the *only* thing holding his life together. He is so fragmented by his memory condition, that the only way to give his life a semblance of order is to organize it around this clear and well-developed project. It is in the service of this project that Leonard undertakes the writing of notes and collecting of polaroids that give his life coherence. If he did not have a unifying project of this sort things would be much different, and arguably he would dissolve into a sequence of distinct 7-min segments.

Indeed, the recognition that his quest for justice is central to the integrity of his self seems to be one of the forces driving Leonard to undertake it. In a very real sense, his life depends upon it. In an exchange with Teddy, a sometime sidekick who aids him in his search for John G., Leonard says that he is going to go after the man who killed his wife, caused his injury, and “took away his ability to live.” Teddy checks Leonard’s pulse and reports: “You’re living,” to which Leonard responds, “Only for revenge.” While this reply is meant metaphorically, implying that the only *purpose* he has for living is revenge, Teddy’s check of Leonard’s pulse suggests the more literal reading it also allows. Leonard, recognizes that the only way that he can have a unified life is to organize all of his activity around a single, all-consuming goal. If he can keep coming back to that goal every time he “awakens” to his surroundings he can

keep a thread going that will hold his life weakly together—something around which everything else can be centered.

This aspect of Leonard's condition is made even clearer in his discussion of the case of Sammy Jankis, a man with a similar affliction who Leonard encountered in his pre-accident job as an insurance investigator. Leonard explains: "Sammy Jankis wrote himself endless notes. But he'd get mixed up. I've got a more graceful solution to the memory problem. I'm disciplined and organized. I use habit and routine to make my life possible. Sammy had no drive. No reason to make it work." He then looks at the tattoo announcing his wife's murder and adds, "Me? Yeah, I got a reason." In his depleted condition the only unity Leonard can have, the only way he can continue as a "self," however tenuously, is to unify himself around this goal.

Leonard's dependence on this project is nowhere clearer than at the end of the film. In a surprising twist Leonard is presented (by Teddy) with evidence that his whole quest is misconceived—the John G. he seeks either never existed, or exists no more, and the assault he is seeking to avenge did not happen as he has been describing it to himself. Rather than recording this information for the benefit of future-Leonard as one might expect, Leonard takes pains to make sure that it will not survive his next loss of memory. He arranges his communications to the future in such a way as to ensure that future-Leonard will take up the (necessarily futile) quest to find John G. If the sole purpose of the quest was to exact justice, Leonard should let his future self know that there is no longer a need to do so. And if the only purpose is to be able to think well of himself (the real story does not reflect well on him) he could communicate to future-Leonard that justice was achieved and that he was the hero who achieved it. The fact that he protects his project as he does implies his recognition that without it he would genuinely fall apart—there would be no "life of Leonard" taken as a whole if there was no extended quest for revenge.

If we accept Leonard's assessment of the role his quest plays in unifying his life, and attend to the contrast with Sammy Jenkis,⁶ who had no such quest and devolved into chaos, we can see that there are real and significant effects of Leonard's conception of himself as a unified, irreducible whole. Thinking about himself in this way results in a particular organization of his life that allows him to carry out extended tasks and enjoy a coherence in his existence that would otherwise be impossible. By thinking of himself as an extended agent Leonard makes himself such an agent, an agent he would not be if he did not have that self-conception. So while Leonard's unity is not a pre-existing metaphysical fact, it is nonetheless a genuine and significant fact, and it is something he accomplishes through a non-constructionist understanding of himself as an agent.

This fictional case can, I think, shed light on the kind of practical unity that can be achieved by taking a non-constructionist view of oneself. In order to coordinate and organize his behavior over time, Leonard needs to think of himself as a persisting self on a mission. By thinking of himself in this way, however, he does become capable

⁶ There is a legitimate question to be raised about my interpretation of the film here, since it is by no means clear that Leonard's account of Sammy Jenkis is reliable. But since this is only an illustration we can ignore this complication if it seems plausible that without a project Leonard's would be considerably less coordinated than it is.

of making things happen in the world over time in a way he would not be able to do if he did not think of himself in this way. It thus becomes reasonable for us, too, to think of Leonard as a persisting self on a mission—a mission that can only be understood as the project of an extended self. Of course, Leonard's case is both fictional and pathological, so it cannot be automatically assumed that it will yield insights into our condition. Having seen this phenomenon in Leonard's case, however, we should be able to read it back into more ordinary lives. It is not an unfamiliar thought that a person needs activities to provide shape and unity to her life, and that a life with no plans, projects, or goals would quickly lose coherence. One might think that the kind of continuity provided by memory could unify a life without agential unity, but as a matter of psychological fact we probably need projects and activities to help us maintain memory connections at least as much as we need memory to help us engage in long-term projects. In any event, we are considering this view as a response to Parfit's claim that there is no meaningful unity in *our* lives—the lives of people with ordinary memory connections. If we find that practical considerations can provide a meaningful unity in Leonard's case, surely they can in our own. We can also explain why the unity found in Leonard's life seems less profound than that found in more ordinary circumstances. It is because his cognitive deficits vastly limit the range and richness of activities he can carry out. Rather than the complex, interrelated activities that make up an ordinary life, Leonard must make do with a central quest and the activities necessary to support it.

We thus arrive at a view that is non-constructionist in a somewhat qualified way, but non-constructionist nonetheless. The picture I am offering is one according to which a person can only behave in the organized way that is characteristic of a person's life by thinking of himself as an extended self with long-term interests. Leonard's interest in future Leonard makes sense only if he identifies himself as the same person as future Leonard. But he does have interests in unification, and so in identifying with the extended self. It is with respect to these issues, however, that my view differs from Korsgaard's, and the ways in which it differs are intimately connected with the way in which my view is not only non-constructionist but also constructionist. I turn to these issues next.

3 Constructionist features of the agential account

I have argued that Leonard unifies himself as an agent by viewing himself as an extended self with interests that stretch over time. Korsgaard might not be as convinced of Leonard's success at such unification as I am.⁷ Her conception of agency is linked intrinsically to autonomy and hence to the authority of reason. In order to truly be an agent, on Korsgaard's view, one needs to be responsive to reasons in a very robust sense, which Leonard obviously is not. Leonard's bit of self-deception at the end of the film is unabashedly instrumental, and he simply has no interest in focusing on the fact that his quest to find John G. is incoherent in his deliberations about what to do. Because of this, Korsgaard might well insist that that he is not really a fully

⁷ I am grateful to Carla Bagnoli and Edward Hinchman for making me see the importance of this point.

unified agent. His actions are heteronymous and hence not fully his own. She might deny also that he genuinely has an interest in hunting down John G., since he clearly has no real reasons to do so.

The sense of agency and of interests that I have in mind is weaker and more basic than Korsgaard's. I do not wish to insist that Leonard is autonomous or that his actions somehow carry authority. When I say that he is an agent I mean only that he is able to intentionally make things happen in the world. When I say he is unified as an agent, I mean that he is able to coordinate his activities over time, and so can make things happen in the world that he could not without such coordination. Finally, when I say that he has interests as an extended self I mean something more thoroughly internalist than Korsgaard has in mind. It is Leonard's *believing* himself to have interests that extend over time that allows him to coordinate his life so that it looks like a life. Thinking of himself as having interests of this sort does have real effects whether the projects he takes on are rationally defensible or not. The unity Leonard achieves by taking himself to have such interests is, moreover, enough for others to recognize him as an extended whole.⁸ While I do not deny that there is much to be desired in Leonard's form of agency, then, I depart from Korsgaard by seeing him as fundamentally successful at unifying himself, and generating long-term interests, despite his unresponsiveness to the dictates of rationality.⁹

This difference is important to understanding the way in which my view is constructionist as well as non-constructionist. The picture Korsgaard offers is one where ideally we are driven by the force of reason to a true unity, in which our actions come from our whole selves and not from one part or another of a divided self. On my view, however, the self remains divided. The unified self does not *replace* the series of temporally limited selves who coordinate their efforts, but sits next to it. Put somewhat differently, on my view a person continues to see herself as a series of distinct, temporally limited selves as well as seeing herself as an extended whole. This is, in fact, what the constructionist intuitions of section one reveal. Both perspectives, that of the extended self and of the temporally limited self are, I argue, co-present. And they can often yield conflicting prescriptions about what to do. Efforts at coordination and unification are thus not so much ways of eradicating the divisions that make us seem, from one perspective, like series of distinct selves, but rather ways of getting the perspective of the extended self to prevail.

It will be easier to see what this means by looking at a more concrete example. One way in which this dual perspective manifests itself is in the mechanisms that must be employed to bring about coordination. Frequently, I argue, in order to achieve unity

⁸ Natalie, one of the characters Leonard encounters, at one point takes revenge on him for an earlier transgression by seeking to thwart his pursuit of John G. When she wants to help him, she gets him some crucial information. Clearly she sees an extended Leonard who can be helped or hurt by aiding or interfering with this extended project.

⁹ I am actually not clear enough on Korsgaard's position to know just how far I am departing from it. Certainly she would not be happy with the methods through which Leonard unifies himself, as the discussion to come will make abundantly clear. Still, I am not certain whether she would say that he fails utterly at self-unification or whether she would say that insofar as he tries to pretend he is acting from reasons he is a unified agent, but a defective one. Either way, however, her demands on unity and agency are far stronger than mine.

we need to recognize ourselves as genuinely divided—and in an ongoing way that is not resolved by simply forming long-term plans. Looking closely, for instance, at *how* Leonard accomplishes this unification we see that at every step he views himself not only as extended-Leonard, but also as a sequence of distinct agents, each with his own will and interests. Moreover, he must do so in order to act on the extended project that unifies him.

One way in which Leonard treats his past and future selves as distinct agents is in his method of communicating with them. The notes, photos and instructions he leaves are the sorts of information we might provide for a colleague with whom we are collaborating—“this is what I’ve got so far and here’s where I think you should go next” is essentially what he tells his future self at each juncture. The extent to which Leonard is required to do this is, of course, a direct result of his cognitive deficits, but to a lesser extent we all sometimes communicate with ourselves in this way—as evidenced by planners, PDA’s, shopping lists and marginal notes. There are, however, also more subtle ways in which Leonard treats his past and future selves as distinct agents, and these are more germane to our present discussion. For one thing, Leonard often leaves future-Leonard more than just instructions about what to do next; he leaves him *reasons* for doing what he instructs. The tattoo he puts on his chest does not say “hunt down and kill John G.”; it says “John G. killed and raped my wife.” There is a nice ambiguity here about the referent of “my”—is it the person who commissioned the tattoo, the person who is reading the tattoo, or is it meant to imply that they are the same? This ambiguity fits nicely into my larger view, but for the time being it is enough to see that, although the reason provided will be a different kind of reason depending on the referent of the first-person pronoun, it is a reason in any case. Whatever Leonard’s relation to future-Leonard, it is not so intimate that he can simply tell him what to do “no questions asked.” Leonard’s conception of his future self as another agent is most clearly seen, however, in his behavior at the end of the film. There Leonard fears that future-Leonard, knowing the truth, would give up on the project that has been his life-line. Future-Leonard, in other words, is viewed as someone who might act against present-Leonard’s interests and wishes, and so as someone who must be manipulated into compliance.

To understand how complex this behavior is, however, it must be appreciated that the interest present-Leonard has in making sure future-Leonard signs on to the project of hunting down John G. is an interest of the temporally extended Leonard and, in a fundamental sense, an interest in unification. Leonard is desperate to get future-Leonard to behave in a particular way because if he does not, the glue that holds him together as a persisting agent will dissolve. In order to constitute himself as a continuing agent, then, Leonard has to view different temporal portions of his life as potential adversaries and strategize about how to get them to participate in a common project. In order to unify himself to live an integrated life, Leonard must view himself as a persisting agent, with irreducibly extended interests, and at the same time as a series of distinct agents whose cooperation must be secured to make the long-term project work.

This complex interaction between a conception of oneself as extended and as constructed out of distinct parts takes a particularly striking form in Leonard’s case, but it is also a feature of our own lives. The very notion of prudence suggests that we have

interests as extended selves and as I have already suggested we, like Leonard, need to take these interests seriously as *our* interests if we are to unify ourselves as agents. But the *need* for a concept like prudence also suggests that we cannot *simply* think of ourselves as extended beings and be done with it. While we do experience ourselves as extended agents with interests that belong to ourselves as enduring wholes, we also think of ourselves as temporally limited subjects with interests specific to temporal portions of our lives. Someone's interest in maintaining a healthy diet, or staying sober, or doing her work in a timely fashion is an interest she experiences when she understands herself as an extended whole—we cannot adequately describe such interests without reference to the extended person whose interests they are. But she may also know full well that there will be a time (maybe it's even now) where she will also have a strong desire to eat something full of trans fats, or get totally drunk, or put off work in the face of a tight deadline.

The sort of problem Parfit addresses—the problem of why someone's present self should make sacrifices for the well-being of her future self—expresses the way in which desires at a time can be experienced as interests of a present self, interests that may be in competition with those of the extended self. The need to find ways to act prudently is, among other things, the need to find ways to act on the interests of the extended self. I only have a prudential reason to resist the trans fat-laden snack or the drink or the work break if I think of the interests of the extended self as my own interests. To the extent there is a conflict about what to do, however, I do not identify *only* with the extended self, but also with my present self, a self with interests of its own.

The kinds of circumstances that give rise to problems of prudence are often thought of as arising from conflicts between present and future desires. I favor a description of them as conflicts between a temporally extended and temporally limited perspective on oneself. To the extent that one views oneself as a temporally extended whole, what we typically consider the prudential course of action will seem more desirable, but to the extent that one thinks of oneself as a temporally limited self, the more immediate desires will be more compelling. When I want greasy snacks despite a resolution to eat only healthy foods the conflict is not between a present self who wants treats and a future self who wants health; it is between a temporally limited present self who wants treats and an extended self who experiences the desire for treats but wants all-things-considered to eat healthy food.

The problem of how to get oneself to act prudently is thus one version of the problem of how to unify the self. To act prudently is to act according to the kinds of extended interests that unify and coordinate our behavior over time rather than according to the interests of the present self. My claim is that this dual identification is a ubiquitous feature of our lives. To act in the ways that unify us we need to act, for the most part, on the interests of the extended self; to *motivate* ourselves to act in this way, we need to identify with our extended selves. When I am facing the temptation of trans fat-soaked goodies I may well feel the conflict between my identification with my present self—who wants to eat the goodies—and the extended self I also am—the self who wants to keep to the healthy eating plan. From a present perspective where I identify with my long-term interests, I have reasons to find a way to encourage a similar identification in a future self who will also crave snacks. I need to do what I

can to make sure that in the future the conflict between the extended perspective and the temporally limited one is resolved in favor of the former rather than the latter. The mechanisms by which we facilitate such identification, like the mechanisms by which Leonard facilitates such identification, frequently involve treating our future or past selves as distinct agents who need to be managed or won over.

Exploring these mechanisms and the way in which they presuppose division is an enormous task, and here I can do no more than mention some familiar strategies for maintaining the extended perspective. One is the institution of habits, routines, or environmental reinforcements that make the extended perspective the default point-of-view. Sartre, who as already mentioned sees us as radically free, points out that many of our possibilities are recognized only as they are being actualized. The possibility of writing, he says, is his possibility as he sits there with pen in hand, at his desk, at the usual hour. (Sartre 1992, p. 80) These habits stack the odds in favor of choosing the possibility that will continue the extended self. By making certain activities rote we, as it were, put blinders on our future selves, relegating interests that would lead them elsewhere to the background.

Sartre himself sees these habits and routines as defense mechanisms that we use to help us flee from the fact of our freedom. By establishing these habits we distract ourselves from our possibilities and pretend that the matter of what we are is settled once and for all, and that we do not need to re-choose at each moment. He sees this as a form of Bad Faith.¹⁰ Others might see these habits as useful tools for maintaining regularity in one's life. Either way there is agreement about the fact that establishing routines can impact the saliency of extended interests in relation to temporally specific ones, and so act to aid unification. This is the difference, after all, that Leonard sees as distinguishing his own case from that of Sammy Jenkis. His need for vengeance gives him one *reason* to unify himself, but his habits and systems are his *means* of doing so. They are, however, a means that involves a certain manipulation of his future self, and this indicates a less-than-complete identification with that self.

A closely related mechanism for giving the advantage to the viewpoint of the extended self is the establishment of explicit policies or rules for future conduct. These policies again provide a default answer to the question of what to do in a given situation, and so tip the balance in favor of the extended perspective. This approach, too, is meant to “break off discussion” and lessen the chance of rethinking our decisions. The policies are meant to keep the future self doing what we—qua extended beings—have decided to do rather than doing what the temporally limited self might decide to do from its own perspective. Michael Bratman has discussed this phenomenon quite extensively. Plans and policies, he says, “play an important role in the constitution and support of continuities and connections characteristic of the identity of the agent over time” (2000, p. 47). He goes on to add:

¹⁰ Nietzsche has his own answer to the question of how we come to be permitted to promise. The memory of the will required comes from the horrible punishments inflicted for unpaid debts. Over time we internalize this punishment and develop a conscience—a mechanism for self-punishment. So here, too, the mechanism of unity and regularity involves viewing other parts of oneself as someone else, someone who metes out punishment.

Indeed, this is part of what plans and policies are for. Such plans and policies have as their function the support of cross-temporal organization and coordination of action in part by inducing cross-temporal connections...and continuities.... A point of having plans and policies is to induce organization and coordination by way of such continuities and connections. (Bratman 2000, p. 47)

This represents another way of trying to fix future behavior and forestall the possibility of too much reflection. We do not want our future selves, except in unusual circumstances, to think too hard about whether they should act in accordance with the wishes of the extended self.¹¹

This kind of strategy for self-control is also described by psychiatrist George Ainslie (2001), who sees the stability of the will as the result of bargaining among our temporally limited selves. Starting from R.J. Herrnstein's "matching law," which says that "rewards tend to be chosen in direct proportion to their size and frequency of occurrence and in inverse proportion to their delay" (Ainslie 2001, p. 35). Ainslie argues for "hyperbolic discounting" in opposition to utility theorists who assume that "people discount future utility the way banks do: by subtracting a constant proportion of the utility there would be at any given delay for every additional unit of delay" (p. 35). To support this analysis, he amasses empirical evidence that the discount curve exhibited in human behavior actually is hyperbolic—that "people devalue a given future event at different rates, depending on how far away it is" (Ainslie 2001, front matter). "This phenomenon," he says, "means that our preferences are inherently unstable and entails our present selves being pitted against what we can expect our future selves to want" (Ainslie 2001, front matter).

Because of this, we need to find ways to get our future selves, who will have their own preferences, to find the perspective of the extended self salient and to act on it. Ainslie says that we do this by making "bundled choices." In addition to making individual choices we can make choices about groups of similar decisions.¹² I do not only make a decision, for instance, about whether to order dessert or have that third drink *now*, but also about what I want to do on the occasions where I am given the chance to have third drinks or desserts, taken as a group. My bundled choices represent the interests of the extended self—how I want to conduct my life as a whole—while the individual choices are choices of selves-at-a-time. When confronted with a dessert option, I have an immediate impulse to indulge. But to the extent that I also have preferences about the *bundle* of choices; I want to abstain. There is, here, a conflict between the two practical perspectives we hold, with the immediate, short-term interest gaining strength from temporal proximity.

Ainslie asks, as we have, "If a person is a population of processes that have grown in the same mind through the selective action of reward, what factors, if any, impose unity on this population?" (2001, p. 40). His answer is that we strike a bargain among the

¹¹ Of course, there must be some flexibility, since the extended self must be able to revise plans, at least sometimes, in response to new information.

¹² Bratman criticizes some of the details of Ainslie's view, but agrees with his basic assertion that we can choose behaviors in bunches. His criticisms have to do with its conception of when these choices are reason giving. On the aspects of the view relevant here, he is in substantial agreement with Ainslie. See Bratman (1999).

different temporal portions of our lives based on the mutual goals they have in virtue of their bundled choices. On his view a person's relation to the other temporal portions of her life is what bargaining theorists call a relationship of "limited warfare"—a relationship in which the participants have some overlapping and some competing interests. This relationship has the incentive structure of a prisoner's dilemma (Ainslie 2001, p. 90). Each member of the population that makes up the different temporal segments of my life (or, at least, many such segments) might have an interest in eating the dessert in front of her. But insofar as the practical perspective of the extended self is present in each case as well, they all share an interest in my sticking, for the most part, to my diet. For each, however, the incentive to stick to the diet depends upon others also sticking to it. If I am going to start gorging on chocolate cake tomorrow and continue for the rest of my life there is no point in resisting tonight. If, on the other hand, my firm resolve tonight sets a precedent that helps to encourage future acts of resistance the incentive to resist becomes greater. "Will," says Ainslie, "is a bargaining situation, not an organ" (2000, p. 90). Extending one's will to future decisions is not exercising a capacity that somehow reaches forward to fix future outcomes; it is a stable situation of mutual cooperation that serves as a solution to an interpersonal prisoner's dilemma. In other words, to unify our lives into extended wholes we need to think also of our future selves as others with whom we must strike a bargain.

Of course, our attempts to lay down rules and fix the environment do not always succeed in making our future selves act on the perspective of the extended self. Sometimes the desires and impulses of the temporally specific self are so powerful that it is hard to get the extended perspective to prevail. This is most clearly true in cases of addiction or compulsion, but can occur in any situation we might call weakness of will or imprudence. When we anticipate that this is likely to happen, we often (acting in the interests of the extended self) take more drastic and coercive measures to keep our future selves on track, just as Leonard does at the end of *Memento*. These strategies can range from leaving the car at home when taking off to the New Year's party, to asking a friend to hold on to one's cigarettes and provide them only at the rate of one per hour, to signing a legally binding document (like Parfit's Russian nobleman), to having one's crew tie one to the mast.

The varieties of self-binding strategies have been usefully studied, among other places, in two books by Jan Elster whose titles make reference to the paradigm case of Ulysses. In *Ulysses Unbound* he says that individuals protect themselves against lapses of will "by removing certain options from the feasible set, by making them more costly or available only with a delay, and by insulating themselves from knowledge about their existence" (2000, p. 1). Whether relying on the efficacy of personal habits and rules or seeking more directly to coerce or manipulate our future selves, we are taking them seriously as distinct agents whose cooperation must be secured, and in this way we continue to conceive of ourselves as a series of independent agents, even as we unify ourselves as extended selves. In order to become unified selves, we need not only *think* of ourselves as enduring beings with extended interests, but also to act on those interests. But if we are going to do that we must also recognize ourselves as sequences of more temporally limited selves with interests of their own, and employ strategies for bringing our future selves on board, encouraging them also to view themselves fundamentally as extended selves.

4 Conclusion

I have described a view that is, in a relevant sense, both constructionist and non-constructionist. It does not claim that as a matter of metaphysical fact we are both irreducible wholes and constructs out of temporal parts—it is hard to know what such a claim would even mean. But it does say (a) that in order to achieve practical unity we must view ourselves simultaneously as irreducible wholes and as constructs out of parts, (b) that taking this view of ourselves allows us to achieve a real, practical unity that we could not without doing so, and (c) that despite achieving this unity there is an important sense in which our past, present, and future selves remain distinct practical units. It is part of the structure of our lives that we take a dual perspective on ourselves. The view I have described is thus a view of practical unity that is both constructionist and non-constructionist.

Behind the controversy between constructionists and non-constructionists is a controversy about the depth and importance of the distinction between individual people. Parfit argues that because of the metaphysical fact of reducibility a person's relation to other temporal portions of her own self is not significantly deeper than her relation to the lives of other people. This leads to the liberating thoughts described earlier, and to a consequentialist ethics that does not need to worry about respecting individual persons as units of distributive justice. Non-constructionists, on the other hand, typically see the boundaries between individual persons as holding immense moral and practical significance. The constructionist/non-constructionist view I have offered thus suggests a third way of thinking about the practical significance of personhood, according to which both the temporally limited self and the extended self should be viewed as holding practical significance.

There is much work to be done before it is possible to say just what the practical implications of this mixed view amount to. Here I have emphasized the value of unifying ourselves, and the interests that we have only as extended selves. In this discussion I have emphasized the practical importance of respecting the distinctions between temporally limited selves as a means to unifying ourselves into extended selves. But if we take seriously the constructionist intuitions, there is also reason to think that the interests of the temporally limited selves should at least sometimes be taken as legitimate interests as well. A life so planned and rigid that the interests peculiar to individual temporal selves are never indulged might be just as practically flawed as one that is chaotic from the lack of a coordinated attention to the projects of the extended self. Moreover, “the practical implications of identity” are not monolithic. The relative importance of the extended and temporally limited perspectives, and the ways in which their claims should be balanced, may be different for moral responsibility, self-interested concern, interpersonal relationships and concerns about survival.

All of this needs to be investigated before the practical implications of the view I have described can be understood. Drawing out these implications is, obviously, a project for another day (and a more extended self). Doing so promises, however, to shed light on some puzzling elements of everyday experience. We do, I think, experience ourselves as both extended and divided, and it would be most welcome if philosophical reflection could help us see how this is possible and what to do about it. At the very

least, the possibility of a mixed constructionist/non-constructionist view reminds us that the question of what is a border and what is a boundary is never a simple one.

Acknowledgements A version of this paper was presented to the philosophy department at the University of Milwaukee at Wisconsin. I benefited immensely from the vigorous discussion there. I am also grateful for conversations with Tony Laden, and extremely helpful and insightful editorial comments on an earlier version from Troy Catterson.

References

- Ainslie, G. (2001). *Breakdown of will*. Cambridge: Cambridge University Press.
- Bratman, M. (1999). Planning and temptation. In *Faces of intention* (pp. 35–57). Cambridge: Cambridge University Press.
- Bratman, M. (2000). Reflection, planning, and temporally extended agency. *Philosophical Review*, 109(1), 35–61.
- Elster, J. (2000). *Ulysses unbound*. Cambridge: Cambridge University Press.
- Korsgaard, C. (1989). Personal identity and human agency: A Kantian response to parfit. *Philosophy and Public Affairs*, 18(2), 101–132.
- Nagel, T. (1970). *The possibility of altruism*. Princeton: Princeton University Press.
- Nietzsche, F. (1998). *On the genealogy of morality* (trans: Clark, M. & Swensen, A. J.). Indianapolis: Hackett.
- Parfit, D. (1984). *Reasons and persons*. Oxford: Clarendon Press.
- Sartre, J. P. (1992). *Being and nothingness* (trans: Barnes, H.). New York: Washington Square Press.
- Wollheim, R. (1984). *The thread of life*. Cambridge: Harvard University Press.