

P.J.J. Delaere
Practical Identity

Practical Identity
An essay on personhood, autonomy, and pathology

Praktische identiteit
Een essay over personen, autonomie en pathologie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
donderdag 2 december 2010 om 11.30 uur

Patrick Jérôme Jules Delaere

geboren te Izegem (België)



Promotiecommissie

Promotoren:

Prof.dr. G.A. den Hartogh

Prof.dr. I.A.M. Robeyns

Overige leden:

Prof.dr. I.D. de Beaufort

Prof.dr. B. Roessler

Prof.dr. J.J. Vromen

*To Helga, Vincent, and Ilse, without whom this book would not be.
And to Michaëla, without whom Ilse, Vincent, and Helga would
not be.*

CONTENTS

<i>Acknowledgments</i>	9
Preface	II
1 Identity and practical reason	19
1.1 Accounts of human agency: a state of play	19
1.2 Actions and persons	23
1.3 Practical identities	27
1.4 Defective action	29
1.5 Unifying agency	32
1.6 The normativity of the principles of practical reason	36
1.7 Some demystification	41
1.8 First findings	50
2 Narrative self-constitution	55
2.1 Narrative self-constitution	55
2.2 Constraints on self-constituting narratives	58
2.3 The missing ingredient	64
2.4 The feeling of what happens	70
2.5 Basic narratives	78
2.6 Trouble for narrativists	82
2.7 Interim conclusions	90
3 Personal autonomy	95
3.1 Introduction	95
3.2 Harry Frankfurt on caring and loving	103
3.3 Some teachings and tensions	113
3.4 Identification and coherence	125
3.5 Bratman's planning theory	132
3.6 The overall picture: a narrative-based account of personal autonomy	140
3.7 Final conclusions	152

4 The fragility of personhood	159
4.1 Introduction	159
4.2 Recurrent psychosis	161
4.3 Ulysses contracts	166
4.4 Psychosis and the someone else problem	170
4.5 Taking coherence seriously	179
4.6 Enter the naysayers	193
4.7 Advance directives and the severely demented	199
4.8 Autonomy at a distance: some second thoughts	213
4.9 The patient's keeper	227
4.10 Pre-emptive suicide as the triumph of autonomy?	238
4.11 Taking stock	247
Summary	253
Bibliography	265
<i>Nederlandse samenvatting</i>	279
<i>Curriculum Vitae</i>	295

ACKNOWLEDGMENTS

Taking the cyclist as a metaphor for doing philosophy, I am much obliged to Theo van Willigenburg who taught me to ride a bicycle, and to Govert den Hartogh who helped me to keep track.

My heartfelt thanks to Lucian Freud for his kind permission to reproduce one of his paintings – *Interior with Hand Mirror* (1967) – on the cover.

PREFACE

Mancher Psychiater hat ausgesprochen, er wolle sich nicht mit Philosophie belasten, seine Wissenschaft habe mit Philosophie nichts zu tun. (...) Aber die Ausschaltung der Philosophie wird trotzdem für die Psychiatrie verhängnisvoll.

K. Jaspers, *Allgemeine Psychopathologie*, 643

In the last decades cross-disciplinary work between philosophy and psychiatry has steadily been gaining momentum. The last few years have witnessed an explosion of what G. Graham and L. Stevens have called ‘philosophical psychopathology’ (1994). This dissertation will link some serious philosophy with psychiatric practice. The starting point was a *cri de coeur* by Bill Fulford that in psychiatry a new paradigm of integration is needed (2000). Following Fulford, two dominant types of psychiatry were crossing the psychiatric landscape as a cart track during the twentieth century: the biological and the hermeneutic (or constructivist) approach. Anti-psychiatrist T. Szasz was among those who led the way in claiming that mental illnesses are not brain diseases but ‘problems of living’. At the heart of Szasz’ project was his conviction that psychiatric diagnosis in essence is an evaluative project (1961, *The Myth of Mental Illness*). The current strongly biological orthodoxy in psychiatry takes another line: modern brain science will unravel the mysteries of the mind. It will reveal the causal disease processes underlying mental disorders. What patients need is new effective medical treatments instead of existential reorientation. Both types of psychiatry lead to what John Locke would have called ‘enthusiasms’, and try to consume one another. Fulford found ‘that neither mind-

only accounts of psychopathology, characteristic of some forms of anti-psychiatry, nor brain-only accounts, characteristic of the more triumphalist among biological psychiatrists, are sufficient' (2000: 68). An adequate account of psychopathological cases requires both.

Kenneth S. Kendler, M. D., in a programmed essay *Toward a Philosophical Structure for Psychiatry*¹ devised a thought experiment to clarify this point: 'Imagine that there are 15 discrete levels, with the mind-brain system between DNA on one hand and the clinical manifestations of schizophrenia on the other. Researcher 1 is conducting linkage and association studies that attempt to directly relate levels 1 and 15 but would provide no insight into the intervening levels. Researcher 2 is trying to understand, at a basic molecular level, the actions of a putative altered gene transcript, thereby trying to move from level 1 to level 2 or 3. Meanwhile, researcher 3 is seeking to understand the neuropsychological deficits in schizophrenia, trying to clarify the link between levels 13 and 15. Although biological reductionists might declare the work of researcher 2 to be more "scientific" and valuable because it is more basic, I hope that this thought experiment makes it clear that we can make no such judgments a priori. There are many links in the chain, and their ultimate value and scientific fruitfulness are unlikely to bear any strong relationship with where on the causal chain (or, more realistically, network) they sit' (438). Kendler calls for explanatory pluralism. We should stop searching for big, simple explanations and developing 'grand theory'. Rather, we should settle for a rigorous integration of multiple disciplines and perspectives on psychiatric disorder. Philosophy is one of the engaged disciplines.

1 Am J Psychiatry 162 (3) March 2005: 433-440.

Philosophers can be found on each of the 15 floors, and one of their main contributions – I think – is that they can help to keep the conceptual house in order. Philosophers can be the window cleaners of the house of research. Philosophers can also challenge the dualism inherent in seeing mental illness as physical illness or as life problem, and propose new ways of integration. To this enterprise I want to contribute. And, like Fulford, I think that a fruitful way of getting started is to examine the notion of agency (Fulford 2000: 65). Here is why. Psychiatry deals with fundamental questions of what it means to be human. It is not just a brain with which psychiatrists are concerned, not even just a mind, but a real person. The relevant unit of analysis in psychopathology is persons, whatever insights may be derived from studies of their component parts and systems. In current practical philosophy, agency is a matter of the highest concern and philosophers are trying to capture personhood in terms of agency and practical reason. Simultaneously, recent work in psychopathology suggests that there are aspects both of our general concepts of disorder (Nordenfelt, 1987) and of specific areas of psychopathology (e.g. delusion and thought insertion in Stephens and Graham, 2000; or akrasia in Campbell, 2000), which are more readily understood in terms of disturbances of action, that is, of the characteristic mode of activity of autonomous agents. A lot of pathology seems to be connected with the experience of failure of ‘ordinary doing’. So here is common ground that I will try to explore. This is my plan.

In the first part of my dissertation I will make it my duty to deliver a philosophically convincing theory of personhood in terms of agency. Like other animals human beings act. But there is something distinctive about the nature of human action. In virtue of a certain type of self-consciousness humans are conscious of the grounds of their actions. They act for reasons and can

make choices. This distinctive feature of human agency, reason, brings with it two capacities. First, human agents can take control of their actions and have to work out how to do that. They have a capacity for normative self-government. Second, whenever human agents take control over their actions, they are making themselves the authors of their actions. They are faced with the task of deciding who to be. As rational agents, human beings also have a capacity for normative self-constitution, i.e., for constructing norm-governed or practical forms of identity. It is this practical form of identity that I am after in my dissertation.

In the first chapter it will be argued that a neo-Kantian reason view of persons, most prominently and eminently defended by Christine Korsgaard, cannot be satisfactory. Such a view had better be developed into a narrative view of the person.

In chapter two I present a candidate for such a picture: Marya Schechtman's narrative self-constitution view. Although I elaborate the theory she offers, the basic narrative ideas are, I think, essentially right, or at least point in the most promising direction I know of.

With that narrative conception before us, my next concern will be with applying my view to specific practical issues. I want to concentrate on the use of advance directives in mental health care. Considerable attention will be drawn to Ulysses contracts. These are written by an author who is competent to consent to medical treatment, stating how he wishes to be treated under coercion in the event of becoming episodically a psychiatric patient (think of manic or schizophrenic psychosis). Also a related biomedical procedure will be considered: advance directives in the case of Alzheimer's disease. Both procedures have in common that they bear upon pre-commitment. Both are cases of an earlier self being in command of a later self, that is, of an earlier self exercising authority and imposing his or her will on

a later self. This raises the question of (respect for) autonomy. So, before I embark on the project of applying my view to psychiatric practice, I will have to make room for the problems of autonomy (and orthonomy) of the person. Developing a theory of autonomy that is related very closely to my theory of narrative identity will be the task of the third chapter. My proposal builds on the work of Harry Frankfurt and Michael Bratman on autonomy. Succinctly put: I think persons are autonomous if they identify themselves wholeheartedly with their self-constituting narrative. As a final step, I claim that Schechtman's basic story about personhood has to take on board some account of autonomy and that Michael Bratman's view on self-government is a good option.

Chapter four will try to solve the pre-commitment puzzle cases in which pathologies might undermine agency and the autonomy of the person. I will argue that my theory of narrative identity provides a criterion that enables to discern pathological states of personhood. Considering Ulysses contracts, it will become clear that they are conceptually dubious and lead us to two problems. First, there is a problem with personal identity. The identity problem centres on the identity relation between the author of a Ulysses contract, and the patient who is taken to fall under the authority of it. There is a second and related conceptual problem. Ulysses contracts are said to derive moral authority from the principle of autonomy. But if it is claimed, as some philosophers do, that the person now is not the person of the coming action, the exercise of autonomy becomes inexplicable. I think that these puzzles of identity and autonomy are solvable and that the use of Ulysses contracts is morally justifiable. Things become more complicated when we consider the parallel case: the use of advance directives by patients suffering from severe Alzheimer's disease. The rest of the chapter will be an attempt at clarifying

these puzzles too. It will turn out that, in my narrativist view, there are no surviving arguments for a moral justification of the use of advance directives in the Alzheimer's case.

Let me go back for a second to Kendler's thought experiment and make two final methodological remarks. The first is that my approach to philosophy will make liberal use of empirical results. I am interested in the scientific progress made on the other floors. As a philosopher I have a distinctive voice, but I do not have my own, independent and immanent access to reality. The philosophical attitude is a scientific one of rational account, rigorous argumentation and openness for discussion of standpoints. Yet, to understand and interpret everyday experience, philosophy can only be practised in critical interaction with one or more of the empirical sciences. In addition to discussions of inspiring positions of classical and contemporary colleagues, philosophers have to relate their insights to authoritative scientific interpretations. They should be open to what they can learn from the other disciplines. This is not undisputed, especially in practical philosophy. It is said, for example, that such an approach confuses the first-personal perspective with the third-personal perspective. To my mind however, the two perspectives are twisted together. A person can never conceive of his own conceptual capacity from a purely third-personal perspective, because he can conceive of it only with that capacity, and hence from a perspective in which it continues to occupy first-person position. Trying to break into the first-personal realm from the outside is like trying to break into the room where you are sitting.²

² I kidnapped this analogy from the dissertation of Marc Slors (2001). For another helpful picture see D. Velleman, in Buss S. & L. Overton (2002): 114.

The second remark is a related one and concerns the readers of this dissertation. Efforts of cooperation with other scientific disciplines are to little avail if presented in inimitable argot. Although my aim is to develop a good philosophical argument, my dissertation tries to be more than a piece of philosopher's philosophy. It will also be readable and useful (I hope) for other academics operating on, as Edmund Husserl would call it, 'a higher level of naivety'.

1: IDENTITY AND PRACTICAL REASON

Our deeds determine us, as much as we determine our deeds.

G. Eliot

1.1 Accounts of human agency: a state of play

What is a human action? Philosophers of action traditionally define this topic by studying the following problem: What is the difference between a bodily movement (a mere happening) and a human action? Or, as Ludwig Wittgenstein famously rephrased the question: ‘What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?’ (*Philosophical Investigations*: § 621). We tend to think that whatever happens is caused to happen by other happenings. But if *we* make things happen, those events owe their occurrence to us, to persons. We are not simply caught up in the flow of events; we seem to intervene in it by producing some events and preventing others. We seem to act upon and change the world and our path in it. But how can a person make things happen in a world where events are caused by other events? What makes an action ours, rather than something that just happens in us? How can it count, after all, as an intervention – or, for that matter, as *ours*?

The standard answer to this question goes like this. We want something to happen, and we believe that some behavior of ours would promote its happening. These two attitudes of desiring and believing jointly cause the relevant behavior. Because these attitudes also justify the behavior that they cause, that behavior eventuates not only from causes but also for reasons. Thus, for

example³: You want to know the time. You believe that looking at the bell tower will result in your knowing the time. These two attitudes cause a glance at the tower, thus manifesting their characteristic causal powers as a desire and a belief. The desire and belief that cause your glance at the bell tower are your reasons for glancing at it; and because you engage in this behavior for reasons, you make it happen. Knowing the time is, of course, only one small example, put to illustrate the claim made for the standard view of agency: in order for a chain of events to constitute a person's making things happen, it will have to constitute his doing something for a reason. The standard model connects the person's role in producing an action to behavior that is based on and performs for reasons. Reasons are closely associated with personhood. One of the premises underlying the standard model of agency is that causation via a person's reasons qualifies as causation by the person himself.

There is a problem, though. The standard view fails to specify the way in which action involves causation by reasons. As Velleman demonstrates⁴, there is a category of ungoverned 'activities' (as opposed to actions) containing the things that one does rather than merely undergoes, but that one somehow fails to regulate oneself autonomously. Slips of the tongue are Velleman's example: one has reason to say something but is saying something different. These 'activities', as Velleman calls them, are in one sense made happen by the subject; in another, they are made to happen

3 Borrowed from Donald Davidson, the foremost exponent of the standard view and first major source of the revival – after Wittgenstein – of the idea that actions are to be explained, causally, in terms of mental states or events.

4 This section partly keeps track of D. Velleman, 2000: 1-31.

despite him. So, what the standard story describes is not human action *par excellence*. Human agency comes in degrees. And full-blown human action occurs only when the subject's capacity to make things happen excludes movements to which a person is impelled by motives over which he has no control. So, the standard view seems to be an account of motivated 'activity'. It specifies the way in which behavior must be caused in order to qualify as a purposeful 'activity', but not how it must be caused in order to qualify as autonomous action. Autonomous action requires something more than motivation by reasons.

The best sustained attempts at a repair of this flaw in the standard view are contained in the work of, inter alia, Harry Frankfurt, Christine Korsgaard and David Velleman. Harry Frankfurt adds a mechanism modifying the motivational forces at work. It surveys our motives, blocking or inhibiting some of them, and reinforcing others. The result is a hierarchical view of agency. Autonomous action, according to this view, is behavior motivated by the desires and beliefs by which the subject wants to be motivated. The difference between 'activity' and full-blooded action is that in the case of the latter, the agent identifies with his operative motives. Frankfurt's candidate that might constitute the agent's self-identification is second-order motives: the behavior that persons make happen is that which is caused by their first-order motives as reinforced by higher-order motives. So, according to Frankfurt, higher-order satisfaction with one's motives makes for agent-origin and constitution of the person. Christine Korsgaard and David Velleman, for their part, take a more rationalistic stance. They define autonomous action as behavior whose first-order motives are perceived as reasons and are consequently reinforced by 'higher-order mo-

tives of rationality'⁵ In their view, persons are somehow identified with their own rationalities.

I take it that up to now the most elaborate avenue of inquiry into personhood and personal autonomy in terms of agency is the work of Christine Korsgaard (1989, 1996, 1999, 2008, 2009). Korsgaard offers a solid and eloquent neo-Kantian⁶ phenomenology of reflective agency and goes on to derive an attractive account of the person in terms of deliberative action. I think that her work is vital to clear up the relationship between agency and personhood.⁷ So, let me first discuss *in extenso* the argument on self-constitution put forward by Christine Korsgaard during the last decade.

⁵ This is Velleman's (obscure) phrase. He wants to exclude movements produced by motives which the agent endorses without regarding them as reasons. What Velleman is after is 'something like a desire to be actuated by the best reasons' (2000: 14).

⁶ The Kantian famously is concerned with what people do and emphasises their agencies. Kant pointed out that as a rational being I can view myself in two different ways. On the one hand, I can view myself as a natural entity, whose behavior is causally determined by certain other natural phenomena. But I also can consider myself as an entity that freely performs actions, an entity that deliberates, chooses, and originates actions. It is from this practical standpoint, Kant says, that we view ourselves as agents, and as the authors of our actions.

⁷ I'm using the concept of a 'person' rather loosely here and as an equivalent of 'self' and 'identity'. What I mean by those terms is personhood in a practical, unreflective sense. In everyday life people take themselves to have identities or to be a self or a person or somebody. What they mean by those terms are the characteristics they have that distinguish them from others.

1.2 Actions and persons

Christine Korsgaard contrasts her picture of the person with what she calls a Combat Model of the person (1999, 2009). In a Humean variant of this Combat Model the person's actions are just the result of the play and combat of the psychic forces within him, reasons and passions among them. The problem then arises that we seem to have lost view on an agent.⁸ For, the agent seems to be something over and above these forces working on him, something that can intelligibly be said to determine himself to action. We could add such a person as a sort of captain who stands ready to stop things that go wrong and to continue things that go well, but then a new problem arises: we are completely mystified by the existence of a separately existing entity that chooses to identify with one of the parts. So, Korsgaard rejects the Combat Model as an unclear picture of the person, and proposes an alternative: her Constitutional Model (1999: 1-3). She baptises it the Constitutional Model, because its clearest appearance is in Plato's *Republic*, where the constitution of the human soul is compared to the constitution of a *polis* or city-state. The agent is something over and above his parts, just as the political constitution of a city-state is something over and above the citizens and officials who live there. A state is not merely a group of citizens living on a shared territory. We have a state only where these citizens have constituted themselves into a single agent that performs actions and so has a life and a history. They have, that is, adopted a way of resolving conflicts, making decisions, interacting with other states, and planning together for an ongoing

⁸ Hume himself denied that we are ever intimately conscious of what we call a 'self'. See his *A Treatise of Human Nature*, Book One, Part IV, Section VI, 252.

future. For a group of citizens to view themselves as a state, or for us to view them as one, all we need is to grant an authoritative status to certain choices and decisions made by certain citizens, as its legislative voice (1989: 114-5). Korsgaard takes the state as an agent defined by its deliberative procedures. A literal political constitution is laying out the city's mode of deliberative action, the procedures by which its collective decisions are to be made and carried out. And just as a political constitution makes it possible for the citizens to function as a single collective agent, the everyday work of deliberative action is constitutive for a human being to function as a person.⁹

What, then, is deliberative action? Korsgaard's answer is as follows. What makes human beings different from the other animals, is that they are self-conscious. They are conscious, not only of the outside world, but also of their mental states, i.e. of the workings of mental forces in them. For non-human animals life is mapped out by their instincts. They pursue what they desire or flee what they fear. Human beings on the contrary are aware of the

9 Korsgaard stresses that we can see the Platonic parts of the soul as the three parts of a deliberative action: appetite makes a proposal, reason decides whether to act on it or not and spirit (*thymos*) carries reason's decision out. If the agent conforms to the dictates of reason, it is because he identifies with his constitution, which says that reason should rule. This line of thought also supports Plato's analogy between the city and the soul. Reason corresponds to the rulers and its function is to direct things; spirit corresponds to the auxiliaries and its function is to carry out the orders of reason; and the appetites correspond to the rest of the citizens (farmers, merchants, craftspeople...) who provide for the city's needs (1999: 6). Note that spirit or 'the sense of honor' as the executive function of the soul requires some hermeneutic effort. It seems to be something with a cognitive and affective aspect (2009: 139-40).

potential grounds of their actions and capable of *choosing* a way of life. Their capacity to turn their attention to their own mental states and activities is also a capacity to reflectively distance themselves from them, and to call them into question. As a human being, I can perceive some situation as dangerous, and find myself with an instinctive impulse to run. But when I bring that impulse itself into view, when I reflect on it, then I can call it into question: ‘Shall I run? Does this situation really give me a reason to run?’ And now I have to decide. So, self-consciousness is making it possible to transcend mere reactivity in my relation to the world. It puts me in control. Instincts no longer exclusively determine what I do.¹⁰ Inner and outer forces make proposals for action, but I may or may not have *reasons* to act in the way they propose. The proposals are something I need to reason about and to decide upon deliberately. As a self-conscious being I act for reasons and I am free, in principle, to endorse or reject any proposal reflectively and to act accordingly. Nothing is a reason for me except what I make a reason for myself. That is, my decision whether to act on a reason or not takes the form of a self-determining action.¹¹ What distinguishes deliberative action from mere behavior and other physical

10 Korsgaard is aware of the fact that some animals (apes, dolphins, elephants, parrots) have rudimentary forms of self-consciousness. The empirical evidence on these matters is, she thinks, inconclusive. But there is no question for her that human beings are self-conscious in a very particular way (2009: 115-6).

11 A similar idea is what R. Jay Wallace defends as volitionalism in *Three Conceptions of Rational Agency*. The distinctive feature of a volitionalist conception of rational agency is the postulation of a kind of motivation that is directly subject to the agent’s control, and independent of the dispositions and desires to which the agent is passively subject (1999: 236-242).

movements is that it is *authored* – it is in a quite special way attributable to the *person* who does it. It springs from and is in accordance with the agent's constitution, that is, the way the mental forces of the agent are put together, like the people in a city, and give the agent the kind of unity he needs to be the author of his actions. So, according to Korsgaard, the work of practical deliberation is not only taking control of what to do, but at the same time it is a work of unification. An authored action has to come from the person working together as a *whole* (1999: 3). It is worth pointing out that Korsgaard's self-constitution model also provides a standard for *good* action with this. Good actions are actions that are most truly a person's own, i.e. actions which most constitute the person as their author and which most fully unify the person.

I realise that I have been summing up different claims very quickly. I will return to those claims later. It suffices that it has become clear by now that persons have a form of identity that is constituted by their reasons and choices. What gives persons their identity is the reasons they autonomously adopt for themselves. Fully developed humans are consciously involved in the construction of their identity and do not exist, so to speak, prior to their reasons. Constructing, creating, (re)shaping, maintaining, improving, in all these ways constituting a kind of identity is the everyday work of their practical deliberation. Of course, the human form as such is given by nature. Humans share it with the species. But nature sets each human being a task. He must make himself into a particular person (someone saying 'I want to make something of myself' is just describing the human condition). He must give and choose reasons for his actions in a way that is self-determining and self-constituting.¹² So, what

¹² Note that there is no paradox here. The picture is not of a craftsman who is, mysteriously, his own product (Korsgaard 2009: 41-2).

gives humans their identity is what they do, i.e. their characteristic activity. Personal identity is in a quite literal way constituted by choice and action. Let me elaborate this a little more.

1.3 Practical identities

Korsgaard's jumping-off point, as we have seen, is the observation that human consciousness has a reflective structure. Reflective distance from our impulses makes it both possible and necessary to decide which ones we will act on: it both enables us, and forces us, to act for reasons. Reflection enables and forces us to make rules¹³ for ourselves, and it makes these rules or reasons obligative. There is a tight connection between the answers we give to the questions that reflection poses us in the practical realm and the conception we have of our identity. An agent adopts reasons in the form of conceptions of himself as someone's friend, or as a doctor, or whatever. The conception of one's identity in question here is not a theoretical one. It is a description by which we want to understand ourselves and by which we value ourselves and find our lives to be worth living and our actions to be worth undertaking. It is a conception of a *practical identity*, that is, a normative conception of oneself that can take the form of an endorsed collection of rules for deciding what considerations count as reasons for us. For an average person there will be a jumble of such conceptions. As examples of practical identities, Korsgaard

¹³ I am avoiding Kantian parlance here. I use reason or rule as synonymous with law. Reasons (laws) have binding (legislative) force.

catalogues: being a human being, a man or woman¹⁴, citizen, mother, friend, student, egoist, catholic; being a member of an ethnic group, family, profession, or movement; being someone's lover and so on (1996: 101-2, 105-7, 113 and 120). And all of these practical identities give rise to sets of reasons, because they determine what 'counts' for us. These reasons express our identity, and establish a relation which we have to ourselves. They direct us to do what we think it would be a good idea to do, and that in turn depends on who we think we are. To violate them is to lose part of yourself ('I couldn't live with myself if I did that!').

There are two questions Korsgaard's argument seems to leave open. First, if we act for reasons that count for us, we take these reasons from different practical identities; but it remains unclear how different identities coincide and constitute one person. To have plans or reasons taken from different practical identities is not to say that one is a unified person. Second, there seems to be a deep element of relativism in her account. Whether reasons can bind us depends on the way that we value our identities. Yet most of the self-conceptions that govern us are contingent. We are born into a certain family and community, perhaps even into a certain profession or craft. Different reasons hold for different identities. And if desired, we can walk away from them. So, why should it matter whether we live up to the demands imposed on us by contingent citizenship, motherhood, or profession? How

¹⁴ Elsewhere Korsgaard rejects the idea that 'gender has to be or should be a deep fact about the identity of a human being' (in 'A Note on the Value of Gender-Identification' in Nussbaum M.C. & J. Glover (1995) (eds). *Women, Culture and Development: A Study of Human Capabilities*, Oxford: Clarendon Press: 401). There is also no mention of race or sexual orientation. This is observed by C.W. Gowans.

can we be bound by obligations which spring from conceptions of our identity which are not in themselves necessary? Korsgaard thinks she can put worries over the unity and the contingency of practical identities to rest. I will return to her discussion of these (related) issues in a moment. Let me first consider her instructive presentation of failing unification.

1.4 Defective action

Korsgaard describes five different kinds of self-constitution (she follows Plato who in Books VIII and IX of the *Republic* distinguishes five different ways in which the human soul may be governed, comparing them to five different kinds of constitutions possible for a city): the good way (monarchy or aristocracy) and four bad ones, growing increasingly worse: timocracy, oligarchy, democracy and tyranny (1999: 17-20). The good constitution is truly unified. Defective constitutions lead to disunity and to that extent undercut agency.¹⁵ I will concentrate here on the democratic and the tyrannical person. The democratic person in contemporary philosophical jargon is a kind of ‘wanton’. He puts his experiences on an equal footing and governs himself only in a minimal sense. The coherence of his life is completely dependent on the accidental coherence of his desires. It is merely an accident that some impulse leads to an action, and it is also merely an accident if that does not happen. To see the problem, consider the story of Anne-Marie:

¹⁵ Korsgaard does not insist on the fact that there are five types of constitution. There are more. She reminds us of Plato’s *dictum* that ‘there is one form of virtue and an unlimited number of forms of vice’ (2009: 163).

Anne-Marie is a 30 years old woman and was born with ADHD (Attention Deficit Hyperactivity Disorder). As far as she can remember her life has been one big chaos. She's a great collector. Her living room is full of knickknacks, piles of CDs, countless books, a prie-dieu with paternoster, and lots of photographs. She's not able to throw away or let go things, because she cannot organise. Her inner life is the very same. Her head, she says, is like a chock-full repository. Sometimes, something is falling out of it unintentionally; she goes on to act or to speak thoughtlessly, and in no time runs into difficulties. Moreover, she is not capable to select or filter sensory impressions. She has to react to every observation: sounds, pictures, smells, temperatures. Willy-nilly. Everything is equally urgent and in need of her immediate response: a car passing by in the street, the conversation with her friend, the noise of the airco, a brainwave, and her clothes that feel like twisted: they all try to be at the centre of her attention simultaneously. Like this, Anne-Marie is skipping and living from one subject to another. In her own words: "Although I'm living in overdrive, I'm spinning around on site. I cannot build up valuable and enduring things. I'm reading three novels at once and never read to the end. I cannot sit out a movie, because I cannot follow the plot; there are too many associations. I cannot enjoy sex, because I continually get up to do other things. As an adolescent I was a rebel without a cause, hunting on thrills and kicks. I had lots of boyfriends, but I was isolated; a long relationship took three days." And the story goes on. Adult ADHD-patients typically are 'people of twelve professions and thirteen disasters', as the Dutch saying goes. Anne-Marie first worked as a teacher (but she forgot names, prepared the wrong lessons and repeatedly lost her students' examinations). Then she became copy editor and again lost her job. Her life was in shatters. At the age of 28 a psychiatrist made her diagnosis, and Anne-Marie was having therapy. She took up her life again and started to read and gather everything on ADHD.

*She worked as an ADHD-expert for the Dutch ADHD Patients Association (called 'Impuls') and went to a school of acting, quite sure that, thanks to ADHD, she will never turn into a grey mouse.*¹⁶

The problem of this 'wanton' is that she is at the mercy of accident. Her choosing by lot is only in a minimal sense different from not choosing at all. She drops her projects in the face of the slightest temptation or distraction. She may be almost completely incapable of effective action. Korsgaard says that what the wanton illustrates is absence of action; the wanton is a case of mere privation of self-government.

The tyrannical person goes in the opposite direction. He is consistently ruled and unified. In a sometimes horrifying imitation of unity and simplicity the tyrannical constitution is governed by one sole desire, which subordinates the entire soul to its purposes, leaving the person an absolute slave to a single dominating obsession. Plato tells us the tyrannical soul is governed by a nightmarish erotic desire, and envisions tyranny as a kind of madness. For modern people it is hard not to think of extreme forms like the addict or the psychopath. Their relation to their obsession is like a psychotic's relation to his delusion: the psychotic is able, and prepared, to organise everything else around his delusion, even at the expense of a loss of his grip on the real world (Korsgaard 1999: note 19). The psychopath may actually view his victim as asking for the violence inflicted upon him, for example, because he needs to see him that way. And for the

¹⁶ Published in the Dutch magazine *Flair*, January 1999. Wantons, in Frankfurt's rendering, are not even interested in the cumulative outcome of their competing first-order desires. So, speaking strictly, Anne-Marie is not a wanton. She has a problem.

addict a house is not full of somebody else's furniture but of things he can sell for the money for the drug. The tyrannical person may act for reasons and it's true that he is governed and unified, but Korsgaard stresses that he is not self-governed. He does not decide what is worth doing. He's more or less condemned to reenact some important scene, some desire that has a hold on him, and that takes over. To that extent he's internally enslaved (2009: 169-73).

1.5 Unifying agency

Let me now elaborate the Korsgaardian idea of the unity of the person in more detail. Why do I think of myself as *one* person? Korsgaard points out two reasons, both practical, for conceiving ourselves as a unified agent at a time as well as over time (1989: 110-112). The first reason stems from the raw necessity of eliminating conflict among my various psychical functions. Because I have only one body with which to act, I must conceive myself as a unity in order to so act. Sharing a common body my psychical functions must work together. Secondly, there is practical unity implicit in my deliberative standpoint. The idea is this: I have various reasons for or against acting on certain competing inclinations, and in deliberating, I weigh these reasons and I choose which ones should count as good or better reasons for acting.¹⁷ Identifying with this way of choosing imposes a practical unity

¹⁷ By choice Korsgaard does not mean arbitrary choice, like choosing a stamp from a stamp booklet, or like wantonly choosing by lot. This raises questions about sufficient criteria on which Korsgaard can set good choices apart from less good choices. I will pursue this complication in section 1.6.

upon me. Choosing to act for a reason is pulling my mental forces together into a unified system, into a single agent (as the classes in a polis make up the unity of a constitution). At first blush, it may seem odd to suggest that my unity as an agent at a single time is explained by my deliberative standpoint. Most of the time there may be neither a conflict among different reasons nor need for deliberation. When I think of myself as one person at a particular time, I may be doing several things at once, e.g. peel potatoes, listen to a Bach toccata on CD, and keep an eye on my four years old son playing on the floor. I certainly consider all these actions to be unified as mine at that time, but I cannot easily feel the unifying force of the deliberative standpoint here. When I turn from synchronic unity of agency to diachronic unity of agency, the unifying force of the deliberative standpoint becomes more plausible. Most of the things I do that matter to me take up time (health, careers, friendships, family lives...). In choosing them I identify with something from which I will derive my reasons: being a father or a good soldier. The choice takes me some way into the future and I need to identify with that future in order to be what I am even now. I, then, both construct continuity of agency and of identity. It is necessary to be one continuing person over time in order to carry out, or even make, life plans. In Korsgaard's words: 'We are one continuing person because we have one life to lead' (1989: 113). Thus, on close inspection, it is practical reason that requires us to construct a unified identity for ourselves. The need for the unity of the person is imposed on us by the necessity of practical deliberation. We would not be able to make plans, choose careers etcetera, if we were not to conceive ourselves as one. This unity is perfectly consistent with drastic changes, Korsgaard argues, provided that the changes are the result of actions by the person himself or reactions for which he is responsible. Where I change

myself, the sort of unity needed for identity may be preserved, even if I become very different. Where I am changed by wholly external forces, the unity is not preserved. Only changes that are of my own doing are authorial changes that preserve my unity as an agent.¹⁸ So, changes made for reason.

I should mention one more thing with regard to Korsgaard's account. Korsgaard adds that authorial changes are changes that respect the fundamental identity that stands behind our practical identities. Let me explain. Though most of a person's practical identities are contingent, the necessity of having some practical identities is based, not on one or another particular contingent practical identity, but on what stands behind them all: a person's human identity as a 'reflective animal who needs reasons to act and to live' (1996: 121). As reflective beings, humans must ultimately come to see their humanity itself as an essential and foundational feature of their practical identities. The reason for conforming to practical identities springs from their 'membership of the party of humanity', i.e. from their identity simply as valuable-qua-human. We can query the importance of a practical identity. But we cannot ignore the fact of the human condition that people have no choice but to be rational and to act on reasons. This involves an inescapable commitment to endorse the value of reflectiveness and – because reflectiveness is a basic characteristic of human beings – of humanity generally, not simply my reflectiveness or my humanity. Ultimately, it is the very fact that I encompass human (ident)ity, that is constituting me into a unified person. Humanity cannot be shed. It seems

¹⁸ Korsgaard also notes that 'when they can communicate, two different people can integrate their functions, and, for purposes of a given activity, become a single agent' (1989:119). We can think of a good marriage. And, as we already know, a corporation or a state also can be a collective agent.

impossible not to value my own humanity. I cannot ‘un-think’ or ‘de-constitute’ myself by claiming that my humanity lacks value; I’m still giving reasons then and thereby presupposing and valuing my own humanity. Humanity is an organising principle that I must follow.¹⁹

I pointed earlier (in section 1.3) to two difficulties in Korsgaard’s argument: the unity problem and the threat of relativism. It is time to enter into the second difficulty: Why should we suppose that obligations that issue from contingent practical identities are normative for us? How can we be bound by reasons which spring from conceptions of our identity which are not in themselves necessary? The content of our reasons clearly is drawn from natural and social resources, from contingent prac-

19 For Michael Smith it is conceivable that my reflective self should want me not to be reflective. Therefore, in his view, Korsgaard cannot be possessed of a premise from which she can draw the conclusion that I have to value my own reflective nature (1999: 393-4). I agree. People have an option not to identify with their reflective nature and not to take it as their ‘most inner nature’.

It is important to note that Korsgaard’s argument leads into moral territory. Valuing one’s own humanity requires valuing the humanity of all persons (1996: 121 and 132-45). Korsgaard calls this ‘the Moral Law’: the principle which demands people to act only on reasons that they can share with all rational beings who live together in a cooperative community. She distinguishes between ‘the Moral Law’ and the purely formal ‘Categorical Imperative’: the principle which demands people to choose only reasons for action they can regard as universal laws (1996: 99; 2009: 80). In her view, there is a path of argumentation between the two. I am sceptical about a conception of practical reason whose core principle is the principle of morality. Vide my discussion with David Velleman about the moral value (dignity) of humanity in 4.10.

tical (social, religious, professional, ethnic, relational, political &c.) identities. Korsgaard's answer in a nutshell: some principles set standards for the form of our reasons and determine what counts as a reason. Korsgaard intimately relates her model of self-constitution to the principles of practical reason. Let me now detail the normativity part of the story.

1.6 The normativity of the principles of practical reason

We saw that it is in the internal world created by self-consciousness, that reason is born. Self-consciousness as the source of reason opens up a space of reflective distance between our impulses and responses. In this space the question whether our impulses give us reasons arises, but, in order to answer that question, we need principles which determine what we are to count as good or bad reasons. As in the state we do not just need some constitution, we need a constitution that results in just laws and righteous policies. In the context of the argument of the *Republic*, the constitution of a state is committed to the principle of justice. On the Korsgaardian story, the basic principles for self-constitution or being a person are the principles of practical reason, namely, efficacy and autonomy.²⁰ Efficacy and autonomy

²⁰ The two principles of practical reason, efficacy and autonomy, correspond to the two Kantian imperatives. On Korsgaard's account, in fact there is just one imperative here: act in accordance with a reason you can will as a universal reason. The choice of actions must be governed by a categorical imperative. The hypothetical imperative merely specifies the kind of reason we are looking for – a causal reason, a practical reason. And that thought is already contained in the idea that what we are looking for is a reason that governs action (2009: 70-2).

set standards for the form of our reasons. Unless we reason in accordance with these principles, we will not succeed in being one person at all. Efficacy or the principle of instrumental reason says that if we will an end, we must will the means to that end. An agent is efficacious when he succeeds in bringing about whatever state of affairs he intended to bring about through his action. Autonomy is the other principle of reason for the domain of action that directs an agent insofar as his actions are in some clear sense self-determined or his own. That is, an agent has to act in accordance with a reason that can, in principle, be adopted by any, hence by all, ordinary agents. In other words, an agent has to act in accordance with a reason he can will as a universal reason.²¹ Reasons are the sort of things that people give and receive, exchange and refuse. If any consideration is to count for

21 Here, we have to do justice to Korsgaard's refinement of the conception of universality. She enters two caveats. Firstly, because reasons are quite specific to the situations at hand, universality may require a high degree of specificity. And secondly, Korsgaard distinguishes three ways in which we can take rules to range over a variety of cases. To start with, a reason can be treated as *general* when we think it applies to a wide range of cases. If we think of a reason as merely general, and we encounter an exception, nothing happens. We expected there to be some exceptions. Second, a reason can also be treated as *absolutely universal*, when we think it applies to absolutely every case of a certain sort. And it can be treated as *provisionally universal* when we think it applies to every case of a certain sort, unless there is some good reason why not. If we encounter exceptional cases, we go back and revise the reason, bringing it a little closer to absolute universality to which provisional universality essentially aspires. The reasons with which we operate in everyday life are, runs the Korsgaardian account, provisionally universal reasons (2009: 73-5).

me as reason-giving, it also has to count as a reason for others (or for a later self) in similar circumstances.

The principle of autonomy avoids two extremes, i.e. the two forms of defective self-governance sketched above. For one thing, it shows that reasoning cannot be a matter of deference to any antecedently given external standard (an obsession for example); external standards can only justify insofar as independent reasons are offered for accepting their putative authority. For another thing, reasoning cannot be a matter of lacking all standards, that is, of anomy or randomness. Korsgaard considers this option at length and comes up with an ‘argument against particularistic willing’ (2009: 72-6). It is her objection against ‘wantonness’. Particularistic willing would be taking some consideration as a reason to do A, thinking that you can use it just this once and then, so to speak, discard it without any implications for any other case. You do not even need a reason to change your mind. Such particularistic willing makes it impossible for you to distinguish yourself, your principle of choice, from the various impulses on which you act, Korsgaard says. It eradicates the distinction between you as a person and the impulses in you. There is nothing left then that is the person or the agent; nothing that is your self-determined will as distinct from the play of impulses within you. You are not one person but a series, a mere heap (in Aristotle’s phrase), of unrelated impulses. To become a unified agent who can be the author of his actions, you have to take your reasons to be intelligible to yourself at later moments and to others. Korsgaard concludes that the reasons of a person must have ‘type’ character. When reasons only have ‘token’ character and the difference between ‘type’ and ‘token’ reasons cannot be made, we can no more distinguish someone who has a particularistic will

from someone who has no will at all. In other words, we must take your reasons for action to be universalisable, or at least provisionally universalisable. Such reasons may be very specific but universalisable they must be.²²

In fairness to Korsgaard, it must be noticed that, according to her view, something more is needed for the unification of the person than merely the formal principles of practical reason. In fact, Korsgaard proposes an Aristotelian-Kantian theory on self-constitution. To the rational constraints for being a person Korsgaard adds the Aristotelian idea that a person, like pretty much any substance or entity, is a functionally organised unity. Persons are so organised as to serve some purpose or function. They want to realise their nature, because that is how a person functions. Aristotelian theory says that the good for a person rests in its well-functioning as the kind of being that it is in circumstances favorable to that well-functioning. So, human persons have capacities and limitations, needs and interests that arise from their general human nature and not just from the particular nature

22 In my estimation, Korsgaard conflates two interpretations of the universalisability requirement. One is that, in order to have normative force, my public reason must be a *shareable* reason, that is, a valid reason for all rational agents resembling me in all relevant respects. The other interpretation is that my reason in order to have normative force must be *acceptable* for all rational agents. Shareability differs from acceptability. It does not follow from the fact that my reason is intelligible and shareable, that other people have to be pleased with my entertaining such a reason and have to accept it. This raises the question of how on a public conception of reasons in terms of shareability alone Korsgaard's universalisability requirement is going to get us into moral space.

(identity) that they have carved out for themselves. Looked at this way, one has to say, as Korsgaard does, that the unification task for a human person has two parts:

- i. to function well in the identity he has chosen for himself, and to be in the circumstances that enable him to do that, and
- ii. to have an identity that enables him to function well as a human being, and to be in circumstances that enable him to do that (Korsgaard 2010a: Lecture Three).

Let us recap. What are the lessons about persons that we can draw from Korsgaard's self-constitution view? Korsgaard believes that persons are in a quite literal way constituted by action, in other words, by the simple fact that they have to act. Action is necessary.²³ To be a person is to be engaged in the everyday work of practical deliberation, that is, to be engaged in constantly choosing and reason-giving. The basic rules for being a person are the normative standards internal to action, i.e. the principles of practical thinking. Reason's own principles (*de facto* universalisability) just are the principles of choosing and acting in a way that constitutes us into a person. What Korsgaard is trying to argue is that the principles of practical reason are constitutive standards of action, and therefore, of us. They unify us into agents that can be the authors of our actions. Normativity and unity, in her view, are closely connected to each other. The principles of practical reason are principles of the unification of agency. Not only do we feel ourselves compelled to act for reasons, we want to be unified persisting reason-givers. Successful or

²³ Korsgaard does not mean logical, causal or rational necessity. It is a necessity humans are faced with (2009: 1).

good²⁴ actions are actions which most constitute the person as their author and which most fully unify the person. The conclusion Korsgaard is looking for is that the necessity of conforming to the (moral) principles of practical reason comes down to the necessity of being one and of maintaining our identity. I will now turn to some worries about Korsgaard's project.

1.7 Some demystification

My worries circle around the unification theme. To begin with, I think Korsgaard's argument that the deliberative standpoint imposes unity on agency and personhood is not successful. Korsgaard overestimates the unifying role of the human rational abilities. How so?

For a given person, many different and incompatible possible practical identities are valuable. Choices have to be made among the practical identities – more precisely among the reasons they generate – that are possible. Moreover, we do not ordinarily suppose that any such identity is as good as any other. Some practical identities are better than others and we ought to seek those that are among the best. Trying to understand the full meaning of *Finnegan's Wake*, for example, is preferable to watching reruns of your favorite TV quiz *Lingo* as often as possible. And we

²⁴ Note that, on the Korsgaardian story, good and bad can be taken here without problem in a moral sense. Korsgaard offers us moral anthropology. Reason's own principle is the principle that really unifies the person and also the principle of the morally good person. In Korsgaard's view, integrity in the metaphysical sense – the unity of agency – and in the moral sense – goodness – are one and the same property.

normally think that a hobby like making mountain trips is preferable to collecting beer bottles. All these activities can be object of concern and become an organising principle in a person's life. Now, to adopt or to give up a practical identity, as reflective beings, we must do so for some practical reason. What source of reasons is left, one could ask, to base a choice on (given that we choose among morally permissible self-conceptions)? At this point Korsgaard's principle of autonomy comes into play, since to act for a reason is always to act on the basis of some 'universal law'. But none of the examples above necessarily would conflict with the universalisability constraint, that is, would be unreasonable or impermissible. So, it remains perplexing how there could be a basis in Korsgaard's normative principles of practical reason alone that explains how one reasonable practical identity could be better than another.²⁵ We cannot rank practical identities in importance on account of making practical reason, in particular the requirements of universalisability, the sole arbiter. Although practical identities are defined to an extent by reference to the categories of practical reason, universalisability cannot per se provide guidance for determining how to choose between different practical identities. Requirements of rationality demand not to violate formal standards like consistency or universalisability. Such logical standards are value-neutral. They do not answer questions of assessing the weight of reasons we have for doing Y or refraining from doing Y. So, the question whether we do have good reasons for valuing and choosing a practical identity cannot depend on the question of fulfilling or violating the formal standards of rationality. It is beyond their scope.

Korsgaard admits that there is a problem here. She sees that

²⁵ C. W. Gowans (2002) also stresses this point.

we need some method of balancing our various ends against one another when they cannot be practically combined. She tags this method ‘the missing principle, something besides the instrumental principle and the categorical imperative, since neither of them will do this job’ (2009: 57). She continues: ‘I have no formulation to offer for a formal version of the missing principle. (...) But it seems clear to me that if we could formulate a version of the missing principle that is formal rather than substantive, then it should be easy for me to make my case about how its normativity is to be established. What I mean is that it seems rather obvious that a formal principle for balancing our various ends and reasons must be a principle for unifying our agency, since that is so exactly why we need it: so that we are not always tripping over ourselves when we pursue our various projects, so that our agency is not incoherent’ (2009: 58).

I’m not sure that there will be much more to reveal about this mysterious ‘missing principle’ in the future. I think we better acknowledge that practical reason often under-determines our choices. Practical reason can select rationally permissible alternatives, but it is impossible for us to realise all permissible reasons at once. And the injunction ‘try to realise as many permissible reasons as possible’ also runs into difficulties, because in many cases our reasons turn out to be incommensurable. There is room for choice and we have to content ourselves with our choice. We can only be ‘satisficers’.²⁶

For similar reasons the formal principles of practical reason cannot per se account for the unity of persons and cannot per se hold together the person as a unity, or so it seems to me. An example may bring this point home. Let us examine how the

²⁶ I’m indebted to Joseph Raz on this point.

principle of universalisability works in a situation of choice, when different alternative practical identities are under consideration. Suppose (I borrow this case from Richard Rorty, 2007: 198) there is a Catholic doctor who thinks she would rather die than kill a fetus, and who now turns out to be a desperate rape victim's only hope. She asks herself: 'Should I retain the practical identity I presently have, or rather develop and cherish the new identity I shall have to assume if I do what my present practical identity forbids?' The doctor may find herself going back and forth between a proposed new practical identity and her old reasons, trying to fabricate a possible identity which will integrate both. All the reasons for action this doctor is weighing certainly will be universalisable, that is, they all will be reasonable for other people in the community. To put the point in Korsgaardian terms: the reasons for being a non-killing Catholic doctor will be provisionally universal reasons (see note 21). In the rape's victim case we may have encountered an exception to those reasons or rules. But the reasons for being a killing doctor in cases of rape victims will also be provisionally universal reasons. What, then, is the absolutely universal reason to which these provisionally universal reasons aspire? It's not at all clear how a principle of universalisability can affect the (dis)unity of this person, who has a very strong urge to pursue a new practical identity that is generally forbidden by her Catholic identity. The physician is unwilling to give up either that new identity or the old ties to Catholic milieu and identity, and does not know what to do. Looking for reasons to guide her, she only faces an arbitrary choice among various universalisable reasons. I do not think, as Rorty puts it frivolously, that the doctor's problem 'simply' is a question about which of the many available suppliers of alternative practical identities she should buy from (2007: 199). But I am also unmoved by the Korsgaardian suggestion that some for-

mal normative principle – as an invisible magnet – attracts and unifies the doctor’s reasons. Formal constraints on what counts as a reason, i.e. the principle(s) of practical reason, cannot provide for substantive guidance to our thinking about the unity of practical identities. I can agree with Korsgaard that the reasons to act which I autonomously choose are *part* of what holds me together. I also subscribe to her idea that what counts as a reason, to be a reason at all, must be a public reason, i.e., it must have normative force that can be shared by all rational human beings. Finally, I can even feel the force of Korsgaard’s master argument that there is one practical identity that I cannot simply doubt: I cannot easily question my human identity as a creature capable of reflective questioning and acting on reasons. I think all this is on the right lines. To my mind, however, the rational capacities that human nature supplies do not by themselves make it possible for human beings to reach rational unity. Here I take a decisive departure from the Korsgaardian picture. Korsgaard overemphasises the unifying force of reasons. Of course, human beings will strive for some unity to realise their different plans, intentions, and concerns in their one and only human life. If we think of the person as a more or less coherent collection of practical identities, as a collection of reasons that somehow fit together into a whole which is sufficiently unified to make action possible, then bringing the aspects of our practical identities together into some kind of an integrated whole is important. The person in question will try to realise some homeostasis, some kind of ‘reflective equilibrium’. Yet, my point is that the basic consistency of a person’s reasons only in a limited way contributes to the person’s unity; and that it certainly is not the only contributor. Working its way into or out of a practical identity is a matter not just of adopting or rejecting principles of practical reason but also of undergoing psychological change. In order to

reach a new ‘integrity’ our Catholic doctor will have to change dramatically. Changing a set of available reasons often requires substantive psychological change.²⁷ Persons are constituted by their reasons, but they also are embodied and embedded entities. They are co-constituted by their psychological make-up and the circumstances.

It should be conceded that Korsgaard tries to deal with this problem. Her considered view is that a person is not solely identified with his reasons, and so regards his incentives as alien things: ‘Neither the incentive nor the principle of choice is, by itself “the reason” for the action; rather, the reason is the incentive as seen from the perspective of the principle of choice.’ (1996: 243). So, it is rather the case that the person is identified with his *constitution*, and that his constitution says that reason should rule. Furthermore, Korsgaard brings her self-constitution view in step with Aristotelian metaphysics: persons are functionally organised unities, who in that way try to preserve and maintain their form of life. Nevertheless, Korsgaard is very clear about the forces that are taking the lion’s share of the unification job: these forces stem from practical reason. The principles of practical reason are the glue that holds the person together in one piece.

The objection might now be raised that Korsgaard has given us a second argument for the unity of persons. As noted earlier, she does not only stick to the unity of the deliberative standpoint, but also stresses that the human body imposes unity on agency and personhood. However, her elaboration of this point is a bit disappointing. To put it briefly, Korsgaard’s self-constitution view remains almost silent on the declared importance of the body. This is, so I believe, a serious flaw in her position. To

27 David Velleman (2006) makes this point in his essay *Willing the Law*.

see why, consider what can happen if one takes a reason view without embodiment to the extreme. That is, if one derives the unity of the person from a rational point of view alone.

How such a line of thought may exactly weaken the unifying role of the body and lead to increasing disunity of the person is what we can learn from Carol Rovane's congenial but much more radical normative analysis of personal identity. We can read her argument as a *reductio ad absurdum* of Korsgaard's emphasis on rational constraints. In *The Bounds of Agency* Rovane offers a psychologically based account of personal identity, which denies that personal identity can be defined in terms of the human body. Neither does Rovane follow the Lockean view in equating a person with a phenomenological viewpoint, that is, with a single locus of conscious experience. We learn from her normative analysis that 'there is one person wherever there is one agent who has its own rational point of view, and who is committed to achieving overall rational unity within it' (1998: 129).²⁸ On Rovane's account, personal identity should be defined in terms of the unity of a rational point of view, and the boundaries of a person in terms of the bounds of a rational viewpoint. As a consequence, in the case of a human being, there is a practical possibility to adopt different rational points of view and to commit to different unifying projects. Humans in principle develop the rational and practical abilities by which they could realise multiple persons, if only they had a reason for doing so. In the case of our Catholic doctor, according to Rovane, there is an option for a double life – a fragmented human life in which two distinct persons figure (contra a life of constant pain and frustration

²⁸ A rational point of view, according to Rovane, is the point of view from which a person deliberates and that constitutes its psychological economy and unity.

as the physician is struggling to pacify two practical identities that are, ultimately, irreconcilable; or struggling to cope with the sacrifice of one of them). Within a single embodied human being with a single consciousness, Rovane argues, distinct persons can be formed, that is, multiple rational points of view, each with its own commitments to its own overall rational unity and unifying projects. There is no preference for integration, as it is not the case that such integration constitutes personhood more fully (assuming that each of the 'alters' satisfies the stipulated definition of a person). Rovane entails the equal status as persons of both multiple and integrated persons.²⁹ Here is her well-known example (1998: 175-177). Suppose you are committed to pursuing the following different activities: philosophical work, a teaching career, and the study of music. Let us take for granted that you are an individual person with a commitment to achieving overall rational unity within your particular human life. This ensures that you never pursue these coordinated activities in complete independence of one another. You will always bear in mind how pursuing one would affect your ability to pursue the others, thereby also ranking the activities in importance. Now imagine that the activities were pursued completely independent of one another without any effort to rank them in importance, and without any effort to arrive at all-things-considered judgments that take them all into account. This sounds as a description of irrationality, or of a chaotic psychological economy. But it could equally well be a description of a case where several distinct persons exercise their agencies in complete independence of one another. The thought is that there might be, cohabiting in the same human being, a philosopher, a teacher, and a musician.

²⁹ Also of group persons, but I leave aside the possibility of group persons here.

Each person might well pursue its unifying project (we can imagine that each wants just three or four hours a day). In this case, it is not clear why it is necessarily irrational to oppose the goal of overall rational unity within this human being. Multiple persons can have reason to enter into certain sorts of cooperative arrangements in order to share a common resource – as can any other individual persons who need to share resources. And their acts of cooperation can be seen as acts of cooperation by three distinct persons, who satisfy the condition of personal identity specified by Rovane’s normative analysis of personal identity.

I think that Rovane’s reason view is very instructive. It puts the bodily and phenomenological ‘givenness’ and ‘separateness’ of persons in brackets with far-reaching and dubious consequences.³⁰ The unity of the person disappears. We can now meet a team of possible persons in one human body: our neighbor can be a music-practicing-person, an integrated person and a group person (as a member of a string quartet) at once. And the human rational and social nature makes it possible to live on both sides of the borders of these overlapping persons and to draw the lines time and again. For Rovane there is no fact of the matter about boundaries between persons. At any given time it is unclear with which person – or even with how many – we are interacting.

I conclude that the rational viewpoint alone cannot account

30 For a similar case against the ‘Cohabitation Claim’ see Eric T. Olson (2003) in *Was Jekyll Hyde?*. Olson rejects the ‘Cohabitation Claim’ because it rules out our being things made of flesh and blood. Thinking carefully about what we are, and especially about what non-mental properties we have, he argues, we will see that there is simply no room for two or more of us within one human being.

for the unity of the person and that thinking so has deplorable consequences. Rovanean persons certainly are not what Korsgaard has in mind. If such indeed is the case, then Korsgaard's reason-based view has to do more to block the proliferation and disintegration of the person. A plausible way to go would be to give the body a more prominent and elaborated role in the unification of the person.

1.8 First findings

Let me sum up the line of thought so far. My plan in this chapter was to provide an account of personal identity in terms of agency. My point of departure was the ongoing debate on human agency, more specifically the camp of David Velleman and Christine Korsgaard, who define human agency as behavior whose first-order motives are perceived as reasons and are consequently reinforced by, in Velleman's phrase, 'higher-order motives of rationality'. I singled out the self-constitution view of Christine Korsgaard, who offers a picture of reflective agency and goes on to derive a very elaborate account of personal identity in terms of deliberative action. Humans are born for action, she states, and agents are moving through a 'space of reasons'. They have the capacity to deliberate, that is, to choose for acting on reasons about what they would do if their reasons were to survive a process of reflective scrutiny. When an agent reflectively does choose which reasons he is to act on, and, when he acts on these reasons, he does act in ways that are constitutive of his self, viz., of his identity. So, on the Korsgaardian account, identity is practical identity and is bound up with practical reason. The *substantia nigra* of self-constitution is the normative force of practical reason. It is the normative

consistency of reasons that is constitutive for personhood and for the unity of personhood. I agree, but that is where praise must end.

As I have tried to convey, Korsgaard overestimates the unifying power of human rational abilities. Her idea that human beings by nature have rational capacities that ought to achieve overall rational unity within themselves is not convincing. I can accept that reasons, to be reasons at all, must be shareable. But reasonable permissibility as such does not ground unity. All that can follow from Korsgaard's account, as far as I can see, is that human beings *can* achieve overall rational unity. The rational capacities that human nature supplies do not by themselves dictate to what extent human beings have to reach rational unity and how to do it. It is true that Korsgaard, taking her lead from Aristotle, adds extra resources to support her unification argument. In spite of this, the unifying role of practical reason, c.q. the Korsgaardian principle of autonomy (universalisability), turns out to be overblown.

I think we have to be modest about the unification business and to accept a more moderate 'reason view' that leaves room for timely incoherence, discontinuities and drastic changes. Think back for a moment to our doctor's dilemma. One of the things this person can do to get rid of her inter-identity conflicts is to give up her commitment to overall rational unity and to construct a new practical identity to suit the occasion. That is, she can accept temporary incoherence or a far-reaching change of her person. In the moderate view I suggest, persons are constituted by a self-reweaving network of reasons, that is, a self-correcting enterprise which can put any claim in jeopardy, but not all at once. We can think here of Neurath's famous boat. All personal reasons are experimental, fallible, and replaceable like the planks of Neurath's ship. This is not to say that persons aren't

committed to unity.³¹ For the most part persons do not want to be like wrecked ships. But to lead the more or less unified lives of more or less unified persons, all these persons need is a Humean reading of instrumental reasoning instead of Korsgaardian ‘principled reasoning’. And in addition to means-end reasoning they can adopt a Humean conception of ‘reflective rationality’, some steady and general point of view in order to correct their idiosyncrasies. I’ll have more to say about the Humean reading of reasoning in chapter three.

For now, it will be clear that Korsgaard’s self-constitution view is leaving us unsatisfied and that it can only be a reasonable start for a viable theory of persons in terms of agency. In closing, let me say something about the direction we need to take. My proposal is to develop the reason view of personhood into a narrative view. Reasons for acting always are the elements of a possible storyline along which to make up what we are going to do. What serves as the rationale for acting, or as the grasp of what an actor is doing, is a whole story. So I propose to connect a picture of self-constitution in terms of agency with, in this case, the narrative

31 Although I think persons cannot definitively banish wantonness. Persons also look much more like a tyrant than Korsgaard seems to accept. The tyrannical person actually remains a puzzle case for Korsgaard in showing that strong unity can stand apart from agency-at-its-best. I think the force of the reasons to reject this despotic type of defective personhood as ‘too great a unity’ and ‘unity achieved in the wrong way’ mainly is a moral one. Tyrants like the mafia kingpin and the addict clearly are harming other persons. But what about the single-mindedness of a passionate scientist, tennis player, parent, artist, monk or animal protectionist? Those we most esteem tend to have some all-consuming passion. More about ‘obsessive’ human lives in chapter three.

self-constitution view of Marya Schechtman. I have at least three arguments for such a move. First, stories are, in all probability, the easiest and most natural way to put in order our experiences and our knowledge. They seem to be a natural place to look for an adequate understanding of personhood. Second, a coherent self-constituting story will do better justice to the multiplexity, the discontinuities, and the changes of the person than unifying reasons can. In a narrative view on persons the unity constraint can be replaced by a demand for coherence, and reason can play a less normatively potent and more instrumental role. Moreover, formal constraints on what counts as a reason can be supplemented with substantive constraints that provide for substantive guidance in our practical lives. A third argument for developing the reason view on persons into a narrative view is that Schechtman's narrative self-constitution view can be linked up better with the human body. It can repair the second failing of Korsgaard's self-constitution theory mentioned above: the lack of an elaborated theory on the embodiment of persons. Reasons alone cannot be supposed to do the self-constituting and unifying work. Also emotions, desires and the human body do their share. Preliminary work is already done here by Marc Slors in defending the importance of 'basic narratives' (2001). A second avenue of research can be opened by Antonio Damasio's theory of feeling (1999). According to Damasio, neither practical reasoning nor being a person are possible without bodily feeling. So let me now turn to the narrative account of personhood.

2: NARRATIVE SELF-CONSTITUTION

My life, it seems to me, is ridiculously shapeless. I know what makes a good narrative, and lives don't have much of that – pattern and balance, form, completion, commensurateness. It is often the case that a Life, at least to start with, will resemble a success story; but the only shape that life dependably exhibits is that of tragedy – minus all the grand stuff about nemesis, fortune's wheel, and the fatal flaw. Tragedy follows the line of the mouth on the tragic mask (and the equivalent is true of comedy). You rise to the crest and then you curve down to a further point along the same latitude. That's the only real shape lives usually have – and, again, forget about coherence of imagery and the Uniting Theme.

M. Amis (2000), *Experience*, London: Jonathan Cape, 361

2.1 Narrative self-constitution

In *The Constitution of Selves* (1996), Marya Schechtman pushes the discussion on personal identity forward by separating two questions and dealing with each in its own terms: the re-identification question and the characterisation question. The *re-identification* question asks what makes a person at time T_2 identical with a person at time T_1 . More precisely: Can we provide a criterion of personal identity that defines the necessary and sufficient conditions for saying that a person-stage at T_2 and a person-stage at T_1 are stages of the same person? The *characterisation* question asks, most simply put, which actions, experiences and characteristics (beliefs, desires, values and other psychological features) are to be attributed to a given person to make him the person he is. What characterises him as a person? The re-identification question is

the stake of a debate on logical and ontological identity, while the characterisation question is at the forefront of the debate on practical identity.³² For my purposes I will focus on Schechtman's response to the characterisation question here: her 'narrative self-constitution view' according to which persons create their identities by constructing, telling, and enacting an autobiographical narrative – a life story. The question is how this is done.

In Schechtman's view not all individual human beings are persons. Individuals may have rich experiences, but they constitute themselves as persons only by coming to think of themselves as persisting subjects who have had experiences in the past and will continue to have experiences in the future, taking certain experiences in the past, present and future as theirs. In early childhood – at the age of three according to some personality psychologists³³ –, human beings can attribute experiences to themselves that are removed in time. Windows appear as it were

32 The two debates move up to each other but do not find each other. D. Parfit is the owner so to speak of the metaphysical identity-question; C. M. Korsgaard, M. Bratman, the late B. Williams, D. Velleman, M. Schechtman and H. Frankfurt are prominent workers on the practical identity-question.

33 Since the mid-1980s personality psychology has witnessed a strong upsurge of interest in personal narratives and life stories. The turn toward narrative began with Silvan Tomkins and the development of 'script theory'. The most extensively articulated approaches developed in the 1980s are D.P. McAdams' 'life-story model of identity' and H.J.M. Hermans' 'theory of the dialogical self' (McAdams 1999).

It should be emphasized that self-narration presupposes a number of skills. What skills? These are controversial waters. Shaun Gallagher e.g. lists four cognitive capacities for narrative competency: capacity

in the house of human existence – windows at the back that are oriented to what one has been through; and windows directed towards the future. The experienced present, the remembered past, and the anticipated future begin to inform and influence each other; they begin to interact mutually in a way that produces a unified experience. This organising or processing of experiences creates an understanding of oneself as an evolving person. Different elements spread out over time form an ongoing dynamic system – a narrative – that integrates to produce a person that extends over time.³⁴ ‘The formation of an identity-constituting narrative alters the nature of an individual’s experience in a way that extends consciousness over time, producing a persisting experiencer who is the primary experiencing subject’ (Schechtman 1996: 149). So, according to Schechtman, it is the living and telling of an in time extended story (a narrative) that constitutes the self or the person. A narrative unifies experiences and creates a person through processes of attribution and appropriation. While expressing their lived experiences in a connected chronicle, human beings experience a protagonist and identify themselves as this person. They weave stories of their lives and their doing so makes them entering the world of persons.³⁵ To put it metaphorically: persons are like Homer’s

for temporal ordering, capacity for minimal self-reference, episodic and autobiographical memory, and capacity for meta-cognition (2007: 205-11). In *Narrative and Moral Life* Tatjana Tietjens Meyers lists eight skills (among the skills) that contribute to self-narrativity (2004: 301-2). I put these matters to one side here.

34 I’m working here with an unanalysed ordinary understanding of narrative. There is no agreed criterion for sharply defining narratives – and certainly none in terms of necessary and sufficient conditions.

Helena, who – in captivity – was weaving a tapestry showing the battle between the Greeks and the Trojans – exactly that which constituted her identity (*Ilias* 3, 125).³⁶ That said, it will be clear that autobiographical narratives are both fictional and factual – fictional, because we create ourselves; and factual, because we really are the characters whom we create. Our running autobiographies give us considerations that serve as reasons for acting; they guide our speech and our doings. Human lives are narrative quests to live up to self-conceptions.

2.2 Constraints on self-constituting narratives

Of course, not just any narrative will do. A view that any narrative self-conception is identity-constituting would be committed to the obviously false claim that persons cannot be ignorant, mistaken, self-deceived or mendacious about themselves. The narrative self-constitution view avoids this result by placing constraints on the kind of narrative that can constitute a person's identity. Schechtman offers three candidates for constraints on self-constituting autobiographical narratives: the demand for intelligibility, the articulation constraint and the reality constraint.

35 It's important to note that the absence of narrative does not necessarily need to be seen as a negative state, but just as a phenomenologically different one from the having of such a narrative.

36 A similar Homeric example is Ulysses' wife Penelope, who avoided a new identity by drawing her tapestry out into wire every night.

Demand for intelligibility

Individuals constitute themselves as persons by creating self-conceptions that take the form and logic of a conventional, linear narrative. There has to be a coherent plot to produce psychological intelligibility and a well-defined character. Incidents, actions, and experiences do not become intelligible if viewed in isolation, but if they are interpreted as parts of a broader, ongoing story that gives them their significance, i.e. if they cohere with some other elements of the story. The requisite form and logic by which the various parts of the narrative fit together are those associated with the ‘Zusammenhang des Lebens’ (W. Dilthey).³⁷ Narrative self-conceptions look very much like a coming of age-story. Of course, coherence is not a categorical demand. Traits are supposed to support one another, but narratives can be overall intelligible even if there are a few anomalous parts (out of character-elements). Also there is the conflicted story of an adolescent experiencing an identity crisis, or the disjointed story of a person with dementia, or the frozen story of a severely traumatised person. There is no precise measure. The narrative self-constitution view needs to accept a somewhat vague and fluid idea of

37 Philip Roth’s novel *Exit Ghost* (2007) offers a nice pair of complete opposites. Larry, the protagonist’s neighbor, is ‘a meticulous, finicky man who seemed to believe that life was safe only if everything in it was punctiliously planned (...). An only child, Larry was sent to live with relatives on the Naugatuck River southwest of Hartford (...), and there, in a boy’s diary of “Things to Do”, he laid out a future for himself that he followed to the letter for the rest of his life; from then on, everything undertaken was deliberately causal.’ The protagonist himself is desperately cultivating ‘rash moments’ and performing actions that are completely at odds with everything he planned and tried to realise during his life.

a standard narrative. Although a range of narrative styles can fall under the category of a traditional linear narrative, Schechtman also holds that a self-narrative that deviates beyond a certain degree from the form of a conventional story does not constitute a person. The question of just how different a narrative must be before it is removed from the family of narratives making up the standard form, or is ruled out as identity-constituting altogether, is 'largely an empirical one' for her (1996: 105).

The articulation constraint

To have a narrative does not mean to have a fully expressed autobiography. Most people simply live and have no time or reason to articulate full-life narratives. An autobiographical narrative is more than a way of thinking about ourselves in reflective hours: 'How we appropriate actions and experiences to make them part of our consciousness is much more like how we appropriate elements of food to make them parts of our bodies than how we appropriate books to make them part of our library' (1996: 113). So, in the main having a self-narrative is a matter of psychological organisation. Still, Schechtman stresses that an identity constituting narrative should be capable of 'local' articulation. This means that persons should be able to explain why they do what they do, believe what they believe, and feel what they feel. Persons usually can account for their experiences and actions by showing how these are part of an intelligible life story with a comprehensible subject as its protagonist. In other words, the self-narrative constituting one's identity cannot remain totally subterranean. And, according to Schechtman, there is a problem if persons have difficulty articulating their stories; or if they do it inaccurately, i.e., if their reasons for believing what they do may be quite different from what they take them to be ('self-blindness'). Schechtman asks us to think of a man who sincerely

believes he loves his brother, but frequently acts towards him in ways that suggest hostility. This person's hostility may not be part of the *explicit* story the person tells of himself, but it may still be part of his *implicit* self-conception in a very real way as expressed in his behavior, emotions and beliefs. So, Schechtman makes a difference between explicit and implicit self-narratives. In the brother example the implicit and the explicit narrative diverge. The person even explicitly denies what the implicit story tells others. Schechtman: 'Although the hostility is not part of the story this person *tells* of himself, then, it can still be part of his self-conception in a very real way' (1996: 116). Explicitness is not a prerequisite for an element's being part of a person's self-conception. The articulation constraint should not be taken to require absolute transparency. But it also says that elements of a person's narrative he cannot articulate are only partially his - 'attributable to him to a lesser degree than those aspects of the narrative he can articulate. It implies, moreover, that if the inability to articulate one's narrative is sufficiently widespread and severe, the overall degree of personhood can also be compromised' (1996: 116). Because the brother is unaware of his hostility, Schechtman argues, this implicit element is hardly appropriated and only partially his. It is less identity-constituting. Its influence is rigid and automatic (like posthypnotic suggestion); mysterious. But this may not prevent us from thinking that subterranean features of our narratives are revealing of who we are, because they fill in the pieces that can make the incomprehensible elements of our explicit stories intelligible.³⁸

³⁸ Our intuitions often run to the contrary: we are inclined – thanks to Sigmund Freud – to think that forces lurking deep in someone's unconsciousness are more definitive of whom he really is and that the real story often is the one not told.

The reality constraint

The last of Schechtman's constraints is that identity-constituting narratives must match with reality, so as to allow for the contrast with people who are, for example, psychotic, ignorant about themselves, or self-deceived. The facts with which the self-narrative needs to cohere can be split into two basic categories: observational facts (information taken in immediately through the senses) and interpretive facts (conclusions about the meanings or implications of those facts). When my neighbor believes himself to be Charles Darwin he is deeply out of touch with reality. He is making an error of fact and therefore he is not living in the same world as other people, the world in which Charles Darwin died in 1882. Pressing the 'neo-Darwin' by asking him questions will reveal one or another major anomaly in the individual self-narrative. Errors like this decrease the degree of personhood of the individual making the claims.³⁹ Thinking of interpretive inaccuracies Schechtman gives the example of paranoia. Paranoiacs do not make errors about the observational facts, but they see the facts as having implications wildly different from those others make. Their conclusions drawn from the facts are unwarranted. Of course there are light and more familiar cases too: someone believing he is treated unsympathetically, or being taken advantage of, or that his work is appreciated, while it is absolutely not. Such errors of reason as well compromise self-constitution. So, to be a person one needs more than a particular kind of subjective orientation toward one's life. The conception that an individual

39 It is quite uncommon to draw a distinction between various degrees of personhood, as Schechtman does. But having a narrative is a matter of degree. A developing child does not create a self-narrative all at once, nor does a patient suffering from dementia lose narration in one devastating moment.

applies to himself must also mesh with a more intersubjective account of his life – roughly the story that those around him would tell (Schechtman 1996: 134-5). In brief: the reality constraint says that narrative self-constitution is always kept in check by social and cultural practices, and by objective facts about the world. One of these facts not yet mentioned is that persons always are intimately connected with a human body.⁴⁰ Persons literally have their unique place under the sun. They are singular points of view and action. It may well be an empirical fact that all actual personhood is secured for a large part through bodily continuity.⁴¹

⁴⁰ This is also a conceptual point. Human embodiment and self-narrative are conceptually related. Suppose angels have a life story; angelic and human biographies will be as chalk and cheese.

⁴¹ Lynn Rudder Baker also stresses that human persons are constituted by human bodies. They are embodied and embedded (situated in the world). But human persons, in her view, most fundamentally are persons, just as a bronze statue is most fundamentally a statue, not a piece of bronze. The distinguishing mark of persons, according to Baker, is their capacity for a first-person perspective. A first-person perspective is the ability to think of oneself in the first-person without recourse to any name or description or demonstrative, e.g., expressing thoughts as ‘I wonder how I shall die.’ When I speak of narrative selfhood, I am considering a much thicker concept of the self than what Baker means by the first-person perspective. A narrative self is not required for a first-person perspective (as Luria’s famous Zasetzky-case demonstrates). A first-person perspective is necessary, but not sufficient, for an idea of a self in the sense of a coherent and comprehensive life story, of which one is the subject. I agree with Baker that there is no way for an adequate account of a narrative self to avoid the first-person perspective (Baker 2000: 87-88). It makes possible much of what matters to us: inner lives, reflective agency, or self-constitution.

Let me summarise Schechtman's narrative self-constitution view. Human beings characteristically exist in time. They live in the present, relying upon realised possibilities in the past and intending to realise possibilities in the future. They have a special relationship with their own experiences in the present, past, and future. They are time-binding organisms that are reflectively concerned with their own temporal existence. This reflectivity is mediated by narratives. Narrative structure pervades everyday life as the manner in which the flow of experiences and actions is organised over time (Then..., and then..., and then...). And narrative is the organising structure from which a person emerges who experiences and acts. In Schechtman's view, a human being constitutes himself as a person in the unity of an autobiographical narrative. And such person-constituting narratives have to meet three conditions: they must be intelligible, articulate, and real.⁴²

2.3 The missing ingredient

I basically agree with Schechtman's line of thought, although her view is not undisputed. Especially the articulation constraint is

⁴² Closer inspection reveals that Schechtman's three constraints on self-constituting narratives are intertwined. Intelligibility presupposes the reality constraint. Personal experiences can only be intelligible for outsiders against a background of a considerable amount of sense of reality. And how could a self-narrative be intelligible, if not articulated in one way or another? In her 2007 description of the narrative self-constitution view Schechtman does no longer mention the intelligibility constraint. Narrative self-conceptions must meet the reality constraint and the articulation constraint.

much debated. Hilde Lindemann Nelson, for example, argues in *Damaged Identities. Narrative Repair* that ‘in claiming that what we can’t articulate about ourselves isn’t as fully self-constituting as what we can, Schechtman fails to acknowledge that one’s identity isn’t simply a function of one’s self-knowledge’ (2001: 91).⁴³ My thought here is that we have to keep in mind that self-narratives remain partly implicit, unreflected and unarticulated. Otherwise, it is as if I am claiming that if you cannot wear your heart on your sleeve then you have not really got a heart. Neither can we assume that the best selves are to be found on the side of the most explicit of our interpretations. Accepting the articulation constraint does not imply: the more articulated, the better. But above all, we have to accept the fact that persons are not transparent; not for other people and nor for themselves. Schechtman’s description of the ‘self-blind’ brother is an example; it leads to a conflict of interpretations between the story of the brother – let me call him Vladimir – and the story of other ‘experts in normal behavior’, who tell him repeatedly that his behavior

43 I fully disagree with Lindemann Nelson’s conclusion that there is enough reason to jettison all of Schechtman’s constraints on self-constituting narratives and to make a fresh start. Schechtman herself refined the articulation constraint in 2007. She now recognises that narrative approaches to identity correctly have been criticised (e.g. in G. Strawson 2004, or O. Flanagan’s 1996 critique of Charles Taylor’s conception of narrative identity) on the grounds that they tie the conditions of personhood to an improbable degree of articulacy and intellectual reflection. In her 2007 response to Galen Strawson’s critique of narrative identity, Schechtman distinguishes weak, moderate and strong versions of the notion of narrative identity, arguing that strong versions, such as Taylor’s, are susceptible to critique of the articulation constraint. I shall return to the articulation constraint in 2.6.

de facto asks for other interpretations, i.e., other attributions of reasons than he himself can offer. According to the experts, Vladimir's actions speak louder than his words. Vladimir denies.

What I described and endorsed so far as Schechtman's narrative self-constitution view is still under construction. There have been several improvements. The most important one is about the role of emotion in narrative self-constitution. In 2001 Schechtman publishes an article by way of supplement to her narrative theory of persons. She is unhappy with the fact that narrative accounts fail to capture the distinction between psychological changes which are survival-threatening, that is, which bring about a loss of identity, and those which are benign, that is, those which bring about identity-preserving development. The mere existence of a comprehensible narrative of change is not yet enough to preserve identity. That self-narratives meet the coherence requirement – internal coherence and 'external' coherence with reality⁴⁴ – is a necessary condition, but not a sufficient one

⁴⁴ The crucial requirement for self-constitution is that narratives be coherent. Self-constituting narratives require internal coherence, i.e., the demand for intelligibility, and 'external' coherence, i.e., the reality constraint. Apart from 'coherence' Schechtman also uses 'cohesiveness', 'connectedness', and 'consistence'. Consistency usually means that there is no logical contradiction between the parts of a whole. Coherence is a broader concept. It includes connectedness and consistency. Coherence means that parts do go together very well. In Schechtman's narrative self-constitution view coherence establishes the norm in two senses of this term. Firstly, coherence is normal; it obtains for the most part. For most of us, most of the time, things do make sense, hang together, and are going somewhere. And secondly, it is the normative standard, which also determines that which deviates from it.

to identify the characteristics of identity-preserving change. The missing piece, Schechtman argues, is *empathic access* (2001: 101): ‘Empathic access involves a situation where the original psychological make-up is, in an important sense, still present in the later, psychologically-altered person. The earlier beliefs, values and desires are recognised as legitimate, and are given, so to speak, a vote in personal decision-making’ (2001: 102). Memory-connection, e.g., not only is a cognitive relation to the past, but one which is thoroughly infused with affect. What is at the core of empathic access to memories is the affective connection to the past, together with its behavioral implications. More than just cognitive recollection, the emotions that belong to the past experiences are still there. Schechtman thus redefines her narrative self-constitution view as follows: ‘The most satisfying view of personal identity will be a combination of a narrative view with empathic access. The basic idea would be that personal identity over time consists of the existence of a coherent narrative of change which includes empathic access. The narrative provides a basic level of continuity while the empathic access provides the additional ingredient necessary for true personal survival’ (2001: 111).

The point clearly echoes Schechtman’s interpretation, in *The Constitution of Selves*, of John Locke as a self-constitution theorist (1996: 105-14). According to Schechtman Locke provided an insight that yields a plausible understanding of how persons create an autobiographical sense of self. Personal identity, in Locke’s view, is constituted by sameness of consciousness. This claim requires interpretation and it is common to view Locke as holding a memory theory of personal identity. Unfortunately, interpreters of Locke are focusing exclusively on the cognitive aspects of consciousness. Consciousness is thought of as a faculty of knowing. This is not, however, the aspect of consciousness that Locke most emphasises, says Schechtman. The aspect of consciousness

that Locke most stresses in his discussion of personal identity is the affective side. He paints a picture of consciousness as the faculty whereby we experience pleasure and pain, happiness and misery (1996: 108). Locke draws a connection between the affective aspects of consciousness and the appropriation of particular actions and experiences when he tells us what makes a particular body (in the present) one's own. It is the fact that someone experiences what happens to certain bits of matter – that their condition directly affects him in the dimension of pleasure and pain – that makes them part of his consciousness, and so makes them his. The same, says Locke, is true of immaterial substance. Insofar as someone is immediately caused pleasure or pain by the present states of a particular soul, that soul is, presently, hers. Present actions and experiences are made part of a person's present consciousness by affecting his well-being or causing him pleasure or pain. Locke tells us that persons make past actions and experiences theirs on just the same grounds. So, in his view, past actions and experiences become those of a present person if they affect present consciousness, causing the person pleasure and pain in the present. On this reading, we extend consciousness back in time to some past action or experience by caring about it in the appropriate way – by feeling its effects.

The objection may now be raised that it still necessarily is the *memory* of some past event that causes the affect. Schechtman denies. There is more to the story, Schechtman argues⁴⁵, because

45 A disagreement with Korsgaard here. According to Korsgaard's view, until a person consciously incorporates his unconscious nature into his own reflective view, c.q. into his unified identity, that unconscious nature cannot be fully part of the person. See her 2009 comments on Jonathan Lear's Tanner Lectures *Irony and Identity* at Harvard University (2010b).

we have known since Freud that the past can affect present well-being without being explicitly remembered: ‘Locke points out that past events can become part of present consciousness by affecting us in the present along the dimension of pleasure or pain. This can happen in a straightforward way – as when some memory causes guilt or remorse – but it can also happen in a more subtle way, by contributing to a person’s overall sense of self. Our pasts give us our conception of who we are and what life story we are living, and this conception may well intensify or mitigate present emotion. (...) To have a narrative self-conception on the view I am urging is thus to experience the events in one’s life as interpreted through one’s sense of one’s own life story, and to feel the affect that follows upon doing so’ (1996: 112). Locke seems to imply that the future (anticipations) does so as well.⁴⁶

Schechtman’s emphasis on the emotional aspect of narrative self-conception, to my way of thinking, is right. Moreover, it

⁴⁶ More authors are emphasising the emotional component in self-narratives. Peter Goldie is a good example in his *The Emotions: A Philosophical Exploration* (2000); in *Narrative and Perspective: Values and Appropriate Emotions* (2003); or in *Narrative Thinking, Emotion, and Planning* (2009). And Owen Flanagan, providing his readers with a brief sketch of his own life and listing five general features of a self-narrative in *The Problem of the Soul*, says: ‘It conveys a certain way it feels to be me, which is not in every case explicitly stated (nor, perhaps, can it be), but that is sensed by both you and me. What I am able to say about my life comes with emotional coloration’ (2002: 232). Schechtman’s narrative view is essentially Owen Flanagan’s: Persons are beings-in-time; so life-stories have temporal structure. What persons are able to say about their lives comes with emotional coloration. Self-constituting narratives are abstractions; they consist of memories, surmises and interpretations (2002: 226-235).

draws attention to the bodily anchoring of self-narratives. At this point, some insights of Antonio Damasio, a neuroscientist, may clear a further path.⁴⁷

2.4 The feeling of what happens

As a neuroscientist, Antonio Damasio asks: what are the biological circumstances that permit consciousness of a self? Answering this question is anything but easy, but Damasio thinks that there are possible solutions. Starting from empirical data, he outlines the following hypothesis.

Human consciousness is not monolithic. We can distinguish simple *core consciousness* from complex *extended consciousness*. These two types of consciousness correspond with two types of selves: a *core self* and an *autobiographical self*. In addition, Damasio presupposes a third kind of self, called the *proto-self*. Each self depends on its predecessor. The autobiographical self rides on top of the core self, which rides on top of the proto-self.

Damasio begins by stating two fundamental assumptions behind his picture of the self. First, unlike cells in the liver or

⁴⁷ Two remarks here. Firstly, I do not think that philosophy must simply defer to science. However, when a philosophical thesis has empirical implications, and the empirical implications are *prima facie* plausible, then I think the empirical implications speak in favor of the philosophical thesis. Secondly, let me emphasise that what science teaches us cannot be determined from outside the space of reasons. Any attempt to take up a scientific third personal position is itself a first personal intentional act.

lung, which perform their assigned biological business but do not represent any other cells or business, neurons in a brain represent objects or events occurring elsewhere in the organism. Neurons are commanded by biological design to be about other cells and other actions. They are born cartographers of the geography of an organism and of the events that take place within that organism. Second, the brain possesses devices (located in the brainstem nuclei and the hypothalamus and the basal fore-brain) that are aimed at managing the life of the organism in such a way, that the internal chemical balances indispensable for survival and well-being can be maintained continuously. The brain devices that regulate the state of life therefore represent, of necessity, the continuously changing states of the organism as they occur incessantly. Those devices have the natural means to represent the entire anatomy and the current functional state of the whole living organism. So, Damasio's point of departure is that the self has a biological foundation that can be found in the brain devices which represent moment by moment the continuity of the same individual organism interacting with the environment (2000: 18-20).

Let me now discuss in more detail the three selves Damasio wants to discern. The *proto-self* is unconscious and constituted by the 'interconnected and temporarily coherent collection of neural patterns which represent the state of the organism, moment by moment, at multiple levels of the brain' (1999: 174). Key aspects of the state of the organism are: the state of the internal milieu, viscera, vestibular system, and musculo-skeletal frame. We can imagine a lobster or a human fetus in the first trimester as possessed of a proto-self.

The *core self* is conscious. Dogs, cats, and human infants have core selves. The core self inheres in the second-order, non-verbal account that occurs whenever an object modifies the

proto-self. Now the organism, plus the external object (say, a face or a melody) or the internal object (say, a toothache), plus the reaction modifying the state of the organism are mapped as neural patterns. The biological essence of the core self is the representation in a second-order map of the proto-self being modified. The mechanism at work here is the generation of an imaged, non-verbal representation of the object-organism relationship, that is, a 'wordless narrative' – with protagonists (the organism and the object) and a period of time (begin, middle, end) – about the state of the organism, the arrival of an object, and the changing state of the organism.⁴⁸ The mechanism has two clear consequences. The first consequence is the enhancement of the image of the causative object. This object is, so to speak, set out from less-fortunate objects, made pregnant, and becomes a 'fact'. The second consequence is that a sense of self emerges from being able to represent the modification of the proto-self; one acquires a core self, that is, a transient protagonist of consciousness, generated for any object that provokes the core-consciousness mechanism. Any organism possessed of a core self is a subject of experience. Because of the permanent availability of provoking objects (the core self can be triggered by any object and the objects never run short) the core self is continuously generated and thus appears continuous in time. Actually, it is created pulsatively

⁴⁸ Several problems haunt the idea of mental imaging. Bennett and Hacker criticise Damasio's representationalism in the seventh chapter of *Philosophical Foundations of Neuroscience*. Damasio admits that we have no idea how neural patterns can be transformed into mental representations, into a 'movie-in-the-brain'. Here, he concedes, human empirical knowledge still is dramatically lacunal.

and rebuilt again and again out of a stream of ‘narratives’ as a series of ephemere appearances. The mechanism of production of core self undergoes minimal changes across a lifetime. Damasio is quite sure ‘that virtually all of the machinery behind core consciousness and the generation of core self is under strong gene control. (...) The genome puts in place the appropriate body-brain linkages, both neural and humoral; lays down the requisite circuits, and, with help from the environment, allows the machinery to perform in reliable fashion for an entire lifetime’ (1999: 229).

Let me make two final remarks on the core self. First, note that Damasio does not invoke the infamous homunculus-in-the-brain, a ‘mini-me’ who interprets the situation. The second-order representation occurs in neural structures such as the thalamus and the cingulate cortices. It presents within the mental process the information that the organism is the owner of the ongoing mental process. It creates a sense of a first-person, a sense of self. Returning to the well-known metaphor of the mental process as a movie-in-the-brain, Damasio’s solution is that the sense of self emerges within the movie, so that it is a part of the movie. There is no need to posit a spectator for the movie-in-the-brain. The first-person subjectivity is constructed from the ground up, based on the same kind of sensory mapping needed to construct the object representations made conscious by the sense of self.

A second and, for our purposes, more important remark is that the sensory mappings of self are imagined as *feelings*. The sense of self emerges as a special kind of feeling, the feeling of what happens in an organism caught in the act of interacting with an object.

Let me now turn to the extended self. The extended, *autobiographical self* requires maturation of autobiographical memory, which probably requires, at least in the human case, that one

have language or be in the process of acquiring it.⁴⁹ The autobiographical self is constituted by memories of individual experience of the past and of the anticipated future. These memories are a residue of records of core self experiences that became dispositional. Those records can – as internal objects – be activated as neural patterns and turned into explicit images. So, a creature with autobiographical memory partakes of more than just passing states of consciousness. It can see and hold its life, or segments of its life, in view. Invariant aspects of an individual's biography form the basis of autobiographical memory that grows continuously with life experience and can be partly remodeled to reflect new experiences. Sets of memories which describe identity and person can be reactivated as a neural pattern and made explicit as images whenever needed (see note 48). Each reactivation of representations operates as a 'something to be known' and generates its own pulse of core consciousness. The result is the autobiographical self of which we are conscious. So the autobiographical self requires the mechanism of core consciousness in two ways. Firstly, it requires the presence of a core self

⁴⁹ Damasio does not mean to suggest that only creatures capable of acquiring humanlike language can possess autobiographical selves. Chimps and bonobos all have trouble learning human language but presumably utilise 'concepts' in self-comprehension. Note that Damasio does not place consciousness, either in its core or extended levels, at the pinnacle of human qualities. Consciousness is necessary, but not sufficient, to reach the current pinnacle. At the end of the chain, extended consciousness permits conscience. That is what we can find at the top of the complexity heap outlined by Damasio (1999: 230). Unfortunately, the author does not explain why this 'inner voice that warns us somebody may be looking', as H.L. Mencken defined conscience, is so special.

to begin its gradual development and, secondly, each reactivated memory only can be realised by generating its own pulse of core consciousness (1999: 174). According to Damasio, autobiographical memory develops and matures under the looming shadow of an inherited biology. However, much will occur that is dependent on and regulated by the social environment. The autobiographical self is the brain state for which the cultural history of humanity most counts. Self-interpretations are always embedded in larger historical and communal meaning-giving structures.

Like Schechtman, Damasio opts for a modest understanding of the narrative self. Narratively constituted selves are coherent and stable, if less than fully unified and consistent constructions. All in all, we do not have a self sculptured in stone and, like stone, resistant to the ravages of time. We are more like the sand castles on the beaches of our childhood. No component remains the same for very long. Nothing is sacred. In Damasio's words: 'Our sense of self is a state of the organism, the result of certain components operating in a certain manner and interacting in a certain way, within certain parameters. It is another construction, a vulnerable pattern of integrated operations whose consequence is to generate the mental representation of a living individual being. The entire biological edifice, from cells, tissues, and organs to systems and images, is held alive by constant execution of construction plans, always on the brink of partial or complete collapse should the process of rebuilding and renewal break down. The construction plans are all woven around the need to stay away from the brink' (1999: 145).⁵⁰

⁵⁰ More on the narrative self and the core self (or 'minimal self' as he calls it) in S. Gallagher 2000 and in S. Gallagher & D. Zahavi 2008 (ch. 10). A defense and analysis of the experiential dimension of selfhood

What is the upshot of all this for a narrative view on personhood? Let me draw two lessons. Damasio's theory describes the biological antecedents of the self and the neural circuitry for each developmental stage of the self, including the autobiographical self. His informed speculations make clear, in the first place, that narrative selfhood is always embodied selfhood. Narrative selfhood builds on a more primitive and foundational core self, which consists of being pre-reflectively aware of the first-personal appearance of experiential phenomena. The core self is the invariant dimension of first-personal givenness in the multitude of changing experiences. And this self-experience has

[*continuation of note 50*] – the core self – also can be found in phenomenological literature, for example in the work of Edmund Husserl, Michel Henry, and Maurice Merleau-Ponty. This experiential dimension of selfhood is what phenomenologists often call the sense of *ipseity*, i.e., the sense that an experience is my experience (the 'mineness' of experience). Shaun Gallagher (2000) very insightfully delineates three different levels of explanation. There is a level of non-conscious, sub-personal neural processes. There is a level of pre-reflective phenomenal experience of agency. On this pre-reflective conscious level we experience a sense of ownership (the 'mineness' of the experience) and a sense of agency (the sense that I am the one causing the experience) that is tied to bodily movement/motor control and intentional aspects (goals, tasks) of action. In the normal phenomenology of voluntary or willed action, the sense of agency and the sense of ownership coincide and are indistinguishable. In the case of involuntary action, however, it is quite possible to distinguish between the two. E.g., when someone pushes me from behind I have a sense of ownership for my bodily movement; in this case there is no sense of agency, a sense that one does have in the case of intentional action. Finally, there is a sense of agency based on reflective attribution.

to be taken as the self-experience of a world-immersed embodied agent. What is required for self-emergence to take place is an organism with a body that houses a certain kind of brain to live in the natural world with other similar ecologically embodied creatures. A single organism requires that there be one single self, if the job of maintaining life is to be accomplished successfully – more than one self per organism is not a good recipe for survival flourishing. For every person that we know, there is a human body. We never meet a person without a human body. Nor do we meet a person with multiple bodies. It just does not happen. And it cannot happen. One person, one human body – a first principle.⁵¹

The second and related lesson is that Damasio's information seems to do a lot for Schechtman's point that self-consciousness and emotion are intertwined. Although we couch our self-descriptions in languages, our selves are not exhaustedly captured by what we can say about them. There are many things we feel about ourselves, that we cannot put easily into words. However, this is a point of hot debate in the narrativist camp and calls for a little more discussion. Before returning to this debate, let me make a brief digression into the philosophy of mind and add one last general and theoretical point about the role of the body in narrative continuity. The point touches upon Damasio's claim that even on the level of the core self something like narrative self-constitution is involved.

⁵¹ Even in the much discussed case of Dissociative Identity Disorder, the principle is not quite violated since, at any given time, only one among the 'alters' can use the body to think and behave.

2.5 Basic narratives

In *The Diachronic Mind*, the philosopher Marc Slors develops a conception of psychological continuity in actual persons that requires bodily continuity (2001, chapter 4). To refer to the kind of psychological continuity he is looking for, he uses the term ‘narrative continuity’ (N-continuity). The connections he is after are such that psychological contents cannot really be considered self-contained items. They are essentially part of series or sequences of contents. They are essentially part of a psychological process: ‘Individual thoughts acquire their full meaning only as part of a process of deliberation, individual sense perceptions acquire their full sense only as part of a sequence of perceptions portraying a body’s movements through space, individual feelings acquire their full sense only in connection with what evoked them and what they produce, etc. It is the whole of such processes that endow their ‘parts’ with their full meaning. Narrativity refers to relations that hold between events in virtue of which these events can be considered particulars only by abstracting them from the whole of a diachronic process of which they are a part, a process that portrays a ‘story’, in the broadest sense of that term. (...) Narrative continuity between ‘particular’ psychological states is a relation between these states such that one or more preceding states are a necessary prerequisite for another state’s full content and the intelligibility or ‘logic’ of its occurrence’ (2001: 89).

While asserting that practically all of our psychological contents are N-continuous⁵², Slors does not mean to say that our psychological lives are completely logical, coherent, fluid, and

⁵² Note that Slors uses the term narrativity at a sub-personal level here; narrativity applies to strings of consecutive mental states that are only parts of the total of a person’s stream of consciousness.

consistently structured wholes. He is aware of the fact that our mental lives often seem fragmentary, patchy like a quilt, and gappy. However, despite the fact that our psychological lives are sometimes gappy and patchy, they occur against a relatively stable background of an ongoing narrative. There is, as Slors is calling it, a *basic narrative* that is neither gappy nor patchy. This basic narrative is presented – in the case of actual persons – by their consecutive perceptual contents. No doubt other types of contents also play a unifying role, but Slors concentrates on perceptual contents because their unifying role is a relatively transparent one.

To illustrate the point, let me take Harold Noonan's simple example of two non-causally related, qualitatively dissimilar perceptions: At present I have an impression of a desk top covered with sheets of writing paper. If I turn my head to the right, I have the impression of a book case filled with books. The impression of the desk top neither resembles nor is the cause of the impression of the book case (nor is the desk top itself a cause of the book case); yet I regard both impressions as mine. How come? The different impressions described belong to the mind of one person (me) because they acquire narrative coherence in virtue of the fact that I know them to be caused by a body's movements through a stable (not static) physical world with whose character and proceedings I am acquainted. A person's mental life is set against the background of and is co-determined by a basic narrative which reflects the career of an objective continuant in and through physical and social reality – the body.⁵³ Perceptual narratives are 'basic' in the sense that they provide a lifelong narrative. The fact that the one body upon which consecutive perceptual contents are dependent is a physical object whose

⁵³ This example obviously is just a tiny piece of what Slors labels someone's 'basic narrative'.

movements are subject to the regularities of the physical world guarantees the narrative coherence of these contents. This is not to say that a basic narrative in itself is sufficient for full-blown psychological continuity. It is not. A basic narrative needs to be supplemented with emotions, desires, thoughts, motives for actions, values, &c. for that. Conversely, however, a basic narrative secures a degree of diachronic coherence between one's thoughts, emotions, desires, and so on, that is sufficient for these mental states to be narratively continuous in at least one elementary sense.

As will be clear now, Slors ascribes a crucial role to the human body in the narrative coherence of successive perceptual contents of actual persons and I am of one mind with him. Despite the fact that many of our thoughts, deliberations, associations, and so on, frequently appear dissociated, most of the contents associated with one body contribute to an elementary narrative thread running through our 'stream of consciousness'. In the case of real people, this is due to the fact that such processes occur against the stable and continuous background of a perceptual narrative, relating the story of a body's movements through space-time. This basic narrative acquires its continuity from the intelligibility of objective physical reality and our bodies being part of it (2001: 110).⁵⁴

⁵⁴ Slors takes it as an empirical fact that the diachronic unity of mind of all actual persons is secured for a large part through bodily continuity. He does not take it as a conceptual truth. Slors sees no conceptual link between bodily continuity and the diachronic unifying role of the basic narrative. This becomes clear to him when he turns from actual psychological continuity to hypothetical cases like brain-transplantation (2001: 103-5). I leave aside possible worlds where brain-transplants or 'body swaps' are a live possibility. Science-fiction cases aside, bodily continuity is a precondition for psychological continuity in the real world.

What Slors means by ‘basic narrativity’ is a far cry from my narrative conception of practical identity. That does not alter the fact that common-or-garden narratives build on Slors’ basic narratives, as Damasio also tells us, and that Slors (partly) can explain how the human body plays a role in tying up a personal life story. With this theoretical point in place⁵⁵, let me return to the main theme and examine some difficulties within the narrative self-constitution view.

⁵⁵ Narrativity is gaining momentum in the field of consciousness studies. According to Owen Flanagan narrative plays a role in the formation and function of consciousness and can be provided with a credible naturalistic analysis. For growing empirical evidence of images of the functioning brain for a neurophysiological basis of personal narratives see e.g. J. Bickle, ‘Empirical evidence for a narrative concept of Self’ in Flanagan O., G. D. Fireman & T. E. McVay, jr. (2003), *Narrative and Consciousness. Literature, Psychology, and the Brain*, Oxford: Oxford University Press, 195-208. For a thought-provoking narrative approach to folk psychology see D. Hutto 2007 and 2009. Hutto’s Narrative Practice Hypothesis says that our capacity for a sophisticated, everyday understanding of intentional actions as performed for reasons (i.e. our folk psychology) may itself best be characterised as a kind of an essentially narrative practice. And that acquiring the capacity for applying reasons might – at least normally – depend upon having received training with narratives. Children acquire the relevant interpretative skills for achieving this through repeated encounters with specific kinds of narratives, when they are appropriately supported by others. With his Narrative Practice Hypothesis Hutto presents a challenge to ‘theory theory’ and ‘simulation theory’ in the philosophy of cognitive science. A final example: David Herman in *Storied Minds* (2009) underscores the relevance of narratives for the philosophy of mind – and vice versa – by interpreting narrative as a system for building models of action.

2.6 Trouble for narrativists

As Galen Strawson (2004) and Peter Goldie (2004) pointed out, narrativists like Jerome Bruner, Paul Ricoeur, Alisdair MacIntyre, or Charles Taylor are pushing the idea of self-storying too hard.⁵⁶ These authors believe that we just are the narratives we tell or could tell about ourselves. This might seem right for some reasons: self-constituting narratives are open-ended constructions; they involve all manner of revisitation and reconstruction *post facto*; they are pinned on culturally relative themes; they express ideals of what one wishes to be but is not yet, &c. Even so, these philosophers over-emphasise the essentially constructed, linguistic nature of narratives. Narrative self-constitution is different from self-fictionalisation in that there must be a sense in which one's self-storying should be accurate or realistic. Extreme fictionalist versions of the narrative self-constituting view are not open to criticism by comparison with reality. But realism about personhood matters. We want to think of ourselves and our lives as they really are. We want a 'reality check'. Extreme fictionalists do mistake life for art. They tend to push what actually happens to the background. They forget that there are lots of facts about one's character and history that do not depend on one's creation.

For Galen Strawson this answer does not settle the point. Strawson is much more sceptical about the narrative view. In *Against Narrativity* he argues against the dual, i.e., psychological and ethical claim that we are, and ought to be, constantly

⁵⁶A. MacIntyre (1981: 203-4); C. Taylor (1989: 47-52); and P. Ricoeur (1992: 158). Bruner has been putting forward the narrative view since the 1980s and recently in J. Bruner (2002). *Making Stories: Law, Literature, Life*, New York: Farrar, Straus and Giroux.

engaged in making a tale out of ourselves and our lives. The dominant view in the academy, Strawson tells us, is that the dual claim is true. Strawson regrets this current acceptance and argues that the claim is false.⁵⁷ He explicitly makes Schechtman's narrative self-constitution view as a good example of the objectionable standard view his target. So let me elaborate his line of attack a bit more.

Strawson characterises the narrative view as a combination of four features: D, F, S, and R. He correctly classifies Schechtman's account of narrativity as concerned with all four (446). It assumes that we are all diachronic (D). It requires that we be form-finding (F) and story-telling (S) and explicitly so. And it is important, in Schechtman's view, that there be no significant revision (R), that one's self-narrative be essentially accurate.

To clarify the four features: the basic form of diachronic (D) self-experience is that as a Diachronic 'one naturally figures oneself, considered as a self, as something that was there in the (further) past and will be there in the (further) future (...). If one is Episodic, by contrast, one does not figure oneself, considered as a self, as something that was there in the (further) past and will be there in the (further) future' (430).

The paradigm of a narrative is a conventional story told in words. The term attributes a certain developmental and hence temporal unity or coherence to the (parts of) lives to which it is applied. The narrative attitude must 'engage in some sort of construal of one's life. One must have some sort of relatively large-scale coherence-seeking, unity-seeking, pattern seeking, or

⁵⁷ For his defense of the claim that a non-narrative life can be a fully moral or human life see Galen Strawson's paper *Episodic Ethics* in Hutto 2007: 85-115.

most generally [F] form-finding tendency when it comes to one's apprehension of one's life' (441).

If one is genuinely narrative one must also 'have some sort of distinctive [S] story-telling tendency when it comes to one's apprehension of one's life (...). According to this view, one must be disposed to apprehend or think of oneself and one's life as fitting the form of some recognised narrative genre' (442).

Finally, if one is narrative one will also have 'a tendency to engage unconsciously in invention, fiction of some sort – falsification, confabulation, revisionism – when it comes to one's apprehension of one's own life. I will call this [R] revision' (443). Strawson has no doubt that almost all human narrativity is compromised by revision, but he does not think it must be. Although autobiographical memory is an essentially (re)constructive phenomenon, humans are not without exception incorrigible self-fabulists. 'When Bernard Malamud claims that "all biography is ultimately fiction", simply on the grounds that "there is no life that can be captured wholly as it was", there is no implication that it must also be ultimately untrue' (446).

Now, Strawson takes himself to be a non-form-finding, non-story-telling, non-revising Episodic. On the strong form of Schechtman's view, Strawson concludes, he is not really a person. Strawson strikes back by criticising Schechtman's articulation constraint: 'This seems to me to express an ideal of control and self-awareness in human life that is mistaken and potentially pernicious. The aspiration to explicit Narrative self-articulation is natural for some – for some, perhaps, it may even be helpful – but in others it is highly unnatural and ruinous. My guess is that it almost always does more harm than good – that the Narrative tendency to look for story or narrative coherence in one's life is, in general, a gross hindrance to self-understanding: to a just, general, practically real sense, implicit or explicit, of one's nature.

(...) The more you recall, retell, narrate yourself, the further you risk moving away from accurate self-understanding, from the truth of your being' (447). So, for Galen Strawson narrativity is not a necessary part of 'the examined life'. It is even unclear for him that the examined life is always a good thing: 'People can develop and deepen in valuable ways without any sort of explicit, specifically Narrative reflection, just as musicians can improve by practice sessions without recalling those sessions. The business of living well is, for many, a completely non-Narrative project. Granted that certain sorts of self-understanding are necessary for a good human life, they need involve nothing more than form-finding, which can exist in the absence of narrativity; and they may be osmotic, systemic, not staged in consciousness' (448). Strawson reassures those who may now think that the Episodic life must be deprived in some way: 'truly happy-go-lucky, see-what-comes-along lives are among the best there are, vivid, blessed, profound' (449).

I think Strawson is on the right page, when he criticises overstatements of the structure of actual lives and the articulation constraint in particular. Schechtman for her part modified her view in response to some of Strawson's challenges (Schechtman 2007). Strawson's argument is against narrative views in general and Schechtman now only wishes to defend a very particular narrative account that avoids Strawson's objections. She distinguishes between weak narrative views, middle-range narrative views and strong narrative views. At one end of this spectrum the rather weak requirement is that a person's narrative must somehow operate to impact his current experience. The person's narrative need not be in any way accessible to consciousness. Having a narrative is necessary to function at all. In the middle range, having a narrative would require that a person be able, at least sometimes, to become conscious of his narrative and make

it explicit. Now, having a self-narrative is necessary for engaging in complex person-specific activities like prudential reasoning, autonomy, and moral agency. Someone needs a certain understanding of how the events in his history hang together. Such an understanding is mostly implicit but is locally accessible where appropriate, if the person is to be able to engage in the person-specific activities on which he places great importance. Strong narrative views require that in order to have a narrative in the relevant sense a person must actively and consciously undertake to understand and live his life in narrative form with a unifying theme or grand telos, and little or no extraneous material. Schechtman is only interested in the middle ranges (2007: 161). The extremes do not represent the refined narrative self-constitution view Schechtman wants to defend. There is no debate about extremely 'edited' versions of the narrative self-constitution view: Schechtman is happy to concede to Strawson that the strong narrative views are too strong. And although she thinks the weak views are true and interesting, she is also willing to allow that some might find it a stretch to call such views *narrative* views.⁵⁸ Even if this point is granted, though, there remain puzzles and disagreements with Strawson's anti-narrativist challenge.

First, on the truthfulness of self-narratives. Persons normally care about truth and take reality into account, say the facts of the world and the interpretations of others. Strawson agrees, but at the same time he seems to think that the prospects for truth are not good for the narrators among us. I side with Strawson that story-telling can lead, even as a psychological and neurophysiological inevitability, to alteration. But it is a bit of a mystery

⁵⁸ I'm not one of those who find it a stretch.

for me why the alteration business has to lead us to the sceptical conclusion that the less we narrate ourselves, the better. Maybe the more we narrate ourselves, the further we risk moving away from accurate self-understanding and from the truth of our being. But alterations also can go in the direction of better self-understanding and coming closer to truth. Or so it seems to me.

Here is a second reply to Strawson. In fairness to Schechtman, it must be noticed that Strawson's claim that Episodics cannot join the person's club is untenable. It is true that Schechtman's view implies that there are *prima facie* advantages for a person to be strongly connected with the whole of his narrative, a tightly woven self-narrative making for a stronger person than a weaker one. There is value in seeking to maintain connection to as much of a person's life as possible. On the other hand, there is always disunity and incoherence.⁵⁹ Life is not perfectly unified; it is contingent, chancy, incomplete, messy, unarticulated. And personhood is a matter of degree. In Schechtman's basic picture of personhood there is also room for minimal, or punctual persons like Galen Strawson. Admittedly, he is a borderline case. But I do not see why a picaresque, rhapsodic and disjointed life cannot be conceived as a narrative genre and lead to narrative self-understanding and personhood.

My last rejoinder is on the 'deep division' Strawson reveals in the human species.⁶⁰ On one side, the narrators: those who are self-storying, 'Homeric', in their sense of life and self. On the

⁵⁹ For a critical discussion of narrative unity as a condition of personhood see J. Christman 2004.

⁶⁰ Also in his review of Jerome Bruner's work in *The Guardian* of January 8, 2004.

other side, the non-narrators: those who live life in a fundamentally non story-telling, solely form-finding fashion. In dividing the human species so dramatically Strawson exaggerates, I think, the importance narrativists attach to being articulate. A reminder is important here. Like Schechtman, I think so called ‘weak’ narrative accounts of personhood – the claim that having a self-narrative is necessary to function at all – are true and important. In the narrative account I advocate ‘weak’ narrativity and ‘middle range’ narrativity are made in one piece. Self-narratives are not exclusively conceived as a linguistic-cognitive phenomenon; they are co-constituted by emotive and conative components as well. Self-narratives are constrained by and consistent with developmental psychology and human neurobiology as well as linguistic structures and narrative traditions. To put it simply: self-narratives are strongly influenced by the body. They depend on somatic input. Persons are embodied beings with rich bodily feelings and emotions that are incorporated into their self-narratives.⁶¹ So, my own view on the division between

61 A good place to look in the philosophical literature for a theory that emphasises the role of bodily feeling is the work of William James. Antonio Damasio presents a modern-day version of some core ideas of William James’ project. Let me add here an interesting observation of David Velleman on emotion and narrative. In *Narrative Explanation* (2003) Velleman makes an issue of the application of narrative in practical reasoning. His question is whether stories help us make sense of things because they convey causal-psychological information or, alternatively, because they have an explanatory force of their own. His answer is that stories exert their own explanatory force that results in a radically different kind of understanding. A stretch of discourse qualifies as a story by reliably initiating and resolving an emotional sequence in us, by ‘completing an emotional cadence in the audience’. Stories have an emotional structure that enables

narrators and non-narrators is that it is rather shallow. The narrative attitude includes much more than telling a life-story. It is as much a way of experiencing and organising a life as a way of storying it. It does not merely capture (aspects of) the person for description, communication, and examination; it constitutes the person. It is one thing to say that our lives cannot be fully articulate, or, as the short story writer V.S. Pritchett observed, ‘that we live beyond any tale that we happen to enact’. It is quite another to hold that narrativity is completely merged with articulation. I do not think it does. To my mind, narrativity is an inescapable structural requirement of human life. Narratives structure our conscious experience and are deeply rooted in the material reality of the human body. Articulation is only part of it.

us to assimilate events to familiar patterns of how things feel. We have an experience of *déjà senti*, because our emotional sensibility naturally follows the ups and downs of the story, and arrives at a stable attitude toward them overall. We understand what the narrated events mean in emotional terms. According to Velleman, we should not discount the importance of narrative understanding. The narrative intelligibility of episodes in our lives is what gives them meaning. Intellectual self-understanding (rationality) and emotional self-understanding (meaning) usually come hand in hand. In a telling footnote, Velleman borrows from Ronald de Sousa the idea that the earliest stories in our lives are about the vicissitudes of our emotions; and that the shape of those stories is determined, in the first instance, by the nature of human affect. The pattern is biologically programmed. We understand stories viscerally, with our bodies (De Sousa himself famously analyses emotion in terms of narrative, hypothesising that each human emotion has a ‘paradigm scenario’). Velleman also stresses that emotions are essentially diachronic. Their nature consists in how they unfold over time and it ultimately must be defined in terms of the arousal and resolution of emotion.

2.7 Interim conclusions

What I am after in this dissertation is a theory of persons in terms of agency. In chapter one I scrutinised the Korsgaardian account of self-constitution. Christine Korsgaard showed us that becoming a person is bound up with reflective agency and so with practical reason. I partly rejected her reason view on persons and proposed to think differently about the normative powers of reason. My proposal came in two parts.

Firstly, we have to move to a less principled, instrumental role for reasons that are constitutive for personhood and the unification of personhood. Believing we need unifying reasons in a strongly Korsgaardian, formal sense to be persons is a mistake, – we can be persons with far less. By implication, we have to weaken the unification requirement. We better take reasons as an incipient story about us, that is, as a ‘rationale’ that makes our actions intelligible as coherent developments in a life story. Therefore, and this was the second part of my proposal, I suggested to transform the reason view on self-constitution into a narrative self-constitution view. The promise of narrative theories of personhood was that they can avoid empty formalism and work with substantive reasons; that they can accept coherence as a requirement for personhood and that coherence does better as a standard for self-narratives than unification; and that narratives can be better linked up with the embodiment of personhood.

In the second chapter I tried to develop such a narrative view on persons in the wake of Marya Schechtman. A person’s identity is now to be found in narrative activity, in the capacity to keep a particular narrative going. Such a self-narrative organises and integrates disparate elements of human experience and action synchronically and diachronically into a more-or-less coherent storied whole and in this way constitutes a person. Internal

coherence, being in step with the facts of the world, and consensus with interpretations of others are important standards for self-narratives.

A Schechtmanian narrative self-constituting view has many advantages over the Korsgaardian picture. It avoids the Korsgaardian overestimation of the formal, unifying powers of reason and settles for the more substantive, vernacular criterion of coherence of self-narrative. Coherence is meant here in a not too rigid and demanding way. In Schechtman's view, unified selves or omnibus editions of the self are not in the foreground.

A narrative self-constituting view is also a substantial improvement of the Korsgaardian approach, because it clearly takes self-constituting narratives as embodied and embedded structures. Persons are incarnated beings. Korsgaard's account of the role of the body was not very informative and possibly mistaken. To put it in terms of political constitution (her favorite analogy): in Korsgaard's view it seems to be perfectly possible to constitute different states within one and the same territory. However, such a picture underestimates the role of the territory, if we apply it to the realm of the constitution of selves.

To unpack this idea of the embodiment of self-narratives, I have taken an empirical line. Antonio Damasio demonstrates that self-constituting narratives have a bodily, neuro-physiological basis. They build, *inter alia*, on the presence of a core-self. Moreover, he clears the way for emotive constraints of self-narrative that keep self-narrative on track.⁶² Damasio makes clear

⁶² This is an important result. The role of emotion in human agency has long been neglected in the philosophy of action. Neuroscientific research establishes that emotions significantly affect action generation, action control and action interpretation. See Jing Zhu & Paul Thagard (2002).

that autobiographical narratives are intertwined with emotion and feeling. Persons are partly and importantly constituted by how they feel, by how it feels to be them. Their self-narratives are peppered with emotions. Schechtman for her part is aware of the significance of affect and the role empathic access plays in the making of self-narratives.

Antonio Damasio also is fully conscious – as is Marya Schechtman – of the embeddedness of persons. A full explanation of self-constitution needs to go not only beyond the brain to the rest of the body, but also into the world. Embodied persons are permanently interacting with their environments. How one feels about or thinks about one's self is partly constituted by the (social) world. Persons are situated beings. As they are socialised into human culture, they are taught to operate with a background conception of themselves as continuing individuals leading the lives of persons. Moreover, self-narratives are confined by observational and interpretive facts. In a deep way they are co-determined by the social world and constructed collaboratively. The individual person does not have the final say, so to speak.

I think a Schechtmanian narrative self-constitution view can do serious work in our understanding of the constitution of persons. All the same, Schechtman's view has to be slightly adjusted and enlarged. First, the articulation constraint has to be nuanced. There is a suggestion in it that persons have to talk themselves into existence. This was also one of Galen Strawson's worries. Schechtman replied: 'The articulation constraint demands that self-narration be more than the subpersonal, background operation of knowledge about one's past or projections of one's future, but constructing a self-narrative is also not conceived as something that must be undertaken as a conscious and active project' (2007: 163). So, the requirement is not that, in order to become a person, one must have a fully worked-out and explicit account

of why everything in his life is as it is, but rather that a person must recognise a certain kind of explanatory obligation, and be able to meet it for the most part. Another implication is that, although persons couch their self-descriptions in language, they are not exhaustively captured in what they can say about themselves. Self-conception and its operations are largely implicit and automatic. Self-narratives are bio-psychological constructs that are only partly expressible in verbal form. The narrative view I am defending in line with Schechtman's has in view a middle group of people who do understand how their person-lives hang together, most of the time not in so many words. But on being asked they can explain their behavior and locally make their self-narrative more explicit. Those are people who can deploy typically personal complex activities like autonomous or moral behavior.

And this brings me to a second task for a Schechtmanian narrative self-constitution view. What we can learn from it so far is clearly apropos, but at the same time quite trivial. It offers us the basic cream, sugar, and eggs for becoming a person. Unlike for example Korsgaard's view, the narrative self-constitution view says very little about deliberative action and reflective agency. Here a Schechtmanian narrative theory of personhood is in need of a supplement. We have to dig deeper into the subject. Persons do not want to be some arbitrary actors. They are able, if not compelled, to authorize their actions in their own name. Persons want to see their own signature in their actions. They want to be 'authentic', 'true' or 'real' persons. So, a narrative theory of persons in terms of agency has to make understandable how actions can be self-governed and authenticated actions of a particular person. It is to these problems of self-determining action and personal autonomy that I shall now (re)turn.

3: PERSONAL AUTONOMY

I can't run my life according to whether or not you like what I do. Not any more. You behave as if everything I do is part of the story of your life. You are the main character, I am a minor character who doesn't make an appearance until halfway through. Well, contrary to what you think, people are not divided into major and minor. I am not minor. I have a life of my own, just as important to me as yours is to you, and in my life I am the one who makes the decisions.

J.M. Coetzee (1999), *Disgrace*, New York: Viking Press, chapter 22

3.1 Introduction

One might have observed that my Schechtmanian picture of narrative self-constitution has drifted apart from the Korsgaardian reason view on self-constitution understanding persons as self-determinating agents. And one might say that the conditions of narrativity, as construed in the preceding chapter, will be trivially met by all individuals and do not inform us of any understanding of persons as self-governing agents. I agree. Schechtman's narrative self-constitution view addresses some commonplace features of people's lives and persons. And an understanding of persons as autonomous agents is beyond the scope of the modest narrative identity view I proposed in chapter two, taking my lead from Schechtman. So, let me make an effort at development here.

Schechtman's view thus far is this. In line with John Locke she insists that narrative identity must be defined in terms of 'sameness of consciousness', rather than sameness of

substance.⁶³ Schechtman develops the Lockean view by adding to Locke's recognition of the importance of memory and brute self-consciousness a recognition of the importance of being intelligible to ourselves. To have such a 'self-understanding view' one 'must see her life as unfolding according to an intelligible trajectory, where present states follow meaningfully from past ones, and the future is anticipated to bear certain predictable relations to the present (...W)e are constantly self-monitoring, keeping track of how we are feeling, what we are doing, and what we are like. This self-monitoring is mostly implicit (...). (U)sually we are caught up in the activity of living (...). On the self-understanding view, it is this self-monitoring that gives us our sense of continuation and coherence as a self, and so provides the kind of self-conception (...) that constitutes personal identity. (...) One quite simple implication of having such a self-conception is that a person can generally answer questions such as, "Why do I feel this way?" or "Why am I doing this?" should they arise. If she cannot answer them, she should be motivated to look for an answer' (2005a: 18-19). Importantly, what Schechtman has in mind here is quite modest. As noted earlier, she does not expect a person always to answer questions like 'Why am I doing this?' in terms of an articulated self-conception. The different parts of a life do not need to cohere in the sense of being wrapped up in a package but rather in the sense that someone

63 Locke tells us that person is a 'Forensick Term appropriating Actions and their Merit; and so belongs only to intelligent Agents capable of a Law, and Happiness and Misery' (J. Locke 1979: 346). This sets out two features that Locke takes to be unique to personhood. First, persons are capable of a special sort of self-interested concern. Second, they are moral agents, capable of taking actions for which they can be held responsible.

views oneself as having gotten from there to here somehow. And a person should not be totally blank when asked ‘Why did you do that?’, as if he did not understand the question.

Schechtman’s narrative self-constitution view is able to define practical identity in such a way that it entails a unified and persisting subject of experience. Practical identity, thus conceived, clearly is connected to agency. It is taken for granted that there is a link between agency and identity, between being a person and being an agent. The kind of psychological organisation that makes an individual a person on the narrative self-constitution view is the kind required for being an agent. Narrative identity primarily is about a series of connected actions and an acting human being.⁶⁴

This is, of course, a rather basic account of narrative self-constitution. And for many of us it is clearly insufficient. As self-reflective beings we want more. We do not just want to have a self-narrative and understand who we are. We want to direct our lives. We want our stories to go a certain way. And we want to be a certain way – to be a certain kind of person. This is precisely what psychiatric patients have in mind considering using advance directives. As I indicated earlier, the job of the last chapter of my dissertation will be to demonstrate that the narrative view I have been arguing for can resolve some problems raised by

⁶⁴ In *The Constitution of Selves* Schechtman connects narrative self-constitution to practical implications, especially to four features that play a crucial role in the day-to-day lives of persons and that are linked with agency: self-interest, moral responsibility, compensatory fairness, and survival. In *Experience, Agency, and Personal Identity* (2005b) Schechtman briefly discusses Korsgaard’s view on agency.

the use of both Ulysses contracts and advance directives in the case of Alzheimer's disease. Both types of contracts are written by persons who are competent to consent to medical treatment, stating how they wish to be treated in the event of becoming unwell. Such manipulative procedures of distal self-control have in common that they bear upon pre-commitment. They are cases of earlier selves exercising authority and imposing their will on later selves. This raises the question of (respect for) autonomy. So, before I can embark on the project of applying my view to psychiatric practice, I will have to say more on the problem of autonomy in relation to my theory of narrative identity. Therefore, let me pursue the difference between being a person on the one hand and being an autonomous person on the other.

Among the goings-on in a person's body, some but not others are due to the person in the sense that they are his doing. A person is not a helpless bystander of his doing: he can cause and enjoy control of it. He can be in the driver's seat of his behavior. When a person distinguishes between those goings on which are his doing and those which are not, he appears to do so in terms of their causes, by regarding the former but not the latter as caused by himself. Persons are able to claim: I want or think or do that; this is me and not just the work of an alien mechanism within me. There is a second, related element. Persons do not only want to control and govern their actions, they also want to shape themselves consciously and deliberately.⁶⁵ Being a

65 I don't mean this in an absolute sense. I agree with Jonathan Glover who captures the interplay of self-directed shaping as well as its limits by comparing the person to wood that can be sculpted, respecting the constraints of natural shape and grain. J. Glover (1988). *I: The Philosophy and Psychology of Personal Identity*. London: Penguin, 136.

person of one's own is being directed by considerations, desires, plans and the like that are part of what can be considered as one's self. So, persons not only act and live *self-governed*, they also act and live *self-governed*. They are, in a word, autonomous. And we expect a theory of persons to provide for an account of self-governance and self-shaping.⁶⁶

Now, what does it mean for a person to be autonomous from a narrative self-constitution view? Schechtman does not pay much attention to the problem of autonomy. The distinction between the person and the 'real' person is blurred. The search for autonomy and the search for coherence seem to be the same search for her. A person's experience or action seems to be autonomous if it fits smoothly and coherently into his autobiographical narrative. The case of 'self-blindness' is very instructive here. Remember Schechtman asked us to think of a man – let us call him Vladimir – who sincerely believes he loves his brother, but frequently acts towards him in ways that suggest hostility. Vladimir is unaware of this hostility. Other people are telling him that hostility plays a role. I think Schechtman's discussion of the Vladimir case is conflating two different questions: the question 'Is this protagonist a person?' and the question 'Is the protagonist in charge or governed by rigid and automatic influences?' Vladimir certainly is a person. One could describe him as a weakened person because his self-conception and the conceptions other people have of him considerably diverge. Vladimir denies what other people are holding up. His own story and the story of others lack coherence, in my own rendering. Yet, Schechtman's question 'Is Vladimir

66 The focus in this chapter is on personal autonomy and – as far as possible – not on moral autonomy. Personal autonomy is conceived here as a basic, neutral trait that individuals can exhibit.

really in charge or is he governed by “implicit” forces?’ is a different question. Now, the question is: ‘Is Vladimir an *autonomous* person?’ To be an intelligible, articulate, and coherent person is not, *ipso facto*, to be an autonomous person. Schechtman’s narrative self-constituting view seems to imply that the (more) autonomous person simply is the (more) coherent person, but that is, of course, not how we always experience it. Coherence cannot explain autonomy. Coherence sometimes even indicates a lack of autonomy. Imagine the following clichéd example.

A woman can try to overcome the fragmentation of her existence by the assumption of the rigidly laid-out social role of a housewife. This woman has a very coherent story to tell about herself that perfectly meshes with the views of her un-emancipated environment. Yet we will hesitate to call this woman an autonomous person because of so much coherence. Autonomy is not the first qualification that springs to mind when being confronted with people doing as much as possible what they did before and what others expect them to do. So coherence is not a sufficient condition to attribute autonomy. What more is needed? A stamp of uniqueness? Is there an autonomy problem for the housewife, because there may be dozens of persons like her, and because her story has been told and lived many times before? I do not think autonomy is a matter of the uniqueness of the life story. In my view, autonomy is not even a matter of the type of content of the biographical narrative. Here is my proposal: autonomy raises the issue of authorship. Autonomy is bound up with the strength of one’s *identification* with one’s autobiographical narrative. In other words, it is bound up with the active authorial work on the agent’s part.⁶⁷ The autonomy

⁶⁷ Autonomy is understood in terms of sovereignty (governing or managing oneself). It can also be understood in terms of authenticity. As the

of persons is related to the degree of appropriation of their life stories. A person can only be called autonomous, when he closely identifies with his actions and experiences. This ‘identifying act’ makes that the story becomes really his, that it becomes his particular life, and that it reveals who he really is. So, as far as the example of our housewife goes, the answer to the question of her autonomy still hangs in the balance. It depends on identification. Either she identifies wholeheartedly with the kitchen – and then we can call her an autonomous person, conventional as it may seem – or her identification is defective and then she loses autonomy.

In conclusion, although Schechtman does not raise the matter of autonomy explicitly, I think there is room in her narrative self-constitution theory for a notion of autonomy.⁶⁸ The conception

Greek word ‘*authentikos*’ suggests, authenticity points at originality in the sense of ‘concerning the first cause’, ‘from first hand’, ‘made with one’s own hand’, self-made. An authentic Van Gogh is a painting or drawing made by Vincent van Gogh. I take the two manifestations of autonomy – sovereignty and authenticity – as slightly different and complementary. Autonomy as sovereignty is controlling oneself, autonomy as authenticity is shaping and structuring one’s life over time. See Van Willigenburg & Delaere (2005).

68 In her 2007 reply to Galen Strawson, Schechtman separates her narrative self-constitution view into two distinct claims: a narrative account of persons and a narrative account of selves. In order to constitute oneself as a *person*, one must implicitly organise one’s experience according to a narrative. In order to constitute oneself as a *self*, one must have a narrative in which one experiences the past and the future as one’s own as part of a stronger kind of narrative: one does need to identify with it

I propose is that a person is autonomous when he endorses or encompasses (parts of) his narrative, i.e. when he identifies with (parts of) his narrative. This, of course, needs further development and support. There is, however, little agreement among philosophers as to how to unpack the problem of autonomy in any detail. The best attempt at an explanation of adequate conditions for self-determining action and personal autonomy is contained in the work of Harry Frankfurt.⁶⁹ I take his hierarchical view as a start of my search for narrative based personal autonomy.

[*continuation of note 68*] or care about it or take an interest in it (2007: 167-71). Schechtman's distinction resembles my contrast between a narrative account of persons and a narrative account of autonomous persons. But Schechtman does not connect it to autonomy.

69 Frankfurt advanced his central intuition in his 1971 'Freedom of the Will and the Concept of a Person', reprinted in Frankfurt 1988: 11-25. Gerald Dworkin was the second founding father of the hierarchical view of personal autonomy (in his 1970 'Acting Freely', *Nous* 4, 367-83). For a recent overview of the debate, see J. Anderson & J. Christman: 1-12 and J.S. Taylor: 1-29.

3.2 Harry Frankfurt on caring and loving

According to Frankfurt, reflexive⁷⁰ consciousness is the hallmark of human beings. By exercise of this faculty creatures like ourselves are capable of introducing a division within ourselves, which in a way separates us from our own conscious states and activities, and then of coming to acquire an attitude regarding those activities and states. Self-awareness makes us susceptible to an inner division in which we separate from and objectify ourselves. This puts us in a position to assess the motivating forces by which we happen to be impelled, and to determine which of them to accept and which to resist. We want our actions to be motivated in certain ways, and not in others. We want our motives and choices, however we acquire them, to be the motives and choices that we really want and therefore in no way alien to us. Frankfurt's idea is that agents will identify with an action, A, and will be able to own what they do, to the extent that they have a second-order volition to do A. They will have such a volition to do A, he says, so far as they want to be controlled by the desire to A; they want it to be the desire that moves them effectively to act (1988: 15). So, Frankfurt deploys two distinct ideas here. The first is that agents are all capable of having, not just first-order desires to do this or that, but also second-order desires of one kind or another. The second idea is that among an agent's

70 'Reflexive' means 'self-conscious'. 'Reflective', conversely, comes to mean 'thoughtful'. Human beings are able to take a stance on their own attitudes. And in a way they have to, as it is nicely put into words in Rilke's *Eightest Elegy*, that fathoms the human condition: "Dieses heisst Schicksal: gegenüber sein und nichts als das und immer gegenüber." R. M. Rilke (1923), *Duineser Elegien*.

second-order desires may be a desire that he or she be effectively moved by a certain first-order desire: that that first-order desire, as Frankfurt says, be his or her will. This distinguishes those second-order desires that count as second-order volitions from more idle second-order desires: from desires just to have and experience certain first-order desires, but not necessarily to be moved by them to action.

In his early writings Frankfurt thinks of volitional control on lines that parallel the notion of rational control. Rational control requires that the agent is disposed to form beliefs and desires in a rational pattern and is disposed to act as they rationally require. 'It is only in virtue of his rational capacities that a person is capable of becoming critically aware of his own will and of forming volitions of the second order. The structure of a person's will presupposes, accordingly, that he is a rational being' (1988: 17). So, the early Frankfurt construes his theory as a theory of rational-cum-volitional control. And it is rational-cum-volitional control that constitutes an agent's personhood. For the later Frankfurt, on the other hand, what the person really wants is what he 'cares' about. Things he cares about, or things he 'loves', are things with which the person identifies himself, and which he accepts as expressing what he really wants. It is this act of selective appropriation or identification that occupies me here. At issue is the question of what it is to identify with one's actions, – where identification is understood as a condition of autonomous agency and personhood, and not as a condition of freedom or moral responsibility. As we just saw, Harry Frankfurt's work has different incarnations. I want to take Frankfurt's current position on identification as a point of departure. Let me first summarize this view.

What are the things that should be important to me and that

I should care about? According to Frankfurt, all efforts to conduct a rational inquiry must fail here. The question of what one should care about must already be answered before a rationally conducted inquiry and deliberation aimed at answering it can even get under way. That question can sensibly be asked only on the basis of a prior answer to the factual question of what one actually does care about or love. Some cares are innate in us, e.g. care for the continuation of our lives or the care for the well-being of our children (2004a: 27-30). To be sure, it is not only living things that we are capable of loving. The object of love can be almost anything – a way of life, a person, a group, a mother tongue, a quality of experience, ideals, a tradition, whatever. Now, what does Frankfurt exactly mean by love? This is his answer: ‘Love is a particular mode of caring. It is an involuntary, non-utilitarian, rigidly focused, and – as is any mode of caring – self-affirming concern for the existence and the good of what is loved.’ (2006: 40).⁷¹ Love is *rigidly focused* in that there can be no equivalent substitute for its object, which is loved in its sheer particularity. Love is *non-utilitarian* in that it cares about the beloved for its own sake. Love is *involuntary*, in that it is not under the immediate control of the will. We cannot love – or stop loving – at will, merely by deciding to do so. Neither is loving the rationally determined outcome of even an implicit deliberative or evaluative process. It is not essentially a matter of judgment or of reasoned choice. Love requires no reasons, and it can have anything as its cause.⁷² However, love is a powerful source of reasons: ‘Loving someone or something essentially *means* or

⁷¹ See also in 2004a: 41-47 & 79-80. To put it differently: love is a central case of volitionally necessary caring.

⁷² On the inconsiderableness of causal history see also 1988: 54.

consists in, among other things, taking its interests as reasons for acting to serve those interests' (2004a: 37). In spite of this, it is not essential to love that it is accompanied by particular affects or thoughts. The heart of the matter is neither affective nor cognitive. It is volitional. It has to do with contingent 'volitional necessities'. It is in the nature of the lover's *self-affirming concern* that he is deeply invested in his beloved. The beloved configures and limits the lover's will. Another way of putting this is that the lover identifies himself with what he loves. Love defines what the lover most intimately is. It sets the boundaries of his practical life and thus fixes his shape as an active being. It profoundly shapes his personal identity and the ways in which he experiences his life.⁷³ Let me proceed to explain the Frankfurtian idea of identification with a problem case.

It may happen that a person truly loves something but that, at the same time, it is also true that he does not want to love it. There is a part of him that is opposed to his loving it. In a word, the person is ambivalent. To overcome ambivalence the person has to become finally and unequivocally clear as to which side of the conflict *he* is on. As soon as he has definitively established just where he stands himself, his will is no longer divided and his ambivalence is therefore gone. He has placed himself wholeheartedly behind one of his conflicting impulses, and not at all behind the other. The tendency that the person opposes is in a sense extruded and rendered external to him. It is dissociated from his will and thereby becomes alien to it. Sometimes, however, the person is unable to make up his mind once and for all which side to be on. He cannot bring himself to identify

⁷³ Frankfurt clearly is breaking a lance here, contra Kant, for 'self-love' (2004a: 61-62).

decisively either with one or with the other. In such cases, the person is volitionally fragmented. His will is unstable and incoherent, moving him in contrary directions simultaneously or in a disorderly sequence. He suffers from a radically entrenched ambivalence, in which his will remains obstinately undefined and therefore lacks guiding authority. This person is at odds with himself. He is deficient – at least with respect to his volitional faculty – insofar as he is not wholehearted. *Wholeheartedness* turns out to be a key notion here for Frankfurt. Being wholehearted means: having a will that is undivided. The wholehearted person is fully settled as to what he wants, and what he cares about. With regard to any conflict of dispositions or inclinations within himself, he has no doubts or reservations as to where he stands. He lends himself to caring and loving unequivocally and without reserve.⁷⁴ Thus wholehearted identification means that there is no ambivalence or any other sort of conflict, instability or confusion in his attitude toward himself, i.e., that there is volitional unanimity. What counts above all is the quality of the will – that is, its integrity – not, for example, the quantity of its objects. One has to be fully intact, one's will has to be purely his own. What, then, one could ask, is so special, urgent, and wonderful about integrity?

One thing in favor of an undivided will, Frankfurt says, is that

⁷⁴ Frankfurt makes a difference between necessities of love and pathological constraints upon the will, such as compulsions, obsessions, addictions, and the like. They differ in that we only submit unwillingly to the latter. Their power over us is external, and merely coercive. The power of love, in contrast, is not like that. Frankfurt admits that it is very difficult to grasp people accurately in full depth and detail, so far as love is concerned (2006: 43-5).

divided wills are inherently self-defeating. Division of the will is a counterpart in the realm of conduct to self-contradiction in the realm of thought. It leads to defective action. By the same token, enjoying the inner harmony of an undivided will is tantamount to possessing a fundamental kind of self-assurance. An undivided will is also important for us, 'because it is the same thing, more or less, as being satisfied with ourselves. (...) It is a condition in which we willingly accept and endorse our own volitional identity. We are content with the final goals and with the loving by which our will is most penetratingly defined' (2004A: 97). So, what is central to identification is the agent's satisfaction with the relevant attitudes. Satisfaction with an attitude is a structural feature of the psychic economy within which that attitude is embedded. One is satisfied with such an attitude when there is no relevant 'tendency or inclination to alter' it (1999: 104). Frankfurt stresses that wholeheartedness (i. e. satisfaction) is only a structural characteristic, which has to do with volitional unity or integrity. It has no essential evaluational vector. The value of what a person loves is irrelevant to the question of whether he is wholehearted in loving it. Love is not necessarily a response grounded in awareness of the inherent value of its object. It may sometimes arise like that, but it need not do so. It is entirely possible for a person to be caused to love something without noticing its value. This leaves open the possibility that someone may wholeheartedly love what is evaluationally non-descript, or what is evil. Being robustly wholehearted is quite compatible with being dreadfully and irredeemably wicked. The function of love is not to make people good. Its function is to make their lives 'good for them to live'.

Let us take stock. What we can learn from Frankfurt on identification is that the defining characteristic of persons is that they are creatures who care about and love things. Echoing Hume,

Frankfurt insists that whatever we have reason to do or refrain from doing depends on what, in particular, we care about. The latest Frankfurtian model is consistent with his earlier thoughts on identification in at least two respects: the model is hierarchic – identification supposes reflexive consciousness – and it is conative. Furthermore, caring is linked up with self-constitution. To care (wholeheartedly) is the key to self-integration. The things – the volitional necessities – we cannot help caring about and we put ourselves behind are constitutive of who we are. There is a close connection with self-governance too. Volitional necessities also have agential authority. On the question of what is it that gives – to put it in Michael Bratman's terms – 'some particular wiggle in the psychic stew agential authority, and not others?' Frankfurt's answer is: that the agent is caring wholeheartedly about it and is satisfied with it, that is, that it is embedded in a psychic system in which there is no relevant tendency to change.⁷⁵ Self-government or making the government up to the self, in Frankfurt's psychological account, is one and the same thing as being identified wholeheartedly. The notion of wholehearted identification provides the hierarchical theorist with the resources to account for autonomous agency. An autonomous agent identifies himself fully and uninhibitedly with the configurations of the will that define his final ends. He accepts them as commandingly authoritative. There is no part of him with which he identifies that is opposed to or that resists his loving what he loves. The lover may be unable to resist the power his love exerts, but it is by his own will that he does what the necessities of love require.

⁷⁵ Frankfurt refers to Baruch Spinoza's 'acquiescentia in se ipso' (2006: 17).

Needless to say that Frankfurt's view is not undisputable. Most of the comments⁷⁶ on his interpretation of identification centre around the question of how it can make clear that identification is conceived as authoritative, that is, as normative and binding for what a person really wants. Frankfurtian identification seems to entail a moment of endorsement or reinforcement of a deliberative conclusion, of some conclusion about what is the best thing to do, what is most choice-worthy, or what is important. Frankfurt does not leave open the question of whether this is a matter of decision or, rather, a matter of evaluative judgment. Frankfurtian identification with volitional necessities must be understood in a *non-cognitivist* and *non-voluntaristic* way. Let me expand a bit on these two characteristics.

In recent replies to his critics Frankfurt clarifies the first one: 'In speaking of these matters, I have regrettably made use of terms – such as "endorse" – that naturally suggest a positive evaluation. However, what I have actually intended to convey by referring to "endorsement" is not that the agent *approves of* what he is said to endorse, or that he considers it to merit his support, but nothing more than that the agent *accepts* it as his own. The sense in which he accepts it as his own is quite rudimentary. It is free of any suggestion concerning his basis for accepting it and, in particular, it does not imply that he thinks well of it' (Frankfurt in Buss & Overton: 87). Frankfurt cannot see that identifying with a desire always requires that the desire provides a reason for such a choice: 'Our most elementary desires come to us as urges or impulses; we are moved by them, but they do not as such affect our thinking at all. They are merely psychic raw material.

⁷⁶ E.g. the comments of G. Watson, R. Jay Wallace, M. E. Bratman and T. M. Scanlon in Buss & Overton 2002.

A desire provides us not with a reason but with a problem – the problem of how to respond to it. Impulses and urges have power, but in themselves they have no authority. They move us more or less strongly, but they make no claims on us. One problem of how to respond to it. Impulses and urges have power, but in themselves they have no authority. They move of the thematic preoccupations of my work in this area has been a concern with trying to understand just how to account for the authority – as distinct from the mere impulsive force – that certain desires appear to have. (...) A person's elementary urges and impulses do not become authoritative for that person because he has favorable second-order attitudes toward them, or even because he has second-order volitions manifesting his desires that those desires and impulses move him effectively to act. What gives a desire authority is, rather, that the person identifies with it or commits himself to it' (Frankfurt in Buss & Overton: 184). In other words, Frankfurt's definition of identification points at brute psychological facts, psychic raw materials that a person may or may not accept for integration into himself. His assumption is that the desires important to identification are (non-judgmental) matters of inclination, attraction, or repulsion. To be sure, people are not love machines. In a way, it is up to us that this or that appeals to us; we make up our minds that we want it. A human being must determine, with respect to some desires, whether to identify himself with it or whether to reject it as an outlaw and hence as not a legitimate candidate for satisfaction. But the vocabulary of reasons comes on the scene only later. For someone to identify with a desire means merely that – for whatever reason, or for no reason whatever – he joins himself to the desire and accepts it as his own. A person may be led to accept something about himself in weary resignation, as well as in welcoming approval, or in recognition of its merit. The fact that he

accepts it, in other words, entails nothing about what he thinks of it. This leads to the second characteristic mentioned above: Frankfurt's non-voluntarism.

Frankfurt clearly offers us a volitional account of practical normativity. Normative authority cannot be established by invoking rationality. There can be no rationally warranted criteria for establishing anything as inherently important. It is possible to ground judgments of importance, Frankfurt believes, only in judgments concerning what people *de facto* care about. The normative question of what people should care about can be answered only on the basis of a prior answer to a question that is not normative, but straightforwardly factual: namely, the question of what people actually do care about. Does this mean that it is all simply up to us – that what is important to us depends just upon what goes on in our minds? Yes and (especially) no. Yes, because answers to normative questions are up to us in the sense that they depend upon what we care about. Frankfurt's idea of second-order attitudes presupposes that it must be possible to influence at least some of our carings. However, what we care about is not *always* up to us. We cannot have whatever will we want, simply for the asking. There are some things we cannot help caring about. Our volitional character does not change just because we wish it to change. It may change. But love is not a voluntary matter. We cannot bring ourselves to love at will – that is, merely by choosing or deciding to do so. When we decide⁷⁷ or make up our minds, we are mainly reacting and neutrally accepting the persistence of our carings. And when we change our cares, we are moved by cares once more beyond our direct, immediate control. To decide, so construed, is more like to discover.

⁷⁷ To 'decide' – Frankfurt sometimes uses the word – is a debatable term in this context.

3.3 Some teachings and tensions

Frankfurt is critical of the philosopher's tendency to exaggerate the authority and control of reason over the person. In fact, his criticism makes him nearly silent on the relevance of reasoning. To illustrate this point he discusses Bernard Williams's example of the man who sees two people on the verge of drowning – only one of whom can be saved – and who must decide which of the two he will try to save. One of them he does not know; the other is his wife. We do not expect this man to consider questions of impartiality or fairness and to add the thought that in situations like this it is permissible to save one's wife. As Bernard Williams famously admonished: that would be 'one thought too many'. Frankfurt is sympathetic to Williams' line of thought but proposes two corrections: 'For one thing, I cannot help wondering why the man should have even the one thought that it's his wife. (...) In the circumstances that the example describes, any thought whatever is one thought too many. In addition, the example is seriously out of focus in a rather more fundamental respect as well. It actually can't work in the way that Williams intends if we stipulate nothing more than that one of the people drowning is the man's wife. After all, suppose that for quite good reasons he detests his wife. (...) Specifying merely a bare legal relationship between the man and the drowning woman appears really to miss the point. So let us put aside the matter of their civil status, and stipulate instead that the man in the example loves one (and not the other) of the two people who are drowning. (...) The fact that he loves her, just in itself, entails that he already takes her distress as a more powerful reason for going to her aid than for going to the aid of someone about who he knows nothing. (...) Loving someone or something essentially means, among other things, taking its needs and interests as reasons for

acting to serve those interests and needs. Love is itself, in other words, a source of reasons. It creates the reasons by which acts of loving devotion are inspired. As a matter of fact, that's precisely how it is that love makes the world go around' (2001: 1-3).

The point is well made.⁷⁸ From the outset it has been clear for Frankfurt that *reflexive consciousness*, *reasoning*, and the *ability to love* are essential human faculties. But love strikes him as more fundamental and important than reason. Frankfurt's idea is that making up our minds (e.g. to form intentions, to commit ourselves to certain ideas or desires, to identify ourselves with an inclination) is primarily volitional activity. Identification is not basically an ability to perceive formal relationships, or to fashion or grasp concepts. Nor does it aim at understanding, or believing, or recognising, or at any prerogative of reason. Reason is not even necessary. In Frankfurt's judgment, identification is a volitional state that creates reasons, but that does not otherwise depend upon them. We can identify with various psychic elements, without having any reasons for doing so. Identification is a non-rational, partly unconscious activity, – and Frankfurt regrets not having acknowledged this more explicitly (in Buss & Overton: 297). Likewise, to act autonomously is being moved by the character of your own will, that is, being moved by love. I think this point has a good deal of force. Unchosen and non-reasoned contingencies play a role in determining whom and what we love, and are giving substance to our lives. We more or less 'fall' into our loves, in processes that elude complete deliberative planning. We are rather hostage to the contingencies of the empirical which we do not control, even in our most resolute declarations of will. And when we try

⁷⁸ For a similar critique on Williams' concept of 'ground project' see Frankfurt 2006: 36-7.

to guide ourselves, there always remains the empirical question of whether we have indeed succeeded in making our will what we want it to be. On the other hand, these contingencies also profoundly affect deliberative reflection, providing us with reasons for actions that we would not otherwise have, and appropriately altering thereby our conceptions of who we are and what it would be best to do. I think all this is on the right, naturalistic lines. Volitional necessities are an essential condition for autonomous life-guidance and decision-making. We can be only what nature makes us. There is a natural mechanism or some sort of a pre-existing nature in place, that guides our attempts to identify and helps to find a resting place in such attempts. A good deal of our identity is already in place by the forces of heredity and environment, before we reach the stage of reflective endorsement. These are forces which shape us in a strictly causal way, and although they have no claim as reasons, they are among the background conditions that allow us to form identifications and ultimately give us reasons. Frankfurt has a point here.⁷⁹

⁷⁹ Many feminist philosophers stress this point. Corporality and emotion are co-determining self-understanding and personal autonomy. One of the merits of Frankfurt's account of wholehearted identification is that it can be extended to identification with unconscious, visceral and affective items. But the fact that these conditions are not of our own choosing and that it is 'unthinkable' for us to give them up does not imply that there is nothing to know about it. Representative here is Diana Tietjens Meyers. In 'Decentralising Autonomy: Five Faces of Selfhood' she argues for the need to redress the overemphasis on rational oversight functions and self-definition to the neglect of the role of self-discovery in autonomy. Whereas self-definition is a matter of self-analysis and reflective endorsement, self-discovery is more a matter of sensitivity and openness

However, his approach seems to take things too far and comes at a price. Here is a final observation of Frankfurt's apparent non-cognitivism, before I turn to a critical discussion of this point.

What Frankfurt thoroughly rejects is not rationality but the doctrine of 'normative realism'. This doctrine holds that volitional necessities are responses to an independent normative reality. On this account, certain things are inherently important; and there are objective reasons for people to act in accordance with the normative reality, whether they know them, or care about them, or not.⁸⁰ Frankfurt's view is different. In his judgment, normativity is not a feature of a reality that is independent of us. It is grounded in us. This is not to say that what we are to care about is a matter that can be settled arbitrarily, or that appears to rule out the possibility of providing any rational basis for deciding what we are to care about. Quite the reverse, the final ends by which we govern ourselves require authentication by some decisive rational warrant. This is, for example, how Frankfurt comments on a much-discussed passage in David Hume's *Treatise of Human Nature* that "Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my

[*continuation of note 79*] (in Christman & Anderson: 27-55). For another exploration of the role of bodily emotion, see Robert Solomon (2004) on 'judgements of the body' in *Thinking about Feeling*, Oxford: Oxford University Press, 84-8.

80 In this case, autonomy is determined by the agent's correct grasp of the truth and the good. Self-government requires not only that one acts in accordance with the reasons one believes to have, but in accordance with the right reasons there objectively are. Defenders of such a view are S. Wolf (*Freedom within Reason*, 1990, Oxford: Oxford University Press) and B. Berofsky (*Liberation from Self. A Theory of Personal Autonomy*, 1995, Cambridge: Cambridge University Press).

finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the lest uneasiness of an Indian or person wholly unknown to me.'⁸¹ Frankfurt: '(I)t is true that this preference involves no purely logical mistake. (...) In this purely formal sense of rationality, his choice is not at all irrational. But what would we say of someone who made that choice? We would say he must be crazy. In other words, despite the unassailability of his preference on logical grounds, we would consider both it and him to be wildly irrational. (...) His irrationality is not fundamentally a cognitive deficiency at all. He is volitionally irrational. He has a defect of the will' (2006: 29-30). Being volitionally rational is not just a matter of choices that a person actually makes. It involves being incapable of making certain choices and of crossing certain external borders. It requires that certain choices be utterly out of the question. A volitionally rational person may think that destroying the world is appropriate; but, when the chips are down, he cannot bring himself to implement his judgment. In virtue of the necessities by which his will is constrained, his choice is not among his genuine options. It is precluded by volitional constraints that cannot be eluded. We cannot help caring about e.g. going on living, 'about avoiding crippling injury and illness, about maintaining at least some minimal contact with other human beings, and about being free from chronic suffering and endlessly stupefying boredom. We love being intact and healthy, being satisfied, and being in touch. We cannot bring ourselves to be wholly indifferent to these things (...T)hey are elementary constituents of volitional reason itself.' (2006: 38). Are these constraints 'objective'? In one sense

81 David Hume, *A Treatise of Human Nature* (ed. L.A. Selby-Bigge), Oxford: Clarendon Pres, 1978, 416.

they are obviously not. They are grounded nowhere but in the character of our own will. That means that they are subjective. On the other hand, they are outside our direct voluntary control. This warrants them as objective, despite their origin within us. The objectivity is relative in part to the common nature of human beings and in part to individual experience and character. Still, despite this relativity, there are plenty of ways that our loving can go absolutely wrong, Frankfurt acknowledges. We have to be alert to the possibility that we do not understand the things that we love well enough. We have to be wakeful to the possibility that other people are required by what they care about to harm or to destroy what we love, and that the conflict between us, so far as reason goes, may be irreducible. This is just a fact of life. It is here, where the power of reason is under discussion, that we arrive at a difference between my view on self-governance and Frankfurt's.

Frankfurt's conception of identification is unsatisfactory, because it distorts the complexities of human agency and identification. My claim is not that Frankfurt's theory is wrong, but that it is incomplete. Reason carries too little weight in it. To improve on it, we need to restore the balance and to move further away from the non-cognitivism that is active in Frankfurt's work. The fact that identification starts with the will does not mean that it has to end with it. From the fact that reason underdetermines a person's answer to the question of who he is, it does not follow that judgments have no roles to play in it. It does not follow that our autonomy as agents no longer can be identified with the point of view of evaluative reflection about what we have reason to do. Frankfurt seems to advocate that we care about what we can, and that it does not matter which cares we commit to, as long as we commit to some. But autonomous persons are not only fully contented human beings, creatures

just revealing their natures and ‘being in flow’. They are also givers and takers of reasons, mongers or traffickers in reasons, as Robert Brandom calls them. To be sure, as we saw earlier, Frankfurt does not reject, but only de-emphasises the role of reasoning. According to Frankfurt, responsiveness to reasons is what identification and wholeheartedness inescapably impose. These ‘provide us with a problem’ that has to be solved, in Frankfurt’s own striking formulation. Although he goes on to stress that our evaluations of what we care about are more often such a response to the caring than its source, our responses are related to reason. Volitional means: programmed to act in some way. There is rationality in it, as Frankfurt admits. (Instrumental) reason matters for Frankfurt. And this leads to the crucial question: At what point is thinking desirable for autonomous agents? This is a question that is mistakenly neglected by Frankfurt. He downplays the requirement of actual reflection.

The critical issue was pressed with insight by Joel Anderson in his 2003 *Autonomy and the Authority of Personal Commitments*. When we accept that the authority of our desires and commitments ultimately stems from the reflexive involvement of the will rather than of the intellect, he asks, how then can we distinguish a change of heart from self-betrayal? Acting at odds with a core personal commitment in Frankfurt’s view seems always to end up being not a violation but simply evidence that one no longer cares. To solve this problem Anderson proposes, as a first step, an inferentialist understanding of wholeheartedness: ‘What makes a state a genuine desire – in the language I am using, what gives it authority – is the degree to which the desiring agent lives up to *inferential commitments* entailed by having such a desire. (...T) he inferentialist point is that my having an authoritative desire is a function of being in a network of affective, behavioral, and cognitive commitments’ (2003: 97). This shift to an inferential-

ist network can make more plausible the intrapersonal approach to explaining how a current desire or commitment can get grip on future behavior, but it cannot solve the problem regarding distinguishing self-betrayal from a change of heart or a previously neglected pattern. Anderson's next step then is 'to take the inferentialist point – about desires having authority (and thus being able to function as reasons) in virtue of their connections to other commitments, desires, dispositions, &c. – and combine this inferentialism with a recognition of the way in which what counts as an *intelligible* connection is constrained by the normative standards built into our shared social practices' (2003: 102). The question of whether neglecting a particular project, relationship, or ideal that someone once cared about is to count as a welcome development or a change for the worse, has to be answered in part on the basis of whether the account the person could give (for why it is not self-betrayal) meets publicly shared standards of an acceptable answer.

Many critics of Frankfurt's conception of autonomy have raised the point of the intrapersonal and hyper individualistic approach.⁸² Communitarians have famously claimed that the emphasis on separated, atomistic decision has obscured the socially embedded nature of personal identity. Contextual accounts of autonomy have been developed to respond to such critique.⁸³ It is obvious – also for Harry Frankfurt, I am sure –

82 Recently by Meir Dan-Cohen in 'Socializing Harry' in H. Frankfurt (2006), *Taking Ourselves Seriously & Getting It Right* (ed. D. Satz), Stanford: Stanford University Press, 91-103.

83 See, for example, C. Mackenzie and N. Stoljar (eds.) (2000), *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, New York: Oxford University Press.

that individuals cannot constitute themselves as persons *ex nihilo* and that personhood is parasitic on lasting emotional attachments to others and on socialization and enculturation. In the present section, however, I largely put to one side Frankfurt's a-social approach. More important, for my purposes, is that Anderson and other critics make clear that persons in order to be autonomous persons need explanations to tackle self-betrayal⁸⁴, and that these explanations will be explanations in terms of reasons rather than a fit with volitional patterns. This takes us beyond the idea that a certain degree of integration and equilibrium within the volitional make-up can do the self-constitution job alone. The required stability is generated, at least in part, by non-volitional considerations.

With this understood, it will be clear that Frankfurt's proposal does not accomplish what we need. On the latest Frankfurtian model, autonomy ultimately stems from the reflexive involvement of the will rather than of reason. The autonomy of one's actions comes from how they are connected to one's volitional make-up. What makes actions autonomous is not that they have

84 John Christman makes a similar point criticising the static approach of the hierarchic model of autonomy (the 'synchronism-objection'). He develops an explicitly historically based version of the hierarchical approach to analysing autonomy: Persons are autonomous with respect to a desire, emotion, character trait and the like, if they competently reflect on it and endorse it *in light of the processes by which it developed*. See e.g. J. Christman (2001). 'Liberalism, Autonomy, and Selftransformation' in *Social Theory and Practice* 27 (2), 185-206. One of Christman's concerns is the problem of manipulation that threatens split-level accounts of autonomy. The manipulation problem, like self-betrayal, cannot be solved without reasoning. We need a (his)story.

been endorsed in critical reflection, but rather that they are subject to imperatives of the will, that is, to certain volitional necessities. Such a non-cognitivist interpretation of self-governance, in my view, is incomplete and ultimately unacceptable. Unless we can articulate what it is about the course of action that seems desirable, our leanings will appear essentially inscrutable, more like an itch or a sensation than a real desire. Just as much as volitions, the individual's evaluative judgments are definitive of personal autonomy. Volitions are not the only kinds of entities through which to bring practical deliberation to a satisfactory solution. If we understand carings as practical identities⁸⁵, that is, the kinds of commitments and concerns that most fundamentally color our personality as agents and structure our ordinary deliberations, we cannot but *reason* about these patterns of concern (predilections, final ends, ground projects). They are the subject of our critical reflection and a matter of normative articulation. In short, they play a role in practical reasoning. Reason as much as love is crucial for personhood and personal autonomy. Reason is not an epiphenomenon. We inevitably want to argue about the importance of what we care about. Moreover, as we have seen, purely volitional accounts of autonomy run

85 I can think here of Bernard Williams' ground projects (e.g. personal relationships and significant work) or of Korsgaardian practical identities. Like M. Betzler I think this can be done in fairness to Korsgaard. I realise that Korsgaard's conception of practical identity is slightly different. In Korsgaard's view practical identities are roles or principles generating constraints. The tenor of her argument is that it is our humanity, to wit our identity as reasonable beings, that makes us think that some ends are important to us. To value a practical identity goes back to valuing our humanity, in Korsgaard's account.

into serious difficulties. Frankfurt is realistic in stating that rational judgments better cannot be prioritised above thoughtless attitudes, as widespread practices of wishful thinking, rationalisation, delusion and akrasia demonstrate. But there is no more ground to prioritise volitions in our definition of autonomy because of our satisfaction with the establishment of a certain internal harmony and constancy of the volitional necessities ('acceptance'). Volitions can also dazzle us. A volitional account cannot make clear, for example, how to demarcate volitional necessities from brainwashing. It cannot explain, let us say, the difference between (autonomous) identification as a child loving her father and (heteronomous) identification as a member of the sect of Sister Rosetta of the Heavenly Gate, brainwashed by her leader Do. This seems to be a serious deficit for an account of personal autonomy. So, we have to extend Frankfurt's approach. My proposal is to develop Frankfurt's view without losing sight of his key insights that the authority personal commitments have cannot be reduced to *contentions* about the importance of what one cares about, and that our embarking on the path of critical self-evaluation is itself an action grounded ultimately in cares. I also want to avoid a pan-rationalistic revision of Frankfurt's position.⁸⁶ My more modest claim is that a volitional approach alone cannot account for self-governance. Identification with volitional necessities is usually contaminated with evaluative

86 M. Betzler, G. Watson, R. Jay Wallace, T. van Willigenburg, and J. Raz are making moves in that direction. Here is an illustrative quote: 'We are ourselves and we lead our own life so long as we see ourselves as rational agents, so long as we conduct our life under the semblance of rationality', from J. Raz (1999). *Engaging Reason. On the Theory of Value and Action*, Oxford: Oxford University Press, 19.

judgment. In our daily lives acting autonomously will involve self-conceptions playing a decisive role, as the sorts of organising principles that serve to privilege some reasons rather than others. How one thinks of oneself becomes a principle for giving order to the pattern of one's behavior, intentions, desires, emotions, and the like.

Let us return for a moment to the sect member. How does my proposal (a partial rehabilitation of reasonableness) work out in the case of a brainwashed person? After all, sectarians may be quite reasonable persons. My claim here is not that in order to be autonomous one must be able to reflect on his volitional necessity and can only endorse it if it has a causal history of the right sort. My more moderate claim is that reasons have to play a role in the exercise of autonomy. It is not trivial for members of the sect of Sister Rosetta of the Heavenly Gate to be able to realise what causal path led them to their identification with leader Do. It may give them reasons to try to get rid of the sect's influence. But there is a second route to avoid heteronomy for our sect member: he may be able to realise fully what causal path led him to his identification with leader Do and endorse it without giving him reasons to change his behavior.

Section 3.3 concludes. The situation so far is as follows. The hallmark of personal autonomy is agentic reflexivity. What matters for autonomy is the ability to draw away mentally from our own mental states (i.e. what makes up our stream of consciousness: perceptions, desires, beliefs, emotions, recollections, daydreams, intentions...). We are able to 'supervise' our mental states, forming attitudes that have as their intentional objects these mental states. As Keith Lehrer emphasises, we have meta-minds, minds capable of turning upon themselves. Desires, emotions, and beliefs present us with a problem of what to do with it: identify

with it, or reject it ‘as an outlaw’? This ability to step back and to reject or identify is crucial to control our behavior and the direction of our lives. Harry Frankfurt highlights identification with volitional necessities as a basic condition for self-mastery as well as for forming the boundaries of the self. Frankfurt points at identification with volitional necessities as a psychological fact, that is, as a depth investment in ourselves, rooted in our bodies (that are essentially related to the world). And it seems very important, indeed, to get ‘in touch with our volitional necessities’. It is about us, not about heteronomous forces. But we cannot stop here. Caring and loving are not the only aspects that are relevant and conditional for identification. Reasoning, interpreting, and evaluating are also part of it. Identification turns out to be an *aspectual* psychological phenomenon. Frankfurt basically agrees but does not take great pains over the rational part of volition and the role of evaluative judgment. This flies in the face of the phenomena. People reason a lot about their patterns of concern, i.e., the kinds of commitments and activities that most fundamentally color their personality as agents. The phenomenon is quite familiar in hearing the shorter stories that people tell about themselves in the pub or over the dinner table. To identify is also, as Michael Bratman puts it (a bit cryptically), making reflective decisions to treat our desires as reason-giving (Bratman 1999: 185-206). So, we need a more complete story of identification.

3.4 Identification and coherence

Let me insert first a quick scan of the actual debate on self-governance and then specify where exactly my own view comes in. *In globo* the running debate on self-governance is a debate between

‘hierarchical theories’ of self-government and ‘reasoning responsiveness theories’ of self-government. Harry Frankfurt and Gerald Dworkin are defenders of the hierarchical or so-called ‘split level theory’ of autonomy. The reasoning responsiveness view is represented by, among others, G. Watson, A. Mele, and J. Raz. Their claim is that self-governing or autonomous action is: to act in accordance with what one takes to have most reason to do. Autonomy, in their view, is exercise of reason. From Frankfurt’s point of view, reasons are instrumental for being autonomous and autonomy supposes the formation of a desire. For reasoning responsiveness theorists autonomy supposes reflection, and reasons are essential for being autonomous.⁸⁷

The narrative view on autonomy I am proposing plainly tries to claim middle ground between the two main theories of autonomy. It goes second-order and it gives reasoning responsiveness its due. Autonomy now stands for: identification with (parts of) a self-narrative. In the previous chapter I listed some important standards for self-narratives that do constitute personhood: self-constituting narratives must be articulate and coherent. Narrative conceptions of personhood arrange actual, past and planned

⁸⁷ One of the recurrent problems for reasoning responsiveness theories is: how does reason motivate us? Possible answers are: people have a disposition to act in accordance with their better or best judgment; they have a generic desire to act as they judge best (A. E. Mele in 1995 *Autonomous Agents. From Self-Control to Autonomy*, Oxford: Oxford University Press, 25-30). In a similar vein, J. Raz holds that almost all people have a desire to be rational (1999: 17). In his radical view, it is not even a contingent fact that people strive for the good. The mental act of knowing the good and desiring the good are identical: *Quidquid appetitur appetitur sub specie boni* (quoted by Raz).

actions in ways that cohere into an acceptable story, i.e., a story that can be made explicit, that is internally coherent, and that is in accordance with the facts of the world and the interpretations of others. Coherence is the constraining power of self-narratives. Suppose I am right about this. How, then, to bring identification in step with the standards of coherence?

As I have indicated, my narrative view on personal autonomy leads to an account of identification different from the discussed views on autonomy of Christine Korsgaard and Harry Frankfurt presented earlier. All the same, in both their views coherence (integrity) was also at the centre stage. For Christine Korsgaard, a process of self-reflection amounts ultimately to autonomy. A reflective self that acts from above human commitments (ordered in practical identities) chooses commitments that make up the agent's identity. Acting autonomously is allowing the part of the self that is above the commitments, reflecting on these commitments from a universal point of view, to be in control and to give laws. Only by identifying with that reflective part of the self can an agent be autonomous. And it is this reflective capacity that allows him to maintain a coherent sense of identity and so of agency. Acting autonomously means acting on rational, universalisable principles. The guiding rational principles one chooses and endorses *ipso facto* are unifying the person and provide him with a coherent set of reasons on which to act.

On the coherence account I advocate, reasonableness plays an important part. Yet, I reject the special, free-floating status of reasons in the Korsgaardian picture. Reasons, emotions and desires are firmly entwined in the same causal net. I am also less optimistic than Korsgaard about the unifying power of the human rational capacities. As regards my own account, coherence settles for a less constraining power of reasons. A reason is authorised or has agential authority for an agent just in case his

other endorsed reasons can be used in its defense. That is, if they support his adopting or maintaining it; if his set of reasons, so to speak, form a loop: a coherent and interweaving structure of mutual support.

Harry Frankfurt sticks to an account of identification as acceptance or wholeheartedness (in Buss & Overton: 160-1). For him, wholeheartedness is a way of avoiding conflict and progressing towards integration. This comes very close to my own coherence view. Frankfurt's wholeheartedness requirement can be interpreted as coherence among all the significant items of the agent's psychological complex. Unfortunately, Frankfurt concentrates on volitional necessities and argues that certain volitional necessities are not only compatible with autonomy, but even essential to autonomy. He does not take much trouble to put the importance of reasoning on the screen. Matters are, however, more complicated. Not only volitional necessities establish a person's autonomous take on the world. Also reflective endorsement and critical evaluation do.

Identification, on my coherence account, now may be understood as follows: a part of a self-narrative is personally authorised – or sanctioned as one's own – when it coheres with the other parts of one's self-narrative. Incoherent elements of a self-narrative are part of the peripheral self rather than of the real and central self. Does that mean that there is also an unreal self, then? Why is it appropriate to call some elements of a person more or less central and real? Why think that cohering elements are more real and central than non-coherent ones? Let me give three considerations⁸⁸ for doing so:

88 In L. W. Ekstrom 2005: 59-60. To be sure, my concept of autonomy leaves room for resistance, change and emancipation.

i. Coherent elements are *long-lasting*. They are abiding, because they support each other. They hold together firmly, displaying consistency and mutual support and remain relatively immune to change; altering one such element requires change in others. This seems intuitively correct. The core of one's character system should be relatively stable through time. We tend to view the more enduring aspects of a person's psychology as central.

ii. Coherent elements are *fully defensible* by the person in the face of external challenge and neurotic doubt. It is only the cohering elements of the character for which we can mount a wholehearted defense. They fit with – *de facto* constitute – the person's 'party-line'.

iii. Coherent elements are features that one is *comfortably owning*. One tends not to have distress over attitudes that cohere with one's other considered attitudes. One is not frustrated or in tumult. The lack of conflict generates a state of inner tranquility, a kind of liberating comfort with oneself.

This may seem natural and persuasive enough, but here are a couple of skeptical remarks: maybe the craving for coherence is unnecessary. Or maybe it is an outlandish concern with psychic tidiness. Alfred Mele e.g. claims that we can describe what self-directed action is, without positing the existence of any (real) self within the agent.⁸⁹ A human agent is simply a human being who acts. Might not a single anomalous attitude, Mele asks us, represent 'who one really is'? David Velleman stresses a similar point calling an account of autonomy requiring wholeheartedness a 'defensive fantasy' (Velleman in Buss & Overton: 109). Why cannot we say in a certain sort of

89 In his 2003 *Motivation and Agency*, New York: Oxford University Press, chapter 10.

case of explosive conflict with another: 'It was my resentment speaking, not I'?⁹⁰ What tempts people to an integration condition on psychological states is merely a perverse wish to be unconflicted. Moreover, Velleman continues, are not people all conflicted to a greater or lesser extent? People simply live with these conflicts and are able to act autonomously in spite of their inner turmoil.

This skepticism has something to go for it. People are conflicted to some extent. There is indeed inner disorder. There are moments of serious disruption, breakdown, and regression. Granted, but within a coherentist picture of the autonomous self there will be room for conflicting attitudes, as long as these conflicts spur the person to restore the unity instead of keeping him in permanent frustration. Also notice that, on the coherentist conception, we need not view the self as a fixed entity. But even if our identities are in flux, and even though we are full of ambivalences and unresolved conflicts and sometimes overwhelmed by a strong feeling, we are autonomous only if we can claim that our decisions flow from us as the author of the ongoing struggle to negotiate those tensions, to reduce the frustration, and to (re) establish some coherence. Time slice attitudes that fail to bear the proper relation to that ongoing personal project of character development that reflective agents are all engaged in, do not manifest autonomy in a crucial way. The project of character development and reflective self-constitution is a temporally extended process. So, it seems undeniable to me, as opposed to Alfred Mele, that actions or attitudes constitute what an agent is like, a character, along with the faculty for (re)forming that character.

90 D. Velleman (1992). 'What Happens When Someone Acts?' in *Mind* 101, 464-5.

And that a character has more constant, central features and more peripheral ones. Obviously, the agent can be mistaken concerning the central status of his reasons. And other people, as well, may be mistaken concerning whether an attitude is part of that agent's real self.⁹¹ But a person is something like a real character – not a single anomalous attitude – together with a power for fashioning and refashioning that character.

The point we have reached in this section is that for the practical, narrative based account of autonomy I am after, standards of coherence are crucial. Autonomous persons are not totally free to identify with whatever narrative elements they want. This is going too far. Although there is some elbow room, it is crucial to the exercise of autonomy that persons identify with memories, emotions, intentions and the like as their own in ways that cohere into an acceptable self-narrative. Autonomy arises out of an ongoing narrative.

Let me now proceed by subjecting the Frankfurtian view on autonomy to some further refinement. Michael Bratman's ideas on self-governance, generally known as his 'planning theory', will be guiding me. My hypothesis is that my account of autonomy can benefit from Bratman's discussion of the topic in two ways. First of all, his essentially Frankfurtian picture of self-governance balances the role of volitional necessities and the role of reasoning. Furthermore, Bratman attaches great importance to the temporal dynamics of autonomy, in particular to the role of intentions, policies, and plans.

⁹¹ One could think here, in line with Keith Lehrer on epistemic justification (see J. S. Taylor 2005: 150-1), of a game or contest with a skeptical interlocutor who questions one's reasons. Coherence, then, is a matter of mutual defense in the face of external challenge or internal doubt.

3.5 Bratman's planning theory

According to Frankfurt, love provides reasons for action. Bratman's proposal is slightly different. He can grant that what people care about 'is the ultimate touchstone and basis of our practical reasoning' (Frankfurt's phrase in his 2006: 28). Our practical reasoning needs to be grounded in what we love or care about. Simply thinking something valuable is not a sufficient ground for practical reasoning, because we still may not care about it. Bratman can also grant that when we do care about or love something, this may not be explained by a judgment on our part that it is a good thing. (Value) judgments underdetermine a person's sensible answer to the questions of how to live and what kind of person to be. But it does not follow from this that in figuring out how to live (value) judgments have no roles to play. Bratman tells us: 'Put it this way: Two theses that are central to Frankfurt's theory are, first, that the psychological functioning characteristic of inner harmony involves treating what one loves as a reason, and second, that there can be this harmony even if what one loves is bad or evil. A broadly Frankfurtian theory could hold both these views and still go on to say (though Frankfurt does not) that our talk of reasons – that is, our talk of normative reasons – has two faces: it tracks such functioning, and it tracks judgments of value. If we were to take such a view, we would then be in a position to say that, in certain cases, love for what is bad does not suffice for reasons.' (2006: 84).⁹²

⁹² This is not to push Frankfurt in the direction of normative realism. 'It is worth noting that an expressivist meta-ethics can also seek to make sense of this critical thought', Bratman continues in a footnote (2006: 115 note 3). It is obvious that Frankfurt would balk at a realistic, 'Platonic'

Bratman does not say here that there are no volitional necessities in Frankfurt's sense. He just wants to put them in their place. He finds it plausible that they provide background constraints, but they do not settle how to live any more than does value judgment. They significantly underdetermine our answers to this question (2006: 86).

The point will be clear now. The Frankfurtian model 'highlights determination by higher-order attitudes that have agential authority, and the relation of these higher-order attitudes to judgments about the good is not central (... But t)he Frankfurtian model seems to understate the role of value judgment in self-governance' (2007a: 9). Bratman wants to avoid this distortion in his own theory of self-governance⁹³: 'We acknowledge the potential relevance, within a psychic economy that exhibits self-governance, of judgments of the good. Nevertheless, we also, in a broadly Frankfurtian spirit, reject the idea that the connection to these value judgments is at the heart of agential authority' (2007a: 9). As a consequence, Bratman's proposal does not, like Frankfurt's, primarily appeal to satisfaction (wholeheartedness). He does think, though, that his proposal needs to be supplemented by something like Frankfurt's idea of wholeheartedness: *reasonable stability*. Frankfurt has a strong form of stability in

approach. And so does Bratman: 'The Platonic theory overstates the extent to which we can expect intersubjectively accountable value judgment to determine the shape of a person's life' (2007: 9). Frankfurt and Bratman stand for a psychological approach.

⁹³ Bratman leaves open the possibility that there are different modes of psychological functioning that can constitute self-governing agency. There is no uniqueness-claim, only a realisation-claim.

mind, Bratman continues, ‘one that involves an incapacity of the will. In contrast, I seek a model of reasonable stability in self-governance that does not require (though it does not preclude) such an incapacity of the will’ (2007a: 11). Let me now, very roughly, give a sketch of Bratman’s theory of self-governance.

Bratman takes self-governance as a distinctive kind of non-homuncular psychological functioning. The use of ‘self’ in talk of ‘self-governance’ is not a way of referring to a special entity, the ‘self’, but rather a way of alluding to forms of unified psychological functioning (2009b: 236). Agents have a practical standpoint, which consists of attitudes that constitute their stance with respect to relevant practical issues, and, as Harry Frankfurt would say, speak for the agent. When that practical standpoint appropriately guides, the agent governs. Now, according to Bratman, what can constitute an agent’s practical standpoint?

One idea is to appeal to the agent’s judgments about value or reasons. Bratman, as we have just learned, is skeptical about such a Platonic idea, because it faces at least a trio of related challenges. First, sometimes such judgments do not seem to be part of the agent’s relevant standpoint. Huck Finn’s judgment that it would be best to return the runaway slave Jim is a famous example. Second, some practical commitments are part of the agent’s standpoint but are not themselves judgments about value or reasons. Think here of the sort of cases that have been highlighted by Harry Frankfurt. A parent’s love for his child may not be itself a judgment about reason or value, though it may ground judgments about reasons and, Bratman adds, be constrained by some such judgments. Third, sometimes we see our choice as underdetermined by our prior judgments of reasons and value. This can happen in ‘Buridan’ cases, in which one sees one’s con-

flicting grounds as equal in weight or as incomparable. A classic example is Sartre's case of the young man who must choose between the Free French and staying with his ill mother. Bratman's conclusion is that we cannot characterise the agent's standpoint simply by appealing to his judgments about value or reasons. We need to appeal to attitudes that need not be tied to value judgment in ways that would reintroduce the trio of problems. But how? Bratman thinks that a basic step here is to embed those attitudes in a wider role of knitting together the agent's practical thought and action both at a time and over time, thereby helping to constitute the agent for whom the attitudes speak. Let me explain.

Central to our understanding of human agency is that it is quite frequently temporally extended. Human agents are not simple goal-directed agents. Actions take time. And in many cases an agent's present activity involves his grasp of how it is embedded in what he has been doing earlier and what he is on his way to doing. Agents play out their activities over time. And their grasp of the larger temporal arc of their activities is an important element in their guidance of those activities both at that time and over time. Part of Bratman's project is to understand how past, present and future thought and action are tied together in temporally extended agency. And his conjecture is that a fundamental ground for the human capacities for temporally extended agency are human capacities for planning agency.

Planning agency, as Bratman sees it, is a distinctive kind of goal-directed agency, one that involves attitudes of *intention*, many of which are future-directed. Intentions settle relevant practical matters and are normally embedded in *policies* – where a policy is an intention that is appropriately general – and in hierarchically structured larger *plans*. Intentions are

plan-states.⁹⁴ They are guided by the agent's (at least, implicit) acceptance of distinctive norms. Central among these norms are norms of intention consistency and means-end coherence. The former enjoin consistency among one's intentions given one's beliefs; they are responsive to pressures of consistency of the many different things one intends with each other and with what one believes. The latter require that one's structure of intentions be filled in with specifications of means as one sees this to be necessary in the pursuit of intended ends (2009b: 227). These are norms of synchronic planning agency. There is also a norm of rational stability of intentions and plans over time. Although intentions, policies and plans are subject to revision, they nevertheless have a characteristic stability at a time and over time.

Bratman reminds us of the fact that the idea of cross-temporally more or less stable, inter-woven and referentially interlocking attitudes is familiar from the Lockean tradition of reflection on personal identity over time. A central idea of that tradition is that identity over time essentially involves overlapping strands of continuities of attitude. The standard functioning in planning agency of attitudes of intending involves such cross-temporal *Lockean ties*, Bratman says.

So far so good. But why does it matter whether we conform to the norms of stability, consistency and coherence of intention?

⁹⁴ Bratman clearly distinguishes intentions from ordinary desires and beliefs. Ordinary desires are not subject to the same rational pressures for consistency. Desiring things that are not co-possible is all too human. And a belief that one will be doing something later – in contrast with an intention to do it – need not require that one settle on the means to doing it.

Bratman rejects the proposal that these norms are, at bottom, theoretical norms of stability, consistency and coherence of belief. He thinks we should see these norms as norms of *practical* rationality. At bottom is the thought that it is a complex of practical roles of intentions that lies behind these norms. A general disposition to conform to these norms is part and parcel of being a coordinating and planning agent. Conformity to these norms will make it more likely that, for example, one will not trip over oneself and will achieve one's ends. Yet, the best way to make sense of the normativity of intention stability, intention consistency, and means-ends coherence of intention is to regard these norms as stemming from reasons of self-management.

So, let me return to self-governance. Bratman's thought was that self-governance is embedded in the attitudes of knitting together the agent's practical thought and action both at a time and over time, that is, by way of kinds of Lockean continuities and connections. The idea is that a fundamental way in which planning agents like us take a stand – a stand that is involved in temporally extended self-governance – is to go beyond various conflicting desires and concerns and settle on coherent and consistent intentions, policies, and plans. If an agent intends E inconsistently or fails to intend believed necessary means to E, there is no clear answer to the question of where he stands with respect to E. Bratman stresses that tying practical thought and action together in Lockean ways need not be what the agent himself does. It is primarily in playing this broadly Lockean role in organising practical thought and action over time, that an attitude gets a *prima facie* claim to speak for the agent it thereby helps constitute, and so to be such that its guidance can constitute that agent's self-governance.

If this is Bratman's picture of self-governance, it will be plausible that central among the attitudes that constitute the agent's practi-

cal standpoint will be policy-like commitments to give weight⁹⁵ to certain considerations. Such policy-like commitments may be associated with value judgment; they may invoke various forms of affect. But at their core they involve plan-states; and they are embedded within the agent's planning system. Moreover, it will also be plausible that for planning agents like us, self-governance requires consistent and coherent coordination of relevant plan-states at a time. The same goes for the norm of diachronic rational stability of intentions. Cross-temporal (defeasible) constancy of intention will be a normal aspect of self-governance over time. Bratman lists several forms of support for stability over time. First, acting on an intention normally changes the world and there can be a snowball effect because in acting on an intention one gets closer to its target. Second, reconsidering a prior intention takes time and uses various mental resources. It implies costs and risks. Given our limited mental resources, we prefer to depend on general habits and strategies about when to reconsider. Third, one's prior intentions rationally have a default status in later practical thinking. I think it is safe to suppose that attitudes playing the roles that are central to self-governance at a time normally induce associated attitudes over time. Synchronic self-governance is tied to self-governance over time in Lockean ways.

I foreshadowed an improvement of Frankfurt's position and I think we can now see how Bratman's picture of self-governance is a refinement of Frankfurt's approach. To be sure, there is partial agreement between Bratman's theory and a Frankfurtian hierarchical theory on self-governance. Both stress the importance of the conative dimension. Self-governance is volitionally infused.

⁹⁵ Bratman during a discussion in Leiden 2006: 'I have no theory on the idea that something has weight.'

Bratman's intention-based theory is a modest theory of the will. Moreover, his intention-based theory of self-governance does share with the Frankfurtian theory an appeal to conative hierarchy, although the fundamental higher-order attitude is intention (instead of volition) and refers to justifying considerations. Nevertheless, it still may be that the relevant self-governing policies that speak for the agent are higher-order attitudes. New and, in my view, a change for the better is Bratman's move in the direction of reasoning responsiveness. Bratman emphasises that higher-order attitudes are shaped in part by rational pressures and by that he takes a course different from the Frankfurtian.

I mentioned a second reason to discuss Bratman's work: the importance Bratman attaches to a specific role of intentions, policies and plans. The point of departure for his discussion of identity and self-governance may remind us of Korsgaard's. Both authors take it that identity as well as self-governance depend on the fact that human beings are practical and rational beings. Bratman's planning theory also contains elements of Schechtman's broadly Lockean narrative self-constitution view. As Bratman presents it, some intention-type attitudes have it as a central role to support cross-temporal organisation of the agent's practical thought and action and they do this by constituting and supporting *Lockean ties*, – ties that are partly constitutive of the agent's identity over time (2007a: 245). In other words, those Lockean ties help constitute the metaphysical backbone of the temporally extended practical thought and action of one and the same person.⁹⁶ And in Bratman's view,

96 This is a claim about the metaphysics of a strong form of agency (what is it in the world that constitutes this form of agency?), not about normative ideal of integrity or the like (not about 'agency-at-its-best', with a quote from Gideon Yaffe).

the authority of self-governing policies also stems primarily from their broadly Lockean role in the cross-temporal organization of practical thought and action. Intentions, plans and policies supporting Lockean ties earn the authority to speak for the agent.⁹⁷ And, last but not least, drawing on his planning theory, Bratman teaches that self-governing intentions, plans and policies engage norms of consistency, coherence and stability with distinctive rational force. He distinguishes a special subset of psychological elements that closely cohere and are constitutive of the person's character and self-governance. To act autonomously, in his view (and I agree), is to act for coherent psychological elements that are characteristic of the person in question.⁹⁸

With the proposals of the previous sections on the table, I now can open up an approach to a narrative-based account of autonomy.

3.6 The overall picture: a narrative-based account of personal autonomy

My entry point is: we have to analyse autonomy in terms of identification. Identification is fundamental to autonomous agency and 'real' personhood. So, clarification of identification (or its contrary, alienation) should serve our understanding of personal autonomy.

97 As Den Hartogh (2004: 32-34) emphasised, it is unclear how Bratman differentiates between policies and plans on the one hand and Lockean ties on the other.

98 One may describe Bratman's planning psychology correctly as a kind of practical conservatism.

Identification may be interpreted in a neo-Kantian way as identification with reasons, which subsequently are binding for us. That is Christine M. Korsgaard's project (1996: 103-4 and 113). It connects with my own view in that it takes reasons as essential for practical identity and the exercise of autonomy. But I reject the Kantian 'noumenalism' of reason that says: being a person is being determined by practical reason, which places every agent under the same, universal laws. Moreover, Korsgaard's approach comes at a high price, which is that there is little room for affectivity and embodiment in it. One may therefore opt for Harry Frankfurt's hierarchical view as a starting point. Frankfurt is thinking along neo-Humean lines and is making a reasonable case for wholehearted identification with 'volitional necessities' (things we care about and love).⁹⁹ These identifications are not always a matter of what a person explicitly wants his will to be. Rather, they form the background context against which any such wants are formed. In general, identification is a not well-reasoned mental activity. So, Frankfurt's approach also comes at a high price, which is that the role of reasoning responsiveness remains underexposed. Identification in a Frankfurtian fashion is essentially staying willfully connected to the world and establishing and maintaining some volitional environment and unity. Subliminal though it is, we all know that when something goes wrong in the psycho-biological realm of the volitions, this may severely affect the self-governing capacities of the agent in question. Nevertheless, even if it is granted that identification is, for a large part, a process garnering its authority for self-determination from one's nexus of cares,

⁹⁹ Identification conceived as 'caring' looks somewhat familiar with the Humean 'violent passions' which are hardly influenced by reason, but rather by 'calm passions'.

this does not imply that reasoning responsiveness and identification with reasons are not important for personal autonomy. Or so I argue. Frankfurt denies: love gets down to the very essence of personal autonomy; and affect and reason have a derived, instrumental role to play. Frankfurt's solution amounts to a kind of essentialism of the will and I do not want to follow him that far. Let us now turn to the account I advocate.

The volitional model alone cannot plausibly account for the notion of identification. It has to be broadened. This is also what Michael Bratman, another neo-Humean thinker, is observing and objecting. According to Bratman, a broadly Frankfurtian theory has to say (though Frankfurt does not) that talk of identification has two faces: it tracks psychological functioning, and it tracks judgments of value. What is authoritative in identification and self-determination, apart from investments in volitional necessities, are the reasons people reflectively endorse. Bratman proposes an amendment of the Frankfurtian framework in giving a substantial role to understanding and reasoning. Apart from volitions and reasons, there is another important undertheorised aspect of identification: the role of emotions. Admittedly, it is true that emotion is included in what Bratman means by psychological functioning, but I think emotion has its own special character.

Self-governance comes with emotional coloration and salience. Persons experience their emotions as profoundly indicative of, and determining a good part of, the kind of person they are.¹⁰⁰

100 D. Velleman (2006: 347-54) proposes a possible definition of identification conceived as empathy and an exercise in imagination (referring to Richard Wollheim's 'Imagination and identification' in his 1974 *On Art and the Mind*; and to the third chapter of his 1984 *The Thread of Life*).

The measure of emotional experiences has much to do with the extent to which persons identify with their emotions, that is, with the degree to which they ‘take things personally’. Human emotions typically are Janus-faced. On the one hand, they are in touch with the volitional, let us say with the more or less sleeping part of the person. Emotions are anything like the perceptions of the loving and caring state of the person. One can plausibly claim that there is an entailment relation from cares to emotions. It seems very natural to suppose, for example, that caring amounts to an emotional investment in the cared-for object. As a loving parent – one of Frankfurt’s favorite models – I am vulnerable to gains and losses (sorrow and joy) when my cared-for child is affected by various events. Caring leads to having certain dispositions to feel. It seems constitutive to emotional dispositions and states that there ought to be some care revealed or reflected thereby. On the other hand, emotions are in touch with the wide-awake (evaluative) judgments of the person. Reason is never dispassionate, but always in league with at least some of the person’s emotions. Here, emotions often can function as a browser for the (evaluative) judgments of the person; and the person’s rational abilities can control, resist, challenge, and counter his emotions.¹⁰¹

So, I end up with a *multifaceted* conception of identification, and hence of personal autonomy: it covers conative, emotive as

101 See Tappolet (2006). Christine Tappolet develops an account of autonomy inspired by H. Frankfurt and D. Shoemaker (2003), according to which an action is autonomous, c.q. an action of the real agent, when it is determined by the agent’s most central cares, where cares are defined in terms of emotional dispositions. I agree that cares and emotions can be related, but I stick to the Frankfurtian notion of care (love) as a certain motivational structure.

well as cognitive psychological functioning and avoids the partialities of the views Frankfurt on the one hand and Korsgaard on the other put forward. And the account I propose is a *hierarchical* one: I take reflexive consciousness as a requirement for identification and autonomy. What is needed is an appropriate psychological structure: a hierarchy of attitudes about attitudes. Candidate psychological attitudes for the role of supporting the agent's autonomy involve a higher-order concern. The focus is on the ability to step back from one's first-order psychological attitudes and assess them in some way. The model of hierarchy here is the one given by self-management. And self-management is self-reflexive in the sense that it is directed to one's own psychology (beliefs, emotions, carings).¹⁰²

Ultimately, I do not want to make too much heavy weather of this notion of hierarchy. E.g., I do not want to take hierarchy as a case in which certain mental states of a certain kind are applied to mental states of the same sort. Consider a self-commitment example in which an agent anticipates that in the future he might fall for temptation and act contrary to what he now judges best. The agent might take steps now in order to avoid falling for temptation at future times. There is a discrepancy of course between the current intentions and the foreseen future ones. However, there is no point here in having a second-order intention about a first-order one. One could rather speak here of a hierarchy in which higher-level attitudes are constituted by large and complex networks of lower-level ones.

Nor is there any need to leave the ground level. One might have thought there is an ascending ladder here. But the ladder is lying

¹⁰² The target of self-management is not a 'deeper' self. Self-management does not give us any deeper access to the self.

on the ground, horizontally. It takes us nowhere. What is needed and what I mean by hierarchy is some distance, some openness, some latitude, the possibility to step back from one's own mental states and to identify or reject. An autonomous person wants to escape being merely in the grip of his mental states. He wants to be able to reflect on why he treats a particular consideration as a reason and to endorse or reject this reason. He does not want to be 'hijacked' by rigid emotion; he wants to assess or reject his emotions. The same goes for willful mental states like intentions, motivations and carings.¹⁰³ Sometimes identification will

103 Although I put into perspective the multitier account of autonomy, I do not side with Agnieszka Jaworska and Richard Moran who want to get rid of hierarchy. I think their projects founder.

Agnieszka Jaworska (2005, 2009) defends a non-hierarchical view on minimal autonomy. She summarises: 'Altogether, our minimal case of autonomous decision making comprises the following core elements. First, the agent cares about a particular object *P*; this ensures that the attitude that guides the decision making is internal—and thus ensures that it is the *self* to whom the government can be attributable. Second, the agent acts in light of seeing reason to pursue what he cares about; this introduces the element of *governance* itself. And, third, the selection of this reason for action takes place against the backdrop of the possibility of first-order reflection. This ensures that the agent is not merely in the grip of treating caring-based considerations as reasons. At no point in this picture does the need arise to appeal to evaluative judgments or to a hierarchy of attitudes about attitudes.' (2009: 98-9). This is not to deny that it would be better for agents to improve beyond the minimal level of autonomy, she adds. In earlier work Jaworska defined autonomy as the capacity to value and established that '(a) value would almost always involve a second-order volition – insofar as a person considers a way of

take place in a very conscious, emphatic, and articulated way; sometimes it will take place more unconsciously, *en passant*, and dispositionally.¹⁰⁴

So far, all this may not sound like a narrative account of self-governance. But it is. We only have to take my approach one step further. I think my account of self-governance comes very close to Michael Bratman's planning theory, as, for my part, such a planning theory can be interpreted as a narrative theory. The following elaboration of Bratman's analysis of human agency should bring this out.

According to Michael Bratman, any reasonably complete theory of human action will need to advert to three core features:

[*continuation of note 103*] acting correct, she would want her desire to act in this way to effectively move her to action' (1999: 115, note 15).

Richard Moran (2001, 2002) advances a 'transparency view'. Following Moran, identification is not a process by which some psychic raw material is elevated into a higher status and inducted into a true self, as Frankfurt-style accounts would have it. He locates identification at the level of ordinary, first-order and transparent deliberation. His basic idea is that mental states and attitudes are genuinely our own, rather than mere occurrences in our psyche, if we take an active stance toward them: in the sense that they are responsive to our sense of the reasons in support of them.

It is a little hard for me to see how 'reason selection' (Jaworska) or 'transparent deliberation' (Moran) can take place without stepping back and creating some order.

104 This is not to say that it is always easy to distance oneself from one's attitudes. In the case of carings, for example, it seems, to put it in a Frankfurtian tone, that we often cannot but be identified with what we care about.

to reflectiveness, planfulness and a conception of our agency as temporally extended (2007a: 21-46). The three come hand in hand. Let me start with the third one. Agents do not simply act from moment to moment. They are not dealing with a kaleidoscopic flux of unrelated events. That would be like tearing out one page of a book and trying to fathom the words without any knowledge of the other pages. Such agents would give rise to the biblical comment that 'they do not know what they are doing'. The common view is that one and the same agent begins, develops and completes cross-temporally organised actions. Our agency is temporally extended. Human agency is diachronic agency. It transcends the present moment. People do not accidentally act from moment to moment. They do not make their decisions on the spot each and every time they face a choice. They conceive of themselves as agents who persist over time and they commit themselves to future directed plans, which they intend should structure and coordinate their actual decisions and activities. Human agency is constituted through mental time traveling. People exercise their capacity for mental time travel whenever they plan a holiday, or when they take up a study, or when they commit themselves to parenting. It includes the capacity to adopt plans which extend over time. That is to say, people remember what happened last time and use that information to create and inhabit a future scenario. They imagine what would happen in the future based on their past experiences. In committing themselves they find reasons for themselves they will then buy into, they emotionally engage with their future and they care for their own future experiences. We can understand this truth with a broadly Lockean approach, Bratman states. The problem of where an agent stands at a time can be tackled by appealing to Lockean overlapping strands of psychological ties (memories, desires, intentions and the like) of the agent over

time. In my account of extended agency this capacity for mental time travel is narrative in structure. The agent's choices are made in the context of an autobiographical narrative. Autobiographical narratives forge a network of cognitive, emotive and conative connections – Lockean overlapping strands of psychological ties, in Bratman's words – that support the agent's identity and cohesion over time and keep the agent on track.

Bratman highlights the central role of intentions, policies and plans in human agency. That is the second core feature. *Plans* of action play basic coordinating, organising roles at a time and over time. They provide agents with a 'background framework' and are subject to distinctive rational pressures of consistency, coherence and stability (2007a: 26). Plans have certain stability. They are steadying the mind (Bernard Williams' phrase). There is, in normal cases, rational pressure not to reconsider or abandon a prior plan without a reason. Plans are also subject to rational demands of consistency and means-end coherence. *Policies*, for their part, are intentions that are appropriately general in their content. They support treating, over time, like cases in like ways, and doing this as a matter of policy. Plans and policies help constitute and support cross-temporal organisation of our temporally extended agency. Again, this is very close to a narrative self-constituting view and a narrative-based view on agency. Agents have intentions, and there is an intimate relation between their intentions and who they are. By fusing together their intentions into policies and plans and finally into a reasonable biographical narrative or 'lifeplan' agents become persons. To be sure, persons do not always act on a complete life plan or on a complete biographical narrative. Most of the time they act on sub-plans or sub-plots. But all reasons, plans, or stories are hermeneutically disciplined. In the end, making sense of someone's present actions requires a narrative understanding

of a whole personal life(plan). There is another related worry over Bratman's planning theory that a narrative account can put to rest. I think Bratman's planning theory somewhat overemphasises the future-directedness of Lockean (*casu quo* narrative) ties and pays little attention to the past. I can imagine persons, especially older ones, who are no longer intending and planning much, and are mainly living well pleased 'in the past', enjoying the memories of their good old days. Let me now turn to Bratman's last related core feature of agency: reflectiveness.

Humans have the capacity not merely to be moved by desires and inclinations. They have the capacity to arrive at assessments of these desires and inclinations and to take them as reasons for action. Bratman calls this capacity to have pro and con attitudes concerning desires the capacity for 'weak reflectiveness'. It goes beyond what is strictly necessary for purposive agency. Purposiveness is a feature humans share with many other non-human agents – cats and dogs, perhaps – who are not even capable of weak reflectiveness. But there is also the capacity for 'strong reflectiveness'. Bratman calls the capacity to take a stand as an agent – that is, to determine where I stand with respect to a given desire – the capacity for strong reflectiveness, leading to strong agency and full-blooded action, that is, autonomous agency. My proposal is to conceive of autonomous agency as identification with (parts of) one's self-constituting narrative including reasons, emotions, and carings. To act as an autonomous person is to identify with reasons, some of them more backward looking (e.g. self-knowledge, knowledge of the past), some others more forward looking (intentions, plans, policies). Taken together we can conceive of all those reasons as parts of one overall biographical narrative. Part of that narrative we do reflectively endorse. Yet, there is a lot that remains under the radar of reason. An-

other part of the narrative is the result of willful identification with things we cannot but love or care about. Here I stick to the Frankfurtian idea that we identify with ‘volitional necessities’, that is, that apart from what we reflectively endorse we also are deeply invested in things we love. Personal autonomy in some respects seems to be something like acceptance of what I really am. It is like a cheque that I can put my name on, a taking possession of. There is a third and underestimated realm of identification: the twilight zone of emotions that bridges the realms of willful and reflective identifications.

I think we now have under discussion all the desiderata for a narrative based theory of personal autonomy. Bratman’s planning theory turns out to go perfectly together with my Schechtmanian narrative theory of personhood. I end up with a Lockean narrative theory of personhood strengthened with a picture of narrative based self-governance: personal autonomy can be conceived of as identification with (parts of) a person’s self-constituting narrative.

The section is nearing its end. Just to avert possible misunderstandings and to prepare for the challenge of the next chapter, let me make three final clarificatory remarks on narrative based personal autonomy. The first is about coherence. What we can learn from the view I posited is that the autonomy of an action is increased by an action’s springing from an attitude that coheres with certain other attitudes of the person’s psychic economy. Persons normally try to avoid an incoherent psychological complex. They have more coherent, central features and less coherent, peripheral ones; and the more coherent features are the more constant, better defensible and more comfortably owned ones.¹⁰⁵

105 On a given occasion, a mental state on the periphery can be motivationally stronger. Yet, motivational strength is not what determines if an

The more an attitude is central, the more it will determine what the agent autonomously wants, thinks or feels. Bratman takes coherence as norms of practical rationality that apply to intentions (intention consistency, means-end coherence of intentions, rational stability of intentions over time). I am completely on his side. Note that these norms of rationality have to be ‘wide-scope’ norms on sets of attitudes. They are norms that enjoin or reject certain combinations of attitudes. Autonomy is also grounded in (non-modifiable) volitional necessities, and in the ‘feel’ that this is where you now stand; that this is what is important for you and what feels settled and stable. A coherent system of self-management needs to cover the whole mental imbroglio of the person.

Here is my second remark. Persons can only become autonomous persons in relation to others. Nevertheless, an autonomous person can only identify with a biographical narrative from a first-personal perspective, that is, on his own. Consider this example. Suppose I am convinced that I am the last living Romanov and that only I know who I really am. Appearances are against me, and Schechtman would call it a case of diminished personhood. And she may be right: in day-to-day living it will be difficult for me to cash my claim that I am a Romanov. But, this does not prevent me from being an autonomous person. I can identify myself seriously and wholeheartedly with my self-narrative as a Romanov. Others may be sceptical, but they have to respect my autonomy and to treat me as a Belgian philosopher who is sincerely convinced that he is a Romanov.

My last remark is about the distal binding powers of identification. I consider a person to be autonomous as long as he is

action is autonomous. It is the one that is strongest in the sense of being most central to the agent’s psychology.

capable, or will be capable at some point in the future, to renew or to revoke his identification as a particular person. As a consequence, we have to consider patients suffering from severe Alzheimer's disease as borderline cases; they sadly lost their capacity to exercise autonomy, and the loss is – to the best of our knowledge – irrecoverable.¹⁰⁶

3.7 Final conclusions

Let me now return to the bottom line of the theoretical part of my dissertation and draw the final conclusions. At stake was a theory of persons in terms of agency. Korsgaard promisingly proposed to bind up self-constitution with reflective agency and practical reasoning. Nonetheless, in chapter one I also criticised her reason view and argued for a less principled and more instrumental role for reasons that constitute personhood. Neither did I buy lock, stock, and barrel Korsgaard's ideal of the unified person. I suggested that coherence may do better as a standard for self-narratives than unification. My alternative to the

106 Luca Ferrero takes this account to extremes. According to Ferrero (2002), autonomous decisions and commitments have no distal executive or binding power, causal or rational alike, on the agent's future conduct. On the highly counter-intuitive Sartrean account Ferrero is defending, commitments determine the conduct of the agent only if renewed at the time of action based on the intrinsic merits of the case. The agent must take the original decision *ex nihilo* at the time of action. Following Ferrero's theory, the agent's past resolutions do not control her future conduct, except by their being renewed at the time of action. I disagree. Luca Ferrero is pushing autonomy too far.

reason view was a narrative view on persons. In chapter two I presented a candidate for such a theory: Schechtman's narrative self-constitution view. Echoing Locke, Schechtman grounds her narrative account of persons on 'sameness of consciousness'. She adds to the Lockean view that persons cannot emerge without being intelligible to themselves. Self-interpretation or self-understanding is required. This is made possible by *narratio*. Self-narratives are accounts of human action organised in time. Personhood is best achievable in and through narrative self-interpretation. Yet, self-constituting narratives have to meet several conditions, for example, internal coherence and being in step with the facts of the world and the interpretations of others. To be sure, Schechtman is thinking here of ordinary ongoing human lives. Ordinary persons are constituted by lived and told narratives, but not necessarily by worked-up narratives. Schechtman offers us a narrative theory of 'baseline personhood', of the familiar sense of personhood that people find germane in everyday life.

We have seen that, although Schechtman's self-constitution view is a very attractive and defensible version of the narrative approach on offer, it under-describes the embodiment of self-narratives. I paid particular attention to the problems narrativist theorists of personhood run into, if they restrict 'sameness of consciousness' to storytelling. Narrative self-understanding, thus conceived, does not capture the whole person. Fortunately, this is not the picture Schechtman has in mind. Being the same sentient or conscious being implies more than being an interpreting and verbalising creature. Schechtman stresses that sameness of consciousness also implies emotional and felt sameness. Sameness of consciousness, self-narrative and emotion are closely interwoven, without yet invoking any theory of consciousness; that is also what we can learn from Antonio

Damasio's neuroscientific research. Damasio's empirical work demonstrates that autobiographical narratives are intertwined with emotion and feeling. Damasio's work also shows that self-constituting narratives have a bodily, neuro-physiological basis. They build, *inter alia*, on the presence of a core-self. Discussing self-consciousness and selfhood, Antonio Damasio distinguishes between an experiential core-self and an extended narrative self. He makes clear that extended narrative personhood presupposes experiential 'selfhood', that is, that it presupposes the pre-reflective 'self-givenness' of our experiences. By contrast, the self as a subject of experiences does not presuppose narrative personhood. Such a core-self is not an abstraction. It is confirmed by empirical science and it can be observed in 'pure form' in pathological cases. Damasio gives the example of an Alzheimer patient: his life-story is scattered, but the patient still enjoys first-personal access to his own experiences. Patients suffering from severe Alzheimer's disease and neonates are 'core-selves'; they do not join the Persons Club (and maybe adult anthropoid apes are initial persons then, if they have rudimental meta-minds and are capable to connect present, past and future experiences). This is of course a departure from the *Körpervergessenheit* of the original Lockean psychological view on personhood. It clears the ground for a more adequate 'mixed' psycho-biological view. It is a mistake, or so it seems to me, to sort psychology and biology in theories of persons and to discuss personhood only after having left the human body outside the front door. Human persons are co-constituted by human organisms. Body and mind are mixed up, as neuroscience ineluctably demonstrates.

There is another, more inconvenient sense in which Schechtman's narrativist theory of personhood was incomplete: reflective agency has not been given its philosophical due. There is

more to human personhood than baseline personhood. Persons want their actions on their side. They want to rule their actions and they want to be the ones, the persons, who are in charge. They want to be autonomous. Schechtman's theory, other than Korsgaard's, cannot explain what makes persons autonomous. So, her narrative self-constitution view needs a harder look and has to be enlarged on this point. This is the solution I propose: a person acts in a *self-governing* manner and is acting as a governing *self* if he encompasses his autobiographical narrative, i.e., if he identifies with (parts of) his self-narrative. Identification is fundamental to autonomous agency and selfhood. In my view, human self-consciousness is characterised by a hierarchy of attitudes about attitudes. This is not second-order magic. By hierarchy I mean that there is mental room for assessment and endorsement (or rejection) of a self-constituting narrative. Human beings are able to take a stance on their own attitudes. This can take place in different ways. Ongoing self-narratives incorporate strands of reasons, memories, intentions, emotions, carings and the like. These 'Lockean ties', to put the point in Bratman's terms, resist hard and fast distinctions. Yet, we can discriminate between at least three aspects of identification. Identification partly is an act of *reflective endorsement* of reasons for action; this was Korsgaard's line of approach. As we have learned from Harry Frankfurt, identification in part is also an act of *acceptance of carings*. Frankfurt has a keen interest in the conative side of autonomy and the limits of reasoned deliberate action. Autonomous persons are both: rider and horse. Often a rider, if he is not to be parted from his horse, is obliged to guide it where it wants to go. Finally, identification is also partially an act of *emotional approval*. Emotions function as relevance-prompts. There is no general intellectual strategy for relevance search. We may notice a lot of what is relevant in our self-nar-

rative by its ‘feel’. And we may approve of that feel. Autonomy has to do with felt intensity, with experiencing oneself (more) intensely.¹⁰⁷

That we humans can live as autonomous, temporally extended agents – navigating through present, past and future – is made possible in part by narrative structure. Autobiographical narratives play a coordinating and organising role in the diachronic practical thought and action of agents like us. Such ‘life-plans’ help constitute and support the organisation of our temporally extended agency that we value highly. A further point to stress here is that the autonomy of an agent is increased by an action’s springing from an attitude that coheres with certain other attitudes of the person’s psychic economy. In Schechtman’s narrative self-constitution view internal coherence (intelligibility) and ‘external’ coherence (with the facts of the world and the interpretations of others) was already crucial for successful self-constituting narratives. Yet, not only the person’s ongoing self-narration has to be relatively coherent. Also the ongoing autonomous endorsement of it has to be relatively stable and reflecting the person’s central, coherent values. Michael Bratman emphasises this point, and I agree. Self-governing narrative-based agency involves norms of practical rationality: pressures of consistency, of means-end coherence, and of reasonable stability over time.

107 As the following quote from William James teaches us: ‘I have often thought that the best way to define a man’s character would be to seek out the particular mental or moral attitude in which, when it came upon him, he *felt* (italics P.D.) himself most deeply and intensely active and alive. At such moments there is a voice inside which speaks and says: “*This* is the real me!”’, in H. James (ed.) (1920). *The Letters of William James. Vol. 1*, Boston: Atlantic Monthly Press, 199.

Again, *coherence* turns out to be an important standard for becoming an autonomous person. Self-governing persons normally try to go beyond an incoherent stew of intentions, emotions and considerations. In the case of incoherent intentions, policies and plans, or in cases of an incoherent self-narrative (that is, plans writ large) the person does no longer know where he stands and how to govern his life as a person.

Now, one could ask: Why make a choice for the norms of practical rationality listed above? How can we justify these norms? Where do they take their normative authority from? I think my account of personal autonomy, like Bratman's, basically is a prudent one. The relevant reason for the coherence constraint seems to be our reason to govern our own lives. Coherence contributes to the effectiveness of our agency and to the richness of our (social) lives. We are more likely to pursue our ends and not trip over ourselves.¹⁰⁸

It is true that, as agents, we do not only wish to govern ourselves autonomously, we also want to govern ourselves *rightly* ('orthonomously'). We want to lead prudent lives, or beautiful lives, or meaningful lives, or moral lives. And we want to be happy, beautiful, or moral persons. How autonomy and orthonomy can both be realised is a longer story and I will not labor all these complications here. In the picture of personal autonomy I offer, the person's choices and own values are central and exclude nothing. The person's autobiographical narrative and his endorsement of it are what count. The person is autonomous as long as he is competent to endorse or reject his own choices and values. And when it comes to the crunch, the person has to exercise his autonomy alone. Autonomy always carries the imprimatur

108 Vide Michael Bratman 2009a: 47-9.

of a first-personal signature. The powers of autonomy cannot be exercised by other people. There can only be encouragement and the creation of opportunities by others.

From the outset it has been clear that my aim in this dissertation was not to provide a final theory on personhood, if we could have one. My hope was to overcome persistent controversies in psychiatry between biological and hermeneutic psychiatrists. In my view, the two actually are partners, not rivals. And my main goal was to find common ground by focusing on the relevant unit of analysis in psychopathology, persons, and developing a philosophically convincing theory on personhood and personal autonomy in terms of agency. In my estimation, my slightly reformed and enlarged version of Schechtman's narrative self-constitution view is such a theory. To demonstrate its significance for psychiatry will be the task of the last chapter.

4: THE FRAGILITY OF PERSONHOOD

At 4.48

When sanity visits
 For one hour and twelve minutes I am in my right mind.
 When it has passed I shall be gone again,
 A fragmented puppet, a grotesque fool.
 Now I am here I can see myself
 but when I am charmed by vile delusions of happiness,
 the foul magic of this engine of sorcery,
 I cannot touch my essential self.

Why do you believe me then and not now?

S. Kane (2000), *4.48 Psychosis*, London: Methuen, 27

4.1 Introduction

This chapter will engage my accounts of personhood and autonomy with specific practical issues. It finds its pretext in the increasing demand for advance directives in mental health care. Currently, the use of pre-commitment directives in psychiatry is the subject of political and public debate. Pre-commitment directives or Ulysses contracts enable psychiatric patients who become episodically disordered (e.g. in cases of bipolar disorder or repeated schizophrenic psychosis) to anticipate future decisions concerning their treatment. Ulysses contracts allow patients to write down that and how they want to be admitted and/or treated, even under coercion, in foreseeable situations of crisis when they paradoxically ‘are not themselves’. Apart from the

evident need for Ulysses contracts for psychiatric patients, such directives raise many fundamental philosophical questions.¹⁰⁹ One problem is that Ulysses contracts presuppose some criteria to determine when persons are ‘themselves’ in a decisive way. They need a theory of persons to explain how in cases of radical changes a person can be designated on non-arbitrary grounds as the ‘real’ one that is best equipped to make long-term decisions on the person’s behalf. A theory of persons that is capable of distinguishing between full and less full personhood can make clear why a less full person, i.e., someone described as ‘not being himself’, is bound by a directive that has been signed earlier by the full person. Another difficult issue that must be addressed here is how to understand (respect for) autonomy in making use of Ulysses contracts. Completing such a contract, we need a moral justification for overriding the actual will of an – in many respects – still relatively competent person on the strength of an earlier will. In what follows, I will argue that my narrative approach can clarify the identity problems of episodically disordered psychiatric patients. And that it offers an improvement on our understanding of autonomy. In the opening sections of this chapter, I will make some general remarks on psychosis and the idea of pre-commitment. Subsequently, I want to discuss the narrative ‘someone else problem’ and turn to the questions of diachronic autonomy. Finally, I will contrast the use of Ulysses contracts with the use of advance directives in cases of severe dementia and make clear that, again, a narrative approach has a lot to go for it. Nevertheless, the dementia case will also prove to be a borderline case.

109 My concern here is not with legal, therapeutic, policy, and implementation problems.

4.2 Recurrent psychosis

Psychosis is a classificatory and descriptive term, referring to a specific range of illnesses and symptoms, the illnesses being those in which the patient's basic competence as a person is called in question, the symptoms being those which seem to indicate some gross disorder of perception and thought (such as hallucinations and delusions). The *International Classification of Diseases 1955*, published by the *World Health Organization*, lists eight specific psychoses. Four of these, the so-called organic psychoses (senile, pre-senile, arteriosclerotic, and alcoholic), are generally agreed to be the result of degenerative changes in the brain. They excite relatively little interest. The other four, the so-called functional psychoses (schizophrenia, manic-depressive psychosis, involuntional melancholia, and paranoia) arouse considerable controversy within the psychiatric profession and great interest with the general public. This is partly because their symptoms are dramatic and partly because research has (as yet) failed to discover convincing causes for them.¹¹⁰ Let me give an example of functional, recurrent psychosis.

Alexandra is 35 years old, married and mother of two young children. She is running a successful family business. Three years ago she was admitted to a Dutch psychiatric hospital with so-called *in bewaring stelling* (place into care). At that time, her psychiatric past was blank. In the meantime, Alexandra has already had a third maniform psychosis. Doctors think she may be suffering from bipolar disease. But they also consider the

¹¹⁰ I borrow this information from Charles Rycroft in G. Underwood (ed.) (2001). *Oxford Guide to the Mind*, Oxford: Oxford University Press, 177-8.

possibility of schizo-affective problems. The first psychosis entered her life more or less unnoticed. Alexandra was moving house at the time and regulating hundreds of practical things. She slept less and sometimes acted in strange ways. For example, she suddenly called her nephew at the other side of the country and bluntly asked him: 'I know our aunt Ada still must have an old Authorised Version of the Bible. Can you pick it up for me, please?' Or she unexpectedly invited an old girlfriend: 'Please come to visit me, I need you.' It was as if the whole world turned around her and only her. Family members and friends often talked to Alexandra and tried 'to keep her in reality'. At the end her husband felt suspicious of his wife and called the family doctor who diagnosed a psychosis. Then, the situation got out of hand. Alexandra became more and more perseverant, confused, anxious and paranoid. Sometimes she covered herself with a blanket in order to avert her fears. She began to see pictures. Finally, the police and a crisis team appeared and put her in irons after a long pursuit and a short battle. Alexandra was hospitalised and treated under coercion for three weeks. Later, she did not remember the arrest. According to her, she was out of this world for six or seven days. Back home, Alexandra became depressed after a few weeks despite all her medication. During months she got up reluctantly in the morning and was glad to go to bed in the evening after another gloomy day. But she had to go on for her family and for business reasons. Then, her mother died and Alexandra collapsed for the second time. This time, the psychosis was less severe. Two years later, Alexandra's depression seemed conquered and she asked her psychiatrist to reduce her medication. The doctor agreed. The thought was that the depression would never return. But, half a year later, Alexandra showed more and more activities. She began to buy all kinds of gadgets. She reorganised the whole house (next to her full-time job). She slept shorter. And

she embraced everything. She was cheerful, sometimes euphoric, and more and more fanatic. She composed a series of remarkable letters, for example, for various newspapers. In those letters she revealed inimitable connections between various facts. For her it was totally coherent. She anticipated the answers of the newspapers and wrote several rejoinders in advance. Her husband is worried, calls the doctor and makes an appointment. Unfortunately, Alexandra refuses to take the prescribed drugs at home. That night, she becomes more and more inaccessible and chaotic. She starts to clear out her study and works all night long. The next morning she throws the complete content of the shower cubicle out of the skylight. That day, Alexandra is admitted for the second time under coercion. Back home later, she reports that she can remember what happened. It was as if she took a train and started riding. People told her to get out, but she did not want to. Later, the voices disappeared and the train gained speed. There was no fear. On the contrary, it was a fantastic and tempting experience. The train 'shot through the universe'. Alexandra can remember how she was shackled and brought to the police office in a police car. That she was waiting in a police cell for the required documents, and how an ambulance brought her to 'her' hospital. There she got a drug injection and slept for 36 hours. When awake, she recognised the institution and some of the local habits. She was contented and did not resist. This time the whole experience was free of anxiety and the quick intervention considerably reduced the disruption and the period of recovery.^{III}

III I adopted this case history from T. van Willigenburg (ed.) (2002). *Zelfbinding in de psychiatrie. Een empirische, wijsgerige en juridische studie naar zelfbindingsverklaringen in de geestelijke gezondheidszorg*, Den Haag/Rotterdam: NWO/Erasmus Universiteit Rotterdam, 15.

Although this is only one example, it can be helpful to understand two main characteristics of psychosis that Wiel Kusters, a ‘hands-on’ expert reflecting upon psychotic experience, offers us.¹¹² Kusters diagnoses that to be psychotic is like wide-eyed dreaming, without the protection of sleep. Unlike normal dreamers the psychotic sometimes collides with ‘the evidence to the contrary’. But just like in normal dreams the psychotic lives with a totally idiosyncratic calendar and compass, that, besides, are the subject of permanent re-adjustments. The partitions between past, present and future disappear in a psychosis. The past gives no longer hold. Past, present and future seem to be files that can be opened and adapted at one’s own discretion. The psychotic becomes totally absorbed in an intensified ‘now’. He is trapped in a stagnant present.

A second characteristic of psychosis, according to Kusters, is loss of personhood. If we take our normal experience of time as a steady river that carries along the person, the psychotic experience of time, in contrast, is like a whirlpool and the person is drawn into that vortex. The psychotic is losing track and disappearing as a person. He is unable to rise above earlier and later moments in time in order to make reasonable decisions that merge all past, present, and future ‘I’s’ into an integrated person, and that take a prudent turning for these different ‘I’s’.¹¹³ The

¹¹² Kusters W. (2004). *Pure waanzin. Een zoektocht naar de psychotische ervaring*, Amsterdam: Uitgeverij Nieuwezijds, esp. chapter 4. It is common knowledge that ordinary mortals are separated from psychosis merely by a thin partitioning wall (some sleepless nights or an appropriate drug).

¹¹³ Apart from the ecstatic absorption in the here now and the loss of personhood Kusters mentions a third characteristic of psychosis in his book: language and thinking become headstrong. The psychotic de-symbolises

capacities for thinking and deliberating are still in place during a manic or schizophrenic psychosis. And, in a way, the psychotic reasons for action are coherent. The point of it is: the psychotic reasons do not mesh with the reasons, intentions and plans before and after the psychotic intermezzo. They seem to be a matter of distortion or self-deception. Although the *status praesens* of the psychotic may look minimally reasonable and competent, after his psychosis the actor often (but not always) considers the psychotic 'alter' as a different, less stable, and impoverished person. In any case, this is the way the psychotic person often (but not always) evaluates his earlier actions.

Kusters' characteristics of psychotic behavior plainly mesh with the way Schechtman may 'derive' pathologies from her narrative self-constitution view. Schechtman says that to have an autobiographical narrative in the relevant sense is to have an understanding of one's history as unfolding according to the logic of the story of a person's life. Such a life story is missing in the psychosis case. The person loses track and disappears. Moreover, what the psychotic is telling us nine times out of ten will be incoherent. And if it comes to expressions that are internally coherent, they will be out of sync with the facts of the world or otherwise with the interpretations of other people or the interpretations of his earlier self. As Schechtman remarks, intelligibility

and dissolves conventional symbolic representations. Symbols do not longer refer to, but begin to behave like objects that seek to enter into coalitions and battles with other objects.

114 A patient who is psychotic 24 hours a day and 7 days a week is rather exceptional. In most cases, some significant other person is capable of getting through to the 'real' reasons of the person now and then. But these moments break off very quickly.

(internal coherence) and correspondence with the facts of the (social) world are ineluctable standards for successful, normal self-constitution. Let us now turn to the idea of self-binding and the so-called Ulysses contract.

4.3 Ulysses contracts

The idea of self-binding or precommitment is very old. It is a self-limiting act for the purpose of achieving a better outcome, as assessed by the agent's preferences at the time of action, than what would have occurred had he retained his full freedom of action. Until the 20th century the idea of self-binding rarely received systematic attention. It became prominent in the work of Jon Elster (*Ulysses and the Sirens*, 1979 and *Ulysses Unbound*, 2000). Let me therefore introduce the theme with his exposition of some rationales and techniques of the concept.¹¹⁵

This is Elster's definition: 'When precommitting himself, a person acts at one point in time in order to ensure that at some later time he will perform an act that he could but would not have performed without that prior act. As I define it, precommitment requires an observable *action*, not merely a mental resolution. Moreover, the action must be one that creates a change in the external world that can be undone only (if at all) with some cost or effort' (2003: 1754). And here is a historical example he gives: Charles de Gaulle wanted to quit smoking and announced his intention to his friends so that his *amour-propre* would keep him

¹¹⁵ See Jon Elster (2003). Don't Burn Your Bridge Before You Come To It: Some Ambiguities and Complexities of Precommitment, *University of Texas Law Review* 81, 1751-88.

from backsliding (1754). Different techniques are available: adding costs, throwing away options, and delaying consequences. What is common to many cases of precommitment is that the individual can entrust his will at moment *x* to external forces, outside his control at moment *y*, that literally make it impossible for him to change his mind at moment *y*. Here are two other real-life illustrations: i. In a cocaine addiction center for physicians in Denver, physicians with a drug habit can write a self-incriminatory letter to be mailed to the licensing authorities in case they test positive for the drug (1775). ii. If I begin saving for Christmas but find myself taking money out of my savings account, I can force myself to save money by joining a so called Christmas Club, which will be deaf to any demands for withdrawal of the funds before December 15 (1759). So, I may gain a strategic advantage from creating an irreversible *fait accompli*, from burning my bridges, so to speak.¹¹⁶

Elster distinguishes three categories of motivations for precommitting oneself: interest, passion, and reason. By interest he means the pursuit of advantage (money, power, status, or honor). Passion includes *inter alia* emotion, sexual desire, pain, craving for drugs, and madness. By reason he means any impartial attitude motivated by concern for the common good or for

¹¹⁶ Ulysses contracts are an example of so called *external commitment*. What makes external commitment external is that it is independent of the decision of the agent at the time of action. The salient explanation as to why Ulysses does not give in to the temptation of the Sirens is the fact that he is tied to the mast. Forms of *internal commitment*, on the contrary, imply that the agent decides to pursue a course of action and subsequently acts as intended, where the earlier decision plays a determining role in the justification of the later action.

individual rights and duties (1755). For my purposes, I will focus on passion. In binding oneself '[t]here are two future-directed intentions: a Time¹ prediction of the state in which he will find himself at Time², and a Time¹ preference for what he should do at Time². If at Time¹ he can perform an action that will ensure that he will do B *and* if the cost of performing that action is less than the value difference between A and B (as assessed at Time¹), he has an incentive to precommit himself. If he is rational, he will' (1755).

The traditional view is that an act of precommitment is an act undertaken in a 'cool' moment of tranquility and calmness to prevent the agent from harming himself (or others) on future occasions, when he might be in a more turbulent 'hot' state. To put it metaphorically: precommitments are chains imposed by the agent when sober on the agent when drunk. This is also what Ulysses, the standard bearer of precommitment, did in order to enjoy the songs of the luring Sirens. Standard cases of precommitment involve interest-based or reason-based precommitments against passion, in this case being enchanted by the Sirens' music. In passing, Elster also warns his readers that, because of its rigidity, precommitment may create problems as serious as those that it is supposed to solve. In practice, for example, we may not be able to draw the line between temptations and legitimate exceptions. Let us now turn to precommitment in medical practice.

Self-binding is a strategy considered to be widely used in medicine and health care. The following four uses of precommitment are more or less accepted: the standard advance directive, the dementia case, the mental health directive (i.e. the Ulysses contract) and the surrogacy case. Standard advance directives, advance directives in cases of dementia and Ulysses contracts are *instructional directives*. They give instructions for what sorts of medical interventions to provide or withhold from the patient

while incompetent (resulting from a coma, severe dementia or mental disease). Surrogacy cases use *proxy directives* that designate a surrogate decision maker to stand in for the patient if he is incompetent. Since surrogates need a basis for making the decision that the patient would have preferred, surrogates can benefit from a clear instructional directive. I will now concentrate on the third application: the mental health directive.

Ulysses contracts in psychiatry aim to ensure treatment in cases of pathology which are characterised by alternating periods of good mental health and severe disorder. Ulysses contracts enable mentally well persons to commit themselves now to a particular course of treatment at a future time, if they suspect they will not be willing or able to follow that course of treatment at that future time. The terms of Ulysses contracts state that if certain symptoms of a relapse appear, the psychiatrist is to invoke the Ulysses contract, regardless of whether the patient consents to commitment at that time, and before the patient meets the threshold of whatever legal criteria are in place in her country, e.g. in the Netherlands danger to self or others. Note that the intervention can be coercive and that, at that very moment, the patient is still legally competent to refuse civil commitment (unless the patient poses a threat to herself or others, in the Netherlands she is legally competent to refuse treatment). As I said earlier, there are difficult issues here, so let me try to take some clarifying steps. As a first move I want to discuss the ‘someone else problem’. It is sometimes argued that in cases of radical changes – like psychosis – the person in question is becoming someone else.¹¹⁷

¹¹⁷ A much-discussed example of the someone else problem is Derek Parfit’s character of the nineteenth century Russian nobleman, a *locus classicus* in the literature on self-binding: ‘In several years, a young

Do we have to uphold that Alexandra becomes another person when her illness resurfaces?

4.4 Psychosis and the someone else problem

The someone else problem touches upon the distinction between narrative identity and numerical identity. Turning to the question of numerical identity is calling for a criterion of numerical identity across time, that is, for a criterion of what makes something one and the same thing as itself at different times.

[*continuation of note II7*] Russian will inherit vast estates. Because he has socialist ideals, he intends, now, to give the land to the peasants. But he knows that in time his ideals may fade. To guard against this possibility, he does two things. He first signs a legal document, which will automatically give away the land, and which can be revoked only with his wife's consent. He then says to his wife, "Promise me that, if I ever change my mind, and ask you to revoke this document, you will not consent." He adds, "I regard my ideals as essential to me. If I lose these ideals, I want you to think that I cease to exist. I want you to regard your husband then, not as me, the man who asks you for this promise, but only as his corrupted later self. Promise me that you would not do what he asks." (Parfit 1984: 327). Parfit portrays the Russian as telling his wife that his younger self is his real self, and that if he loses his ideals she should regard him as effectively dead. Though a tantalizing case, it is fraught with problems and caused a lot of confusion in bio-ethical discussions. For now, let me remark that the case is significantly different from Ulysses or the psychotic: the young Russian does not anticipate that he is going to become irrational or *non compos mentis*. For illuminating discussions of the Russian nobleman case, see Korsgaard 2009: 185 et seq. and Laden 2009.

Turning to the numerical identity question is turning to metaphysics ('what is a person essentially?'), or to what Schechtman coined the reidentification question. My aim is not to discuss this ontological puzzle here, but to make a detour. It may be that narrative identity presupposes numerical identity, but my focus here is on practical, narrative identity: do episodically psychotic persons experience their psychotic characteristics and actions as attributable to one and the same person, to wit themselves, or to someone else? To answer this question let me sketch out some of the results of 58 interviews with patients, proxies, and psychiatrists for a recent Dutch research program into the advisability of Ulysses contracts in mental health care.¹¹⁸ All the respondents have been questioned about their understanding of severe deregulation and personhood. Let us begin with the moments of florid psychosis.

The research results make clear that a severely deregulated patient hardly can be called a person – in the sense of the narrativist theory of personhood I am defending – during an interval of manic or schizophrenic psychosis. The psychotic loses the reflexive temporal grasp on his life story. Psychosis often seems to be an experience of mere sequence, as in dreams or in moments of high fever, that is destroying the internal narrative coherence. Furthermore, the external coherence – i.e. the connection with reality – is also under pressure. The patient lives 'in his own film'. The schizophrenic patient is very convinced that she is Maria Magdalena; and the manic patient imagines himself to be

118 See T. van Willigenburg (ed.), S. Gevers, I. Varekamp & P. Delaere (2002). *Zelfbinding in de psychiatrie. Een empirische, wijsgerige en juridische studie naar zelfbindingsverklaringen in de geestelijke gezondheidszorg*, Den Haag/Rotterdam: NWO/Erasmus Universiteit Rotterdam.

the immensely rich president of a multinational. All the signs are that severe psychosis is a borderline case of personhood. Full-fledged narrative self-constitution is no longer practicable then. In *severe* cases of psychosis the patient temporarily is a 'no-more-person'; the narrative 'cage' collapses, internal and external coherence disappear, and the person falls to pieces. In *less severe* cases psychosis remains a disruptive process that sometimes alters the earlier person dramatically and diminishes the capacities to live the life of a person.

Now, how do patients cope with their psychoses in healthy episodes of life, when their temporal grasp on the narrative is restored? In the interviews of the Dutch research program three possible reactions to recurrent psychosis become apparent: i. Some patients describe psychosis as a brain disease, i.e., a useless individual natural disaster that has to be overcome. ii. Patients, and especially psychiatrists sometimes picture psychosis as a cognitive problem that has to be identified to restore the rational faculties as soon as possible. iii. Finally, psychosis is often perceived as a moment of purification and a step on the way to better self-understanding. For some patients psychosis, besides being a disorder, also mounts a challenge to cope with that disorder. In all cases (perhaps it would be unfair to press this relation too hard for the first group), episodic madness becomes part of the self-constituting narrative. Patients by and large fit their experiences of madness into their narratives. Sometimes psychotic crisis is described as a 'biographical correction' (e.g. of being too well-adapted) or as a 'struggle to find a guideline'. Some patients stand up for the meaningfulness of the psychotic experience that can turn into a reassessment of a person's values and objectives in life. In fact, psychoses typically break out when a person is trying to take his life in his own hands, and very often interfere with life events like leaving

home, divorce, pregnancy, or the loss of a beloved. Psychosis then seems to be an ultimate attempt to process and integrate 'strange' experiences that throw one's narrative out of gear, and to keep the life story together; or, as one patient put it aptly, 'to translate oneself with more coherence'.

A further issue, and the one that occupies me here, is how patients and their psychiatrists interpret the psychological discontinuities and dramatic changes of the person. Do they see the patient in moments of psychosis as another person, that is, as a human being with a different identity? Seventeen interviewed Dutch psychiatrists come to judgments like the following: 'Of course he is himself. What comes to the surface during his psychosis is a part of himself. My concept of a person is a layered one. At the bottom there is a primitive, archaic, unconscious part. On top of this dark morass there is a central control room that is structuring, steering and regulating. During a psychosis, the bottom layer comes up, because the protecting upper layer temporarily is eroded. A part of the person is taking over and performing impulsive actions, without a sense of reality or something that functions as a brake. Such a person is going astray. He is not somebody else. He is giving the crazy part full scope. But he is still the same person.' Psychiatrists are likely to describe psychosis as a loss of rational capacities and a relapse into primitive stadia of functioning. In their view, talk of someone else in cases of psychosis is going too far. The psychotic is the same person, but the person in a different way. This may sound sensible to us, but is it also the way patients see it?

Here is how a patient describes herself in manic times: 'When I'm deregulated, I look like a difficult child. I cannot sleep anymore and I become coercive and annoying like a child that is overtired. I also have the innocence of a child: I am not aware

of the consequences of my actions. I am living very fast without being considerate of others. It is getting on my nerves when they do not have my speed. I am working on big projects and I am enjoying it. Life is very intense in periods of madness. In a short space of time I am coping with lots of past experiences. I am re-experiencing past emotions. It reminds me of the stories of people who have had a near-death experience. These people tell us they overlooked their whole lives in one single moment. During my first manic phases, I was behaving like a pre-school child. In later episodes, it felt like a child of twelve years old. My madness has a processing function. It is a way of dealing with past experiences. Without madness I would become an icicle. Therefore, they have to keep hands off my madness and not to sweep it under the carpet as a symptom of something inconvenient. Am I myself, when I am mad? Yes, I'm simply myself. The child is part of me.' What this answer and many others suggest is that patients, just like psychiatrists, do not attribute practical significance to the notion of a radically other person or a totally different identity in cases of psychosis. The stories they tell about the psychotic patient for the most part are stories of personal survival and continuation.

Patients, however, clearly do not always want to identify with their psychotic experiences. Consider another example of mania, about which the patient during a lucid period reports: 'When I'm manic, I'm going beyond myself. I'm doing hundreds of things at once. They are spinning through my head. That is a quite chaotic experience. I'm constantly prompted to do something, as if I have to act out a story or so. Later on, when I'm discharged from the hospital and back home, most of the things that happened are forgotten. I can only remember fragments. My boyfriend then recalls how I behaved and talked. Sometimes it's hard for me to believe him. Disorder is

something coming over me. It's just embarrassing for me and my friend, and it does not have any function or significance at all. Absolutely not! At that very moment I'm just like a fallen cupboard with the drawers and the contents all over the floor.' For this person psychosis for the most part has the character of a series of mental and physical happenings in her history, rather than a series of her own experiences and actions. Psychosis seems to be an anomaly that has to be excluded. The personal change is experienced as alienating and the person is reluctant to endorse that element of her biographical narrative. She cuts it and denies it 'personal character'. But other patients are doing the reverse. They identify wholeheartedly with elements of their psychotic experiences. As the last but one example revealed, psychosis can be a way of coming to grips with earlier experiences (as a kind of dream work). It may even happen that psychosis for the main part is experienced as positive. Psychosis can be experienced as an attractive exceptional state of inner agility; one is more sensitive, porous, fulfilled, and creative then. Psychosis can even be made artistically productive. In such cases it seems counterintuitive to deny a psychotic episode 'personal character'.

Let me proceed on the assumption that this handful of opinions of psychiatrists and patients on recurrent psychosis and personhood are somewhat representative, and allow me to draw two important lessons from these cases. One is that the vexed claim that during a psychosis the mentally well person is gone, flies in the face of our common personal interactions. No one is dying and no new person is born during a serious psychosis. At least, this is how the environment experiences it. And significantly, it is also how the patient suffering from recurrent psychosis experiences it. There seem to be no practical reasons for supposing a number of different persons in cases of recur-

rent psychosis.¹¹⁹ The psychotic is still the same person, but the person in a different way. He is not literally someone else. Alexandra suffering from bipolar disease may find the prospect of disruptive psychosis ego-dystonic. But she believes and feels that *she* will in fact be that psychotic individual: ‘Whoever that deranged woman would be, after all, it would be me.’ Possibly it will prompt her to take measures. And this leads me to the second and related lesson to be drawn from the Dutch research results: persons suffering from recurrent psychosis seem able to bridge the narrative gap between their non-psychotic and psychotic selves. Let us consider this in more detail.

During Alexandra’s psychotic surges, narrative identity clearly is disrupted. Psychosis usually means a massive infringement on the synchronic and diachronic coherence of the self-constituting narrative. This is a matter of degree. In extreme cases, when the illness resurfaces, Alexandra will no longer be capable of narrative self-constitution. She will have no story at all then. Less

119 Recent findings in philosophical psychopathology seem to confirm that practical approaches to the question of personal identity clarify matters considerably. The concept of personal identity is central to the subject matter of psychopathology. Psychiatrists are puzzled by questions like ‘what constitutes a person?’, ‘what does it take for a person to persist from one time to another?’ and ‘how does mental illness impact on identity?’. A lot has been written on disorders that pose problems for personal identity, especially on schizophrenia, Alzheimer’s disease, gender identity disorder and DID/multiple personality disorder. The outstanding 2006 *Oxford Textbook of Philosophy and Psychiatry* has a whole chapter on personal identity. The authors of this textbook advocate a Lockean ‘self-tracking capacity view’ (766-7). Becoming a person is a matter of know-how, of skills, i.e., of having the proper capacities.

severe cases are different. During mild psychotic surges Alexandra's reasons, emotions, and carings may change. The psychotic Alexandra may seem to be more or less psychologically removed from her narrative-carrying past. Being in a manic state, she may identify with different, atypical narrative elements. Narratively speaking, she may look like a different person. But in lucid periods, Alexandra will think of psychosis as an integral part of her life, though one she has outgrown most of the time. In other words, Alexandra's moments of psychosis will have autobiographical significance. She might try to look for a way to integrate those seemingly contradictory moments of her life into her life as a whole. How so?

Well, this is how persons usually deal with disunity. Persons have a temporally extended nature. They always have to balance different sets of (potentially conflicting) past and future commitments, projects or plans 'with an eye on the whole person', turning them from a collection of disparate commitments, projects and plans into a kind of personal life and a coherent person. To be sure, the nature of the integrity produced can be rather loose. It is not acting in light of a grounding metaphysical fact of integrity, but an attempt to fashion integrity out of seemingly disparate elements. All that is needed is a person whose parts are, as it were, on speaking terms; in other words, whose parts are coherent. Coherence, as described in the sections 3.4, 3.5 and 3.6, is the practical norm of rationality at work here. So, we have to conceive the unity of the person as a practical achievement. It requires ongoing work. Life is full of centrifugal pressure. Persons (and especially psychiatric patients) have to deal with disruptive, 'heteronomous' (to echo Paul Ricoeur) experiences. And, in a very natural way, they strive for a practical guide forwards, for a picture to hold on to, to hold themselves together, to make of their various parts something whole, some one thing: 'me'.

It is such a practical conception that will tell Alexandra what sorts of arrangements and compromises are consistent with her integrity, and thus will help her guide her choices. The various parts of her person can have their say, as it were, in Alexandra's deliberations what to do. Maybe, at the moments when psychosis tempts her, Alexandra's sense of wholeness is precisely what she is lacking: why should she act in line with a standard whose ground does not have a firm hold on her? Or maybe psychosis is being single-minded in her pursuit of one of her aims and failing to take heed of other current aims. These could be reasons for Alexandra to banish the psychotic intruder from the life of her healthy person. But she may also identify with (parts of) her psychotic variant. She may also value the psychotic person stage as a way to restore the balance and the unity. Here we have to be careful not to identify automatically disorder with fractured life and scattered personhood. Which leads us, I think, to the hard problem for a narrative account of the use of Ulysses contracts: the problem of authority. Different stories are brought into play. The mentally well person has a story to tell. The prepsychotic also has a story to tell. The psychotic has a story to tell (except in the case of extreme surges). Why is one story better than the other? What is it that lends the better story its authority? On what grounds, to put it roughly, can we defend that a person's self-narrative at T^1 has authority over the person's self-narrative at T^2 ? I will expand on this critical point in the next section of this chapter.

The upshot of the present section is twofold. We have learned, firstly, that the diverging experiences of psychotic episodes, however dissociated, by hook or by crook immediately and inevitably take part in the overall narrative self-constituting process of the person suffering from recurrent psychosis. And, secondly, that there is no practical reason for presupposing a practical, narra-

tive ‘someone else problem’ in cases of episodically disordered persons suffering from recurrent psychosis.

4.5 Taking coherence seriously

To clear up the problem of the authority of self-narratives in the event of recurrent psychosis, we have to return to the subject of diachronic autonomy. Ulysses contracts are often justified and recommended as instruments of self-control, that is, as a method to get control over one’s whole life, especially in moments of crisis when one has lost one’s grip of the situation. In other words, as a way of enjoying autonomy ‘at a distance’ or as a way of ‘extending’ autonomy. How to make sense of such a notion of ‘remote’ autonomy?

I think the case can be treated on the same lines as the renowned case made by Ronald Dworkin in his *Life’s Dominion* for people suffering from dementia (I will indicate the differences with dementia later in this chapter). Here are the elements of his argument. Dworkin begins with the assumption that the competent¹²⁰ person and the disordered person are one and the same person, that is, the same single object of inquiry. If only one person is involved, we have to decide what is in his overall best interests when he has different and competing interests at different stages of his life. The questions that have to be resolved are questions of intra-personal balance. They have to be resolved

¹²⁰ I do not consider competence as a legal concept here, but as a normative concept. To be competent is to have sufficient mental capacities for a given task, i.c. for the exercise of autonomy. Competence is a matter of degree. And to credit a person with competence is a normative decision.

within the life of that single person. Only *one* person's autonomy and overall best interests are at stake. Dworkin further distinguishes between two kinds of interests people have – critical interests and experiential interests: 'People's experiential interests are their interests in enjoying or avoiding certain kinds of experience: their interests in having pleasure, for example, and avoiding pain. Their critical interests are their interests in having a life that is a good one judged as a whole. Of course experiences that occur in a person's life after he has become seriously demented are not experienced by that person in a competent state. But it does not follow, and it is opposed to most people's own judgment, that what happens to someone after he has become demented cannot affect his critical interests' (Dworkin in Burley: 367-8). In Dworkin's view, a person's critical interests survive serious and permanent dementia or, as in our case at hand, recurrent severe psychosis. Alexandra does 'live with' impending psychosis; in some way or another psychosis affects what kind of life she will have had, that is, psychosis affects her critical interests. These critical interests, Dworkin claims, are the same at her different life stages. When Alexandra dreads living on in severe psychosis, she may dread that a stage of her life that she lives through will soil that life. And she may take this as a reason for trying to guard against it. Now, according to Dworkin, how do we best respect the person's autonomy overall? Autonomy ultimately serves what Dworkin calls the 'integrity-value' of a person. And Dworkin proposes to give powerful weight to an exercise of what he calls *precedent autonomy*, i.e., the future-oriented or prospective autonomy exercised by a person when most competent to exercise it. Alexandra in a mentally well and competent state can decide what set of policies does best to protect her person's autonomy overall. She can lay down this set of policies in a Ulysses contract. People might have other

good reasons for treating the psychotic Alexandra as she actually wishes, rather than as she once asked in her Ulysses contract. But that would violate rather than respect her (precedent) autonomy, in Dworkin's view.

I think the Dworkinian concept of precedent autonomy easily connects with the narrative view on personhood and identification I set out earlier.¹²¹ A competent person is a self-narrator and (where circumstances permit) self-shaper who cares about continuing as such. Such a person wants to consciously, deliberately shape his own personality and life direction. He prefers an inner story whose overall character and direction matter to him. He prefers to have and live such a story and he wholeheartedly identifies with it. He identifies with the things that really matter to him, in other words, with what is true to his self-conception (with his 'critical interests' or 'integrity-value' in Dworkin's terms).¹²² Now, one could argue that, during a psychotic surge, the competence of the person is diminished and identification then results from influences that the person in question, on careful reflection in competent times, would consider alienating. During a psychosis the person is not in the position to identify with his 'integrity-value', so to speak. Therefore, his precedent autonomous identification is weightier than the actual enjoyment

¹²¹ Similar work is recently done by David DeGrazia in *Human Identity and Bioethics*, who is also taking his lead from Marya Schechtman's work on narrative self-constitution. Other theorists who appeal to narrative in discussing advance directives are Howard Brody, Mark G. Kuczewski, Ben A. Rich, and Jeffrey Blustein.

¹²² Notice that, in this picture, identification is clearly evaluatively loaded. Critical interests are referring to some standards. In Dworkin's analysis, critical interests are interests in having 'a good life'.

of autonomy in his psychotic state. But how so? In finer focus, on what specific grounds can we decide that the (pre)psychotic story does not speak for the person in question, and the non-psychotic story clearly does so? How to pick out the authorised spokesman more precisely?

One could think of privileging stable and long-lasting decisions as more authoritative. Or one could think of privileging the decision that exhibits the greatest degree of rationality (this is what, e.g., M. Quante has in mind in 2002: 287-295). My own proposal is to lend authority to identification with one's self-narrative. So, the question now becomes: Why does identification in one case outclass identification in another? In the previous chapters I have argued that *coherence* turns out to be the standard of becoming an autonomous person in two ways. In chapter two, I defended that basic to be a person is to exercise narrative capacities for self-interpretation that brings about the integration of the person over time. And that the constraints Schechtman proposed for successful narrative self-constitution are clearly apropos: internal coherence and coherence with the facts of the world and the interpretations of other people. In the third chapter, I used this notion of narrative identity as building block for my discussion of autonomy. To be an autonomous person is to identify with one's self-narrative: 'This is the person I want to be and the person-life I want to lead.' And I advocated that the autonomy of an agent is increased, again, by an action's springing from an attitude that *coheres* with certain other attitudes of the person's psychic economy. Not only the person's ongoing self-narration has to be relatively coherent, but also the ongoing autonomous identification with it has to be relatively stable and reflecting the person's central, coherent values and interests. So, coherence turns out to be the key. To see how this may work, let us consider again Alexandra suffering from bipolar disorder.

Let us suppose, for the sake of argument, that Alexandra suffering from bipolar disorder¹²³ is issuing a Ulysses contract in order to prevent harm.¹²⁴ The contract is intended to ensure that the once fully competent Alexandra retains her autonomy and right to consent to medical treatment in the event of becoming less competent to consent. Why respect her precedent autonomy in this case? The question is especially urgent while many people in a (pre)psychotic episode have sufficient mental capacity for a relatively competent refusal at the moment of intervention with coercive measures. That is also how the Dutch law system formulates it. The law takes her departure from the idea that a person is competent as long as there is no serious danger for herself or others. So, what is in need of justification here is that we submit a legally sufficiently competent person to coercive measures in the name of her earlier competent decision. In many western jurisdictions, this looks like enslavement of a later self by an earlier self. There is no legal basis for imposing coercive measures on a person one is contracted with, if that person is legally competent and refuses. In such a case, she has simply changed her mind. Her wish to prevent serious harm has disappeared; it is gone and there is nothing left to discuss. Under Dutch law, a competent

123 Patients suffering from bipolar disease go through cycles of splintering mania and subsequent, inevitable flameout. Their minds first pelt out of bounds and then out of commission. Kay Redfield Jamison's essay 'This life, this death' offers a penetrating description of the (dangers of the) manic-depressive condition (2000: 52-69).

124 Let me recall that two types of advance directives are common in medical practice: *instructional* directives and *proxy* directives. My main focus will be on written instructional directives planning advance treatment for mental illness.

person is free to refuse the deal and walk away without coercion. It is precisely the ability to entertain second thoughts and to change direction that is deeply bound up with what makes a person an autonomous human being. So, an extra justification for imposing coercive measures is indicated here. Framing the problem in terms of autonomy one could say: there is a conflict arising within Alexandra's autonomy. She wants to direct her own life autonomously and in her psychotic state refuses compulsory treatment. But she had an earlier autonomous wish being 'truer' or more deeply held that also commands respect. What must we take as Alexandra's most authoritative will? During her psychosis, her rational and deliberative faculties possibly will be knocked out of action. This is of course a serious condition. But psychoses do not always lead to cacophonous disorder. Maybe some capacities for reasoning and identification will remain in place during her psychosis; maybe her psychotic reasons for action will sound pretty coherent. Nevertheless, while executing a Ulysses contract, Alexandra knows that these reasons do not mesh with her plans, intentions and reasons before and after her psychotic state. She considers them as a matter of distortion or self-deceit. The psychotic Alexandra may think of herself as very reasonable and competent, but looking back when the psychotic episode is over she may evaluate her psychotic 'predecessor' as a cloudy person. In healthy times she knows that she cannot (completely) trust her psychotic reasons and that she may cause serious harm to herself and others in doing it nonetheless. She may say: 'Well, of course, this manic person, that's me. But it's not fully me. I don't want to be like that. I do not identify with that part of me. Please, protect me then from what I want.' So, it seems very understandable that Alexandra in good times will look for some means in order to prevent such damage in mad times. That is, in order to control a little her reasons, utterances

and behavior and to preserve some unity in her person-life.¹²⁵

It is beyond serious question that self-narratives can be misconstrued and pathological.¹²⁶ Problems may arise in the same capacities that give people narrative competency. A self-narrative can be impoverished because it is fragmented or because it is too focused (i.e., a monomaniac and rigid narrative). It can become a flamboyant delusional narrative (e.g. a person who deeply believes that his entire mental life is controlled by the Intelligence Service). Pathological cases face us with the difficult task to find some *threshold-value* for self-narratives in order to determine where a self-narrative becomes defective. This is a big topic, so I can only try to sketch in a rough way how I think that story should go in case of bipolar disorder. In general, I would say that at the heart of narrative based autonomy are, as Bratman remarks, the person's intentions, policies, and plans, all of them constrained by rational norms of consistency, means-end coherence

125 This does not imply that a patient composing a Ulysses contract always urges that the wholeness or the unity of his person is his reason for self-commitment. Ulysses contracts (and personal autonomy) do not require mental gymnastics.

126 To the best of my knowledge, there is no natural, principled boundary between normal and abnormal conditions of suffering. It is not possible to disentangle the enmeshed relation between natural, social and personal facts that constitute mental health and psychopathologies. The point of greatest consensus on what can be called pathology is a harm-based approach. Patients putting themselves in harm are driven to the psychiatrist by the poor outcome of their condition and their inability to (socially) function as wished. For a thought-provoking discussion on psychiatric disorder see Derek Bolton (2008), *What is Mental Disorder? An Essay in Philosophy, Science, and Values*, New York: Oxford University Press.

and reasonable stability in time. The case of bipolar disorder may illustrate the point.

To my mind mania and depression both are threatening the coherence of Alexandra's self-narrative. As a depressed person Alexandra may lose her interest in many things that normally make life worth living for her. Depression diminishes the pull of the various concerns that give shape and color to her life. It is gradually losing herself: Why bother about having an occupation? Why bother about the birthday of her children? Why bother about being a mother? Why dressing herself up neatly in the morning? Why go on? Such a growing disinterestedness may destroy her connectedness to the various self-conceptions under which she values herself. It may lead to inconsistent beliefs, emotions and expressions of will. And it may lead to a very unsteady *curriculum vitae*. Alexandra being in a state of mania is a bit of the reverse. Now, she experiences an increase in the sharpness and focus of attention. Her general level of mental energy grows spectacularly. And she may engage in lots of new projects, some of them mutually inconsistent. This may lead to the same kind of alienation as depression does. New projects and grand ideas are adopted, new concerns fill her life and she seems to lead her life in the highest gear. But she becomes alienated from the concerns and self-conceptions that are constitutive of the person she is. She may devalue her old concerns as parochial, petty and trifle. Like in a depression she will experience a further loss of connection to the concerns and practical identities that are constitutive of being Alexandra. And she will lead a life robbed of stability.

Note that, in my narrativist view, it is the first-person perspective that is authoritative for narrative self-constitution and autonomy. Alexandra's self-narrative has to be an *inner* story. As narrative identities involve *self*-conceptions and autonomy is *self*-

legislation, the privileging of the first-person standpoint is the only reasonable option.¹²⁷ Other people do not have as much authority in determining someone's narrative identity as the protagonist herself does. Others can employ qualifications such as: this person is someone who deeply believes that his thoughts are controlled by the Intelligence Service. Nonetheless, others do have some assisting role to play. What role?

While living in a state of mania or depression, Alexandra's self-interpretation becomes largely unrealistic. She turns into a severely deluded person and part of her condition is that she does not realise it. This is of course defective personhood, which can be corrected by what other people know about the world and about Alexandra. I stressed earlier that self-constituting narratives by definition are co-constituted and kept in check by our fellow creatures. To some extent a self-narrative is a joint construction. Personhood is socially embedded and can only survive in the broader community that supplies much of the content of self-narratives; in the 'space of persons' so to speak. Becoming a person is impossible without bringing a social and cultural background into play, a network of personal relationships (recognition and critique) within a shared living environment. So, while knowing that in a state of mania or depression she will not be capable of correcting her self-narrative on her own, Alexandra has an option to organise it 'with a little help from her friends' and, in our case at hand, to consider a Ulysses contract

127 Agreement with DeGrazia here (86). Narrativist Mark G. Kuczewski argues that persons are objectively part of a larger group. Therefore, if the group survives, so does that person – or at least his or her interests (1994: 42). I reject this view. Personhood requires the first-person perspective of a subject.

that authorises other parties to interfere in her future actions. Alexandra has particular interests, values, cares and the like that confer her individuality and that matter to her. And expectations are that these interests, values, and cares are coherent and do confer and matter to her for her whole life, or at least for a longer episode. Alexandra prefers to live a truthful life, which is to live in the real world and not to be locked up in her own delusion. Other people can keep her in reality, taking their departure from the values, cares, interests and the like that shape Alexandra's life as a whole. Talking to her doctor and her proxies Alexandra may conclude that, e.g., wearing long earrings and staying up all night are decisive indications for 'passing the threshold' and for coercive intervention, in order to stop the disintegration. And she may decide to admit it to her Ulysses contract.

My job here is not to provide a theory for all cases of pathology and pre-commitment. My aim was more to say something about the direction we need to take. In order to balance whichever of two personal stories – a psychotic and a non-psychotic one – has to prevail, coherence is of deciding importance. The person's most coherent intentions, policies and plans – to use Bratman's words – are decisive. At what moment the threshold of incoherence precisely is reached is partly a matter of particularity, but the most coherent story is the authoritative one. In the case of florid psychosis, the psychotic story will lack internal coherence. And in the case of less severe psychosis the internal coherence of the story will be seriously flawed. In both cases, important parts of the non-psychotic story will disappear from sight in the psychotic story, whereas the non-psychotic narrative is keen on the integration of the psychotic story into the self-constituting narrative. Of great importance, moreover, is also that the psychotic story lacks external coherence. Up to a point it does not mesh with the stories of others and with the real world we all live in.

One last observation, before concluding. In medical practice Ulysses contracts are a very reasoned instrument. The conditions for completing Ulysses contracts are very strict. The author has to be diagnosed as a person with an illness that has periodic features (in our case as suffering from bipolar disorder). She composes her Ulysses contract being competent and with her disease in remission. She knows that in the future she may go through episodes of psychosis and she knows what that means having experienced such psychotic intermezzi in the past. Moreover, she has discovered that treatment proved to be helpful and that timely treatment can prevent a lot of the negative effects. She also knows how compulsory intervention worked out in her special case, and that her future refusal of compulsory treatment, in spite of being relatively competent, plausibly will be a sign of her altering state that could escalate into a florid psychosis. Together with her psychiatrist (and a proxy), she describes in the Ulysses contract the particular harms to be prevented by coercive treatment and/or hospitalisation, the types of behavior that signal the outbreak of a psychotic episode (sleeplessness, wasting money, bizarre thinking, euphoria, sexual promiscuity...), and the particular interventions that are necessary in case these signals manifest themselves, even if she then refuses these interventions (clinical commitment, psychoactive drugs, electro-convulsive therapy...). The agreed coercive measures are allowed for only a short time (for instance three weeks). And last but not least: Ulysses contracts sanctioning involuntary treatment and/or admission always have a limited period of validity (for instance a year) and are defeasible. After a psychotic relapse has passed away, there has to be room for evaluation, renegotiation, and revocation or refinement of the agreement. Let me now sum up.

I think it is safe to say that the use of Ulysses contracts can

be justified. The crux of my argument is that for our person suffering from bipolar disease executing a self-binding contract is contributing more to her personal autonomy (integrity) than escalation into florid psychosis will do. Alexandra has a picture of who she wants to be. She wholeheartedly identifies with this self-narrative. Cyclic psychosis is part of her narrative. Therefore, she is concerned with her future psychotic periods. Earlier outbreaks of psychosis were overcoming her more or less as a causal process, not as the result of deliberation and choice. They produced massive psychological change and a lot of material, personal and social losses. Alexandra arrived at the conclusion that her (sometimes tempting) psychotic behavior does not fit into her life, her values and her self-understanding. She feels alienated and takes her psychotic condition as a *Fremdkörper* that has to be banished. In her view, psychosis is a perverting influence on her mental powers that skews her perspective and causes distorted and impoverished personhood. It is threatening her deepest identity conferring concerns. It compromises and profoundly diminishes her person-life. So, this psychotic influence has to be minimised. Therefore, she makes the Ulysses contract part of her life-story and life-project. There is good reason to respect her prior consent to coercive treatment at a later time and to let what is expressed in the Ulysses contract take priority over her actual wishes as a patient.¹²⁸ The moral costs are low. It remains true that the actual autonomy of the hypomanic person will

¹²⁸ This is sometimes called 'self-paternalism' (Dresser, 1982: 789). The term 'self-paternalism' suggests that there are two distinct persons involved. I reject this assumption. At the very most, the psychiatrist is acting paternalistically in our case at hand.

be violated¹²⁹, that is, overruled by the precedent autonomy of the composer of the Ulysses contract. But we have consent-in-advance and one could reasonably expect that in most cases where the person refused treatment at an earlier point in time, this refusal will change later on. Retrospectively most patients will still consent to the compulsory interference. In such cases we respect autonomy by respecting former and subsequent consent.¹³⁰ In general, what lends authority to the former and subsequent exercise of autonomy is that the self-narrative the patient identifies with at that time is the more coherent one. It is the self-narrative that holds together the person's more central values and interests, and that is better kept in check by the stories of others and by the facts of the world.

It goes without saying that Ulysses contracts are a joint endeavor. To begin with, Alexandra needs a minimal supporting network that recognises the symptoms of psychosis in time. She needs a doctor (on the assumption that her doctor will be more

129 There is a lower limit to how much autonomy a psychotic person can exercise. In some cases the person is incompetent and the exercise of autonomy not an issue.

130 More or less related views are proposed by M. Quante: Ulysses contracts lay down the most authentic manifestation of the patient's will (2002: 287-295), R.L.P. Berghmans & G.A.M. Widdershoven: Ulysses contracts are expressions of joint narrative work and derive their moral significance from being an expression of the patient's fundamental values (2001), R. Spellecy: with a proper understanding of the future-directed planning embodied in Ulysses contracts, the enforcement of some Ulysses contracts may be justified (2003), and T. Van Willigenburg & P.J.J. Delaere: the proposal of an argument from authenticity based on 'one's deepest identity conferring concerns' (2005).

than a lever to achieve the patient's aims). And, ideally, the contract proposal also envisions a third party involvement, apart from the psychiatrist and the patient, to ensure that the patient's best interests and crucial values are served. One could even say that in enforcing a Ulysses contract other people are keeping the psychotic person in personhood. They are scaffolding and continuing her life story, and making the story of what happens a chapter of her life story. In the same way other people are helping her in preserving her personal autonomy, when she is in distress and in need of help. This may remind us of the second type of advance directives: the proxy directive. Why exactly cannot we argue, then, that the proxy directive is a better way to ensure that Ulysses contracts accurately reflect the 'value history' of the now less competent person? After all, the proxy can be very intimately connected with the patient – being, typically, her spouse, best friend or offspring. Why not give a proxy *carte blanche* to respond to all the subtleties of the situation? I think this road is heavy going for two reasons. First of all, and unfortunately, not everyone has someone in his circle of acquaintances who would make a proxy directive. Secondly, and more importantly, patients are very reluctant to declare that, during an early stage of psychosis, they are incompetent and the proxy is in charge. Experience teaches us that patients have a difficult time to part with 'the last word' about their future treatment by means of a Ulysses contract, even if they know that they possibly will benefit from it.

4.6 Enter the naysayers

The defence of advance directives is not uncontested. Rebecca S. Dresser and John A. Robertson are well-known opponents.¹³¹ This is how the sceptics try to clinch the case against Ulysses contracts. They argue that the trouble with this strategy leaves us with at least two problems: advance directives like Ulysses contracts are *theoretically tenuous* and they are *morally dangerous*. Let me unpack these worries.

Firstly, the threat of theoretical tenuousness. The current defence of Ulysses contracts presupposes a clear distinction between long-standing unconditional identity-conferring commitments (life plan, central character traits, integrity-value) and momentaneous defeasible commitments. Unfortunately, what is supposed to be a stable, whole life commitment, or ‘critical interest’ (to put it in a Dworkinian term), or ‘categorical project’ (to borrow B. Williams’ phrase) often proves to be rather contingent. Naturally, people form plans and sometimes life plans that are, as Bratman stated, to some extent robust and inert. Such plans should not come up for reconsideration arbitrarily; they have default stability. On the other hand, people can, and often do, change their plans. No plan is sacred. The same holds for central values and interests. These are supposed to be stable, but are they really such solid and overarching structures for life as narrative accounts like mine seem to suppose and want us to believe? Must not we say that people often revise their values and interests? Think of cases of conversion or the simple case of a young man who changes from a ‘Bruder Leichtfuss’ in a

¹³¹ Dresser even objects in principle to the use of advance directives (1982, 1994, and 2002-3).

conscientious family man after the birth of his first child. The value histories of persons are seldom uninterrupted. The overall direction and themes of someone's life narrative may change substantially. Our views of life are often brought to the bar of experience and may then alter. We can never be sure when this is going to happen, as we can learn from Ivan Ilyich who is tormented by the inauthenticity of his entire life – in Leo Tolstoy's story *The Death of Ivan Ilyich*. In one word, we cannot and should not exaggerate the degree to which we can anticipate outcomes and plan and value our futures and lives.

The second attack on advance directives – the moral danger problem – connects with the first and is more about treatment. Advance directives can be morally dangerous. Given the significant psychological discontinuity between the past and the present person, this discontinuity questions the supposed convergence between precedent interests and values and current interests and values.¹³² Being psychologically separated from the earlier person, the psychotic person may have changed his interests and values. He may then have new independent values and interests based on current needs. And these should take priority over the earlier ones and trump the required respect for the patient's precedent autonomy. The less psychological continuity, the more important actual values and interests become, the skeptic argues. Psychological discontinuity makes the values and interests of the later person independent from and more important than the values

¹³² I doubt the underlying ontological claim – put forward by R. Dresser among others – that psychosis can cause such a massive psychological change that the symptom-free person and the psychotic person are in fact different persons. But I do not want to dwell on these metaphysical matters. The outlook that I defend in my dissertation is a practical one.

and interests of the earlier person. Refusal of treatment by the patient, from this point of view, is not *per se* a proof of decreased competence. To suppose that is paternalistic: why not think that the patient's decision to revoke the Ulysses contract is as competent and rational as the decision that is written down in the Ulysses contract? A person in a hypomanic condition is not fully incompetent. The person has changed his mind. Why then respect precedent autonomy in these cases? It could even be dangerous to prioritise the so-called precedent autonomy. Advancing the 'coherent' values and interests the psychotic person no longer cares about could harm her. Moreover, coercive intervention is a debatable practice.¹³³ It leads to very emotional situations and may be experienced (by patients and professionals) as traumatic and as an attack on identity. Whether coercive intervention succeeds is unpredictable and partly a matter of luck. And very little is known with certainty about the therapeutic effects of coercive treatment.

Do these two problems leave us at a loss? Do they pose a severe threat to the authority of precedent autonomy? We must go carefully here. But the difficulties can be overcome in the case of Ulysses contracts. Or so I will argue. Begin with the charge of theoretical tenuousness: the assumed stability of coherent values and interests. To claim that an individual's central interests and values are always to be understood from a longitudinal or whole-lifetime perspective, rather than from a momentaneous perspective, is too rigid. Full-fledged personhood does not exclusively imply: to stick to certain values and interests. Full personhood also means: to be prepared to adapt to changing circumstances,

133 See G. Widdershoven & R. Berghmans (2007). 'Coercion and pressure in psychiatry: lessons from Ulysses' in *Journal of Medical Ethics* 33, 560-563.

to learn from them, and to change your mind. On the other hand, drastic changes are rather exceptional. Most people do not permanently change their hearts at a great pace. The stability of their life plans and characters makes it possible that they are predictable, cooperative and trustworthy. Yet there is an obvious reply to doubts about the stability of integrity values figuring in Ulysses contracts. Psychosis is a relatively short *entr'acte* of confusion, one could argue. Afterwards the person can pick up the pieces and continue her life plan. Eventual damage can be remedied. Occurring changes in the person's value system or lifeplan, if any, can be incorporated in a refined Ulysses contract.

What about the moral danger? The treatment problem is linked with a broadly accepted anti-paternalistic approach: for us Westerners the ethical principle 'respect for autonomy' takes priority over the ethical principle of beneficence.¹³⁴ All people

¹³⁴ The importance of personal autonomy is a disputed topic. It seems to be a Western preoccupation. There are other cultures where families and social networks are valued more highly than individual autonomy. In Western philosophy care ethicists, among others, propose to deemphasize autonomy in relation to the importance of trust and responsibility. But they do not dispense with self-determination altogether, so I feel safe in holding on to respect a patient's autonomy. Defenders of the principle of beneficence put forward that in the case of recurrent psychosis precedent autonomy is not the issue. The decline in competence of the patient is temporarily and the preferences and interests of the patient developing a psychosis are not what she truly wants, as in most cases becomes clear when she is no longer deranged. She would be appalled at what we have done, if we denied her Ulysses contract. Supposing that may amount to the claim that Ulysses contracts can be justified on 'soft paternalistic' grounds: the agent does not know what she is doing and 'truly' wants what others are imposing on her.

have a right of self-determination and a duty of respect for the autonomy of others (even if the exercise of autonomy eventually leads to a decrease in autonomy). Accepting the primacy of respect for autonomy as a starting point and taking autonomy as a diachronic matter, Ulysses contracts can be justified forcefully. First of all, enforcing a Ulysses contract we do respect the former autonomous consent of the patient. In addition, in many cases it will be possible to respect the patient's subsequent autonomous consent too. Agreed, failures cannot be excluded and compulsory treatment is a controversial *ultimum remedium*. But expectations are that a short violation of autonomy plus a long period in desirable circumstances after the coercive intervention on balance will produce the best possible self-governance.¹³⁵

In conclusion, in the last three sections I have argued that the narrative outlook on autonomy developed earlier in this dissertation can make sense of the notion of precedent autonomy and underpin a moral justification of the use of Ulysses contracts. The gist of it was that to be an autonomous person comes down to identification with one's self-narrative. We saw that autonomy thus conceived can resist a trio of skeptic arguments in the actual controversy on Ulysses contracts.

The apparent narrative disruption of identity was the first

¹³⁵ A related view is what J. K. Davis calls *diachronic respect for autonomy*: one should act so that there is the greatest overall consistency between what the patient wants and how his life goes over the longest total amount of time (2008: 97-9). Diachronic respect for autonomy is a clear-cut prospective moral justification for enforcing Ulysses contracts, in Davis' view, differing from retrospective justification, soft paternalism, and subsequent consent.

objection raised. Psychosis certainly is a disrupting experience, but the person can bridge the gap by making recurrent psychosis part of her self-narrative and appropriating it as an aspect of herself first; and by prospectively identifying with her future psychotic state. One may object that foreseeing and identifying with future conditions of illness is reading tea-leaves, but the case may be that the patient can show experience with psychosis. This experience may be a negative one, which the patient wants to avoid. It can also be a valuable experience the patient wants to profit from. In the remaining cases, psychotic disorganisation may be an ambivalent experience.

A more difficult second point was the supposed stability and persistence of a person's interests and values. (Contra Dworkin) I do not want to put a lot of work in a whole-lifetime 'obituarist' perspective on self-narratives. What will be salient features, good and bad, for a person's life is a matter of choice, within some limits at least. The criteria for evaluating what to count valuable and what trivial sometimes change within a human biography. Fortunately, in the Ulysses contract case the lifelong persistence of interests and values is of relatively minor importance. The period of validity of a Ulysses contract is limited and occurring changes in the person's value system or life plan, if any, can be incorporated in a revised Ulysses contract.

Treatment and/or admission under coercion of a relatively competent revolting person is obviously morally damaging, so went the last skeptic challenge. To that one may object with the principle of respect for autonomy. Respect for the precedent autonomous consent of the patient composing a Ulysses contract in the first place; and in most cases respect for the subsequent autonomous approval by the patient after the compulsory treatment.

This is, of course, a welcome result for my approach and, more

importantly, for the actual debate on advance directives. Nevertheless, there remain serious problems in the air, as another much discussed example, which leads to quandaries about self-narrative and how to respect the precedent autonomy of persons, may make clear: advance directives in cases of severe dementia of the Alzheimer's type. In the introduction of this chapter I foreshadowed that the dementia case will also prove to be a borderline case. Here is why. One of the things I want to make clear in the following sections is that, in order to be applicable, a view on narrative based self-governance has to meet certain conditions, one of these being that personal autonomy presupposes that owners of self-narratives remain autonomous as long as they are capable, or will be capable at some point in the future, of renewing or revoking their identification as a particular person. Unfortunately, severe Alzheimer's disease puts an end to the capacities to recognise and endorse a self-narrative as your own once and for all.

4.7 Advance directives and the severely demented

Severe dementia of the Alzheimer's type¹³⁶ strikes a lot of people as an illness about which there is only grim news. Decline into

¹³⁶ I concentrate here on only one stage of the gradual nature of dementia: pre dementia, mild dementia, severe dementia, and end stage dementia. For a detailed description of Alzheimer's disease, the most common form of dementia, see American Psychiatric Association (2000), *Diagnostic and Statistical Manual of Mental Disorders*, 4th rev. ed., Washington DC: APA, 154-7. Note that the gradual nature of dementia poses extra challenges to those who favor the use of advance directives.

‘infantilism’. No effective drug treatment. Advanced Alzheimer’s cannot be cured or reversed. To the best of our knowledge, it is a fatal disease. Symptoms include – among others – severe memory impairment and loss of control (apraxia and disturbance in executive functioning). It is humiliating and worse than cancer, H. Dupuis argues.¹³⁷ The main reason Dupuis gives us for giving preference to cancer is that dementia affects the human mind (‘our humanity’). To grow demented (literally) means to lose your mind. The victim is locked up in a brain that is ravaged by ‘plaques and tangles’ and shrinking; one third of the brain matter gets lost. The demented patient does no longer know who she is and where and in what time she is living. She is 78 and she thinks that she is 20 years old. She does not recognise the house where she lived for decades, happily married, and is desperately looking for the – no longer existing – house of her childhood. She does not recognise her life partner. Her emotions are whimsical. Her thoughts incoherent. She cannot discuss plans and decide what to do. She lives between the remnants of the past and without the promise of a future. In the house of doctor Aloïs Alzheimer the lights go out one by one. The person is disappearing slowly, becoming a residue of who she was, and in a sense dying a *mort sans corps* (a living death). Advanced Alzheimer’s disease seems to destroy the glue that binds a person over time and to bring an end to the existence of its victim as a person.

Most people, I suppose, do agree with H. Dupuis and view the prospect of becoming incompetent by advanced Alzheimer’s

137 H. Dupuis (2008). ‘Erger dan kanker? Een essay over demencie’ in I. de Beaufort a.o. (eds.), *De Kwestie. Praktijkboek ethiek voor de gezondheidszorg*, Den Haag: Lemma, 256-60.

disease with sadness and fear. It means the loss of so much in their present lives that gives them value and makes them their lives. There seems to be little value to a future life with severe dementia. For this reason some people decide to write an advance directive determining that life-sustaining treatment be withheld if they were to become incompetent and enter a potentially fatal medical condition, even in the event that they appear contented while profoundly demented. And again, respecting the precedent autonomy of such a patient is not without difficulties. Envision the following standard case as presented by Thomas Buller:

Vera Stuart is a 63-year-old nursing home patient who suffers from Alzheimer's dementia. Although incompetent, Vera seems quite content and could be described as 'pleasantly senile'. In particular, she appears to enjoy being read to (although she does not understand what is being read), sitting in the sunshine, and having her hair combed. Prior to the onset of dementia, Vera was a historian. An extremely knowledgeable person with a formidable memory, Vera loved the intellectual life that she led. The thought of becoming demented horrified her, and several years ago Vera wrote an advance directive requesting that she not receive life-sustaining treatment should she become incompetent. Vera has recently been diagnosed with pneumonia which is life-threatening if not treated immediately with antibiotics.¹³⁸

¹³⁸ Buller T. in Fulford K.W.M., D. L. Dickenson & T. H. Murray (eds.) (2002), 420. The case is based on a description in Allen E. Buchanan & Dan W. Brock (1990), *Deciding for Others: The Ethics of Surrogate Decision-Making*, New York: Cambridge University Press, 108.

Note that at this point Vera is quite contented, cannot remember the advance directive, and, when asked, expresses a preference to stay alive. The case confronts us with a choice between two options: we could decide to respect Vera's advance directive, or we could decide to override it. However, both of these decisions are problematic. The first sacrifices the life of a contented woman, though incompetent, expressing a desire to continue to live, on the basis of past interests and values that Vera might have forgotten. Is it right to allow a contented old woman to die when a simple intervention would save her? The second commits Vera to a type of life that she had expressly sought to avoid in her previously competent state. She conceived such a mentally compromised state as incompatible with how she sees herself. What is in question here is: do we have to regard advance directives as carrying significant authority in decision making for incompetent patients? Most leading literature in biomedical ethics does. I am not that sure. To think this through, let me recapitulate the sceptic objections – the theoretical tenuousness and the lack of moral authority – and try to work out the answers for the severe dementia case. But before moving on I would like to have some background points in place.

For a start, what could an advocate of narrative identity say about Alzheimer's disease? The dementia case may look more serious than the recurrent psychosis case. And in fact it is. Alzheimer's disease causes huge, irreparable brain damage. And the 'deconstruction' of the self-narrative also seems to run deeper. Broadly speaking, the incompetent patient suffering from dementia lacks the capacity for narrative self-constitution and *a fortiori* she is no longer capable of contributing to the coherence of the narrative or to make autonomous decisions. Author and life-story are falling apart. Vera suffering

from severe dementia is no longer living as a person with an unfolding self-narrative. She is a severely diminished version of that person. To the narrator, the victim of advanced dementia ceases to exist as a person. Wanting to live but having no sense of a past or future as one's own is insufficient for full-fledged personhood and personal persistence. What is left, to put it crudely, is a residue of a person. Some people may even think that Vera's demented life is similar to a dog's life. Vera lacks the complex forms of consciousness that constitute personhood, they say. She is definitively a sentient non-person. Needless to expect that such a no-more-person or post-person will be capable of self-determination and identification with her self-narrative. Autonomy is not even an option. Or so those thinkers argue. And they have a point. This is the basic story about Vera's condition. Nevertheless, my view is to proceed with caution here. Different from animals, people suffering from severe dementia once were persons and occasionally they may be able to partly 're-appear' for a moment. There is dim self-awareness. They may have relatively coherent likes and dislikes and some sense of self (they are not in a permanent vegetative state). There may be some ability to complete actions (sometimes surprisingly primed by what Oliver Sacks calls the narrative power of music). There may be some recognition of others left. A doctor reported to me of a severely demented former management consultant who did not recognise his family. When a family member visited him, he whispered conspiratorially: 'You're a member of my team, aren't you?'. And last but not least, there certainly are the possibilities of somatic communication and of sharing experiences with people suffering from severe dementia in the immediate environment. Agnieszka Jaworska (1999) helpfully noted that many demented patients are not yet so impaired that they have lost

all capacity to care and to value.¹³⁹ This may remind us of the fact that being a person involves more than rumination and cognition. Persons are emotive and caring creatures. In fact, we do not know what exactly is going on in people suffering from a demential syndrom, apart from the datum that the experience of dementia may differ considerably among individuals. I think all those things speak against a simple equation of the demential state and the animal state (or even the state of an infant; dementia is sometimes described, in a rough way, as a mirror-image of the human early psychological development). We do better to judge that people like Vera are still ‘regular folks’, even if they are drastically different. And most of us do.

Further background issues are about the use of advance directives, especially about some asymmetries between the dementia case and the recurrent psychosis case. Alzheimer’s disease causes a permanent, substantial decline in competence. So, enforcing an advance directive is different from enforcing a Ulysses contract. In the case of Ulysses contracts the agent’s earlier

139 Jaworska contends that the patient’s capacity to value or to care survives the ability to understand the narrative of one’s whole life. In her view, severely demented persons like Vera and Dworkin’s main case-study, Margo, may well still be valuers, whose efforts to *implement* their values are directly and severely hampered by Alzheimer’s disease. Only end-stage patients lack the capacity to value (1999: 134-5). The same holds for caring. Patients in moderate to severe stages of Alzheimer’s disease may well still be caring creatures. Late-stage patients are capable only of appetitive desires and sensual pleasures (2007b: 531). As a consequence, advance directives for demented persons who are still valuing and caring would be more equivalent to Ulysses contracts.

competence and the later competence may not be on the same level, and Ulysses contracts may enforce one competent expression of will over another (relatively) competent expression of will. In the described standard Alzheimer's case there is great disparity between the earlier competence and later incompetence. The advance directive is enforced when the patient permanently lacks the ability to understand that treatment issues have arisen and that she once made choices bearing on these issues. Consequently, Alzheimer patients often seek control over future treatment completing the two types of advance treatment directives: instruction and proxy directives. They choose someone they trust to make decisions and instruct that proxy to make certain sorts of choices on behalf of the formerly competent person.

On the face of it, there is a case for the use of advance directives in the event of dementia. Narrative identity is a kind of unity, wherein various experiences occurring at different times are unified as part of the same self-narrative, such that the narrative self can identify at any given time with all the experiences. Episodes of dementia can be included in an individual's self-narrative. Although dementia is disrupting narrative identity more seriously and permanently than psychosis, on closer inspection, what holds for cases of recurrent psychosis can also go for Alzheimer's disease. According to Vera, the life-stage with Alzheimer's is part of her self-narrative. It is a chapter of her life, not a codicil. She believed she could become demented, she felt that that individual would be she, and she was deeply caring for her future situation. While the disruption of narrative identity ('I wouldn't be myself') may invite a decision now to forgo life supports at a later time, narrative identity also bridges the gap between the normal and the demented person stage ('After all, it would be me'). Drawing an advance directive is trying to make the last stage of Vera's life a stage of *her* life. It is creating

narrative continuity and wholeness, one could claim. But, let me point at another asymmetry here. Vera's ability to identify robustly with her possible future state is only the prospective half of the identification story. Retrospectively, Vera in her diminished condition cannot identify or connect herself with the earlier person. The basic psychological capacities, which enable her to construct a single narrative out of various sets of past, present and future experiences and to identify with it are beyond retrieval. So, there is a great deal of disunity between Vera during post-competency and Vera signing her advance directive. Recurrent psychosis, by contrast, leaves room for retrospective identification after the psychotic surge. To avoid this problem one could put forward that all that matters for narrative identity is that at some earlier stage Vera was capable of projecting her self-narrative into a future stage. Yet, this is of course a watered-down version of narrative self-constitution.¹⁴⁰

One might even widen the gap and also worry about the prospective part of the identification. One could suggest that Vera's supposed 'robust' prospective identification was in fact more a 'fantasised' identification. How could Vera know, while competent, what life will or may be like as an individual suffering from dementia whose mental life is very different from what it used to be? Given that Alzheimer's disease will cause massive psychological changes, how could Vera be informed about a life with dementia? Definitely, she will know that there is a chance of suffering and loss of decorum. But people do a poor job of

140 DeGrazia distinguishes between two senses of narrative identity: *strong narrative identity* if a person has persisting narrative capacity and *weak narrative identity*, if a person projects her self-narrative into a future in which she no longer has narrative capacity (2005: 180).

predicting their preferences in situations they have never experienced. Again, the case of the psychotic person was different. The psychotic person already gathered *some* experience of what the future psychotic state would be like. The person anticipating dementia has no own experience at all with a demented state. So, how can she grasp her future situation as a dementia patient? This sounds sensible. Vera does not envision the sort of circumstances she will be in later. She does not know what it is like to be an inhabitant ‘of that country’. Alan Jacques, a physician who spent many years caring for dementia patients came to the conclusion that

(t)heir ideas of how they would react reflect more on their current state than on what their *actual* experience of dementia will be years later. Just as a child may swear that she will never marry, yet 20 years later is married and happily settled, so we look forward to old age and see its horrors as if we would experience them as we are now, not as we will be then (Dresser, 2002-3: 1836).

Let me now recap the sceptical criticisms – the theoretical tenuousness and the lack of moral authority – and see what can be said in favor of respect for advance directives in cases of severe dementia.

The charge of theoretical tenuousness was the problem of the assumed stability of a coherent self-narrative. The concern was that an individual’s central interests and values cannot always be interests and values assessed from one stable standpoint of one’s life as a whole; that we have to allow for changes of heart. Think of the famous Saul-St. Paul case. After his Damascus-experience Saul of Tarsus had to give up his whole-life-encompassing interests and values and make room for different whole-life-en-

compassing interests and values. In other words, it makes sense to speak of life stages depending upon an individual's changing system of interests and values. An individual's self-narrative may change along with his nature, character, values, or interests. Accepting that, it becomes less evident to suppose persistent whole-life interests and values surviving radical psychological changes. Remember that in the recurrent psychosis case there was a similar risk that the manic patient would change his mind when the advance directive was enforced. But after the psychotic intermezzo was over, it was possible to test the actual force of the patient's central interests or concerns and to repair the Ulysses contract, if desired. The case of dementia is different. The likely temporary nature of the anticipated condition is out of the question. There is no opportunity for evaluation and revision. In this play there is, so to speak, no Second Act. So, my tentative conclusion here is that the narrativist reply to the charge of theoretical tenuousness in the justification of advance directives in the severe dementia case is flawed, but can still be upheld. What to make of the concerns about the lack of moral authority and the moral dangers of advance directives in the case of severe dementia?

The moral authority of advance directives as a vehicle of precedent autonomy¹⁴¹ now faces significant problems. Two points are in order.¹⁴² The first problem is that strategies enforcing

141 Such a view is defended by Allen Buchanan, Dan Brock, Nancy Rhoden, Ronald Dworkin, Jeff McMahan and David DeGrazia, among others.

142 What follows is essentially Rebecca Dresser's sustained attack on advance directives (setting aside her metaphysical assumptions about personal identity). I concentrate on the ethical arguments, leaving out policy objections and empirical findings that expose problems with the advance

precommitment choices now may conflict with moral responsibilities to protect incompetent vulnerable patients from harm. The second problem, that will be postponed to section 4.9, is the mainstream patient-centered perspective that dominates the debate on advance directives in contemporary bioethics. Let us look at each point *in extenso*.

The first problem, an objection often raised by enemies of advance directives, is that, although a person's statements about future care can be relevant, such statements should be given less weight than other considerations affecting patient care. They see two reasons for overriding the advance directive. The first reason is that the directive gives an unadequate account of its makers' true preferences. It contains an uninformed choice (this is Alan Jacques' point) or fails to communicate the individual's actual treatment preferences. Therefore it lacks the moral weight of an autonomous choice. And here is Rebecca Dresser's rendition of the second reason (put in a Dworkinian register): 'When the capacity to appreciate critical interests is lost, experiential interests should take priority. Competent persons are free to elevate their critical interests above experiential interests. But after they lose decisional capacity, they have a different set of concerns. Experiential interests become central to their lives.

directive approach, although the practical problems revealed by empirical research are significant. The idea of precommitment turns out to be a seductive strategy, that is difficult to implement: 'First, advance directives are rarely completed. Second, most of the directives that are completed fail to convey meaningful information. Third, people making directives often have a poor understanding of what they are deciding. In particular, they may not envision how they could experience their decisions in a future incapacitated state' (Dresser: 2002-3: 1829).

Experiential interests should also be central to decisions about their life-sustaining treatment' (Dresser, 2002-3: 1840).¹⁴³ When patients become sufficiently and permanently incompetent, they lose their earlier preferences and interests. At first blush, it seems logical then to look at the patients' earlier wishes when they were still in touch with the beliefs and values that shaped their lives. But why should previous autonomous preferences and interests be extendable into an incompetent future? The interests and values of demented patients might differ from those they had as competent persons. We cannot respect the former preferences and interests any more than we can vote in an election after the polls have closed (in John K. Davis' words). Precedent autonomy is mythical in cases of severe dementia. To be authoritative, the autonomous decision has to continue to be reflective of the person's character and interests. In the case of Vera the advance directive no longer promotes her interests in her intellectual life. These old interests and values are

143 Dresser adds: Giving effect to a harmful directive, i.e. refusing an intervention that now offers clear benefit, would be inconsistent with the moral judgments underlying the *parens patriae* doctrine. The state has a traditional power and responsibility, under the doctrine of *parens patriae*, to care for and to protect the best interests of incompetent persons. 'Following such a directive could require insensitive and even physically coercive conduct by clinicians, which the law condemns in other contexts. To show respect for vulnerable persons, clinical practice and legal rules should protect patients from prior choices that would harm them in their current state' (Dresser 2002-3: 1840). A constellation of values supports this approach, according to Dresser: compassion and empathy for dependent persons, as well as the belief that people with mental disabilities can have lives of meaning and worth.

forgotten. Vera ceased to care about its objects. She most likely even lost the capacities to care about those objects. Her experiences are vastly altered. Respecting her precedent autonomy is a fiction. Severely demented people cannot be autonomous, no more than 'precedent' autonomous.¹⁴⁴

The worst of it, opponents of advance directives continue, is that to make an appeal to precedent autonomy is not only fictitious in cases of severe dementia, it is also morally troubling. Enforcing precedent interests will possibly harm the patient suffering from a demential syndrom. The incompetent later Vera is cut off from the earlier competent Vera (this being the result of severe brain damage). Such a massive psychological change makes it irresponsible to suppose convergence between the earlier interests and the current best interests. Respecting the interests of the autonomous former person is abandoning and consequently harming the best interests of the present incompetent person. The demented person has independent current best interests based on her current needs and preferences. In cases of severe dementia the best interests standard becomes more important than respect for autonomy. Vera's directive forbids a treatment that would effectively relieve her discomfort as an incompetent patient and that would allow her to continue a life that appears to have value to her.

¹⁴⁴ R. Dworkin believes that critical interests can survive incompetence but he provides no argument for this claim (1993: 222-37). One could object that many of us believe that some interests survive death. This is of course a highly elliptical claim. After death surviving interests rather become normative considerations with a certain legitimacy for the nearest and dearest instead of interests the deceased person has at that time. I shall return to this.

Do these strictures pose a challenge to the narrativist approach? We have two perspectives now: do we have to understand the best interests of our ‘pleasantly senile’ patient as a function of her current time-relative interests, or as a function of a whole-lifetime perspective, assessed from the standpoint of her life as a whole, in which case best interests are determined by the self-narrative and expressed through precedent autonomy? Vera completed her advance directive autonomously and, on the one hand, these autonomous decisions about her future have a role to play. On the other hand, why should the authority of her earlier decision extend to her present autonomy-lacking condition? Why should Vera comply with a preference she no longer affirms or even understands? The considerations for the two opposing ways of construing Vera’s best interests seem about equally reasonable and weighty, canceling each other out.¹⁴⁵ Vera’s best interests are ambiguous. Hence, showing that precedent autonomy has the force to trump the actual interests will require further moral argumentation. I think such a moral argument is exactly what a narrativist approach can afford.

On my narrativist account the further moral claim will be that identification with the whole of one’s life story, including the foreseeable parts of life when one will be severely demented and incapable of identifying, survives the loss of competence and dynarrativia caused by Alzheimer’s disease. Vera is autonomous

¹⁴⁵ This is also DeGrazia’s point (a man called Al standing in for Vera in his argument): ‘There is, I suggest, no uniquely correct answer to the question of whether continuing to live would be in Al’s best interests’ (194). But this indeterminacy concerns only best interests. In DeGrazia’s view precedent autonomy has independent *moral* importance that is decisive.

when she identifies with her coherent self-narrative, that is, with her most coherent identity conferring concerns. When Vera is severely demented, autonomy is not an option. She only has ‘experiential interests’ in those things that she finds pleasurable or exciting as experiences. So, there is no conflict within Vera’s autonomy. The best interest conflict can be resolved in favor of Vera’s earlier autonomous decisions. Autonomous advance treatment choices carry greater moral significance than other considerations, because they encompass authenticated interests and values expressing the person’s design for her life as a whole. Matters are, however, a bit more complicated. Therefore, let me engage in a short excursus on the distal binding powers of identification.

4.8 Autonomy at a distance: some second thoughts

Let us assume that a severely demented human cannot exercise autonomy and is not capable of narrative self-constitution. She is a post-person, so to speak. This assumption may seem to be in tension with the considerations that led me to propose to treat severely demented people as regular folks (and as more than human animals). But it is not. Although even in quite severe dementia there is often something left of the original person, the behavior is not distinctively the person’s own. None of it comes from any project of self-narration. Nor does it come about through identification. The lion’s share of what is left from the person survives in other people, and that is what gives these people good reason to treat a patient suffering from severe Alzheimer’s disease (also) as the person they knew.

Assuming that, I see a problem here. Precedent autonomous decisions are supposed to add value and interest to the person’s life independently of whether it is intrinsically enjoyable. But, in

my view, interests and values, in order to be *someone's* interests, have to be perceived and experienced as interests and values, and are, as it were, made complete by their own perception and experience. Interests and values can only be had by beings who 'can' be aware that they have them (what 'can' exactly means here will become clear in a moment). This is especially important for whole-life encompassing values and interests. These are not wholly impersonal or intrinsic. They are essentially relational. They can only exist as long as there are beings who have evaluative attitudes towards their own condition. And maybe we are confronted here with a fatal flaw in a narrativist defense of advance directives for the severely demented in terms of identity conferring values and interests. These narrativists are pressed to the claim that the identity conferring interests and values are authoritative although they are no longer in touch with, c.q. perceived and experienced by the demented person. Those identity conferring interests and values apparently are supposed to be respected, no matter what. After becoming demented, the patient is flying on an autopilot, so to speak. Vera becomes the hostage of her advance directive, even in cases of life and death, that is, in cases of decisions about her still having interests and concerns at all. Why should we accept that?¹⁴⁶

¹⁴⁶J. K. Davis develops a variant on this view on the authority of precedent autonomy (2007). In some cases, he argues, it is logically and nomologically possible that the person recovers from permanent incompetence. Interests survive if, assuming that it is nomologically possible, the holder would affirm the interest if he or she regained the capacity to do so. The holder does not lose concern, she loses the capacity to concern. An interest can survive a loss of capacity even though it cannot survive a loss of concern. I am not sure Davis' proposal is a moral defense, but, in any case, it suffers from a similar difficulty as the narrativist approach: concerns are provided with a star 'objectively valid', even if no longer subjectively endorsed.

Now, one could reply that the Ulysses contract case has the same implication. When Alexandra was unwell, her actual autonomy was trumped by the values and interests laid down earlier in her advance directive. Values and interests obviously can be kept longer than the moment of identification. Dementia is not a matter of conversion. Vera never revoked her values and interests. She is no longer capable of bearing them in mind. The pattern of values and interests she embraced at the time of the drawing up of the advance directive still counts, even if she is no longer able to take part in it. Granted, Vera's case is different from Alexandra's: the severely demented is beyond recovery. Vera can never reconsider her advance directive. But what is the big deal? Why is the difference between having an opportunity to reconsider the advance directive (in the psychosis case) and not having it (in the dementia case) so essential?

The answer to this objection is implied in my conception of autonomy. In my view, the distal binding powers of identification with a self-narrative are limited. I consider a person to be autonomous as long as he is capable, or will be capable at some point in the future, of renewing or revoking his identification as a particular person. I am adopting here a middle course between, on the one hand, thinkers like Luca Ferrero for whom autonomous commitments have no distal binding power on the agent's future conduct. On Ferrero's account, autonomous commitments determine the conduct of the agent only if renewed at the time of action based on the intrinsic merits of the case. The other extreme I want to avoid is the Dworkinian account that says that autonomous commitments contribute to the 'critical interests' or the 'integrity-value' of the person and his life, even during post-competency as a living (or dead) person. Both extremes are pushing autonomy too far.

I realise that such an avenue of reply, in particular the paral-

lel between post-persons and dead persons, seems to fly in the face of widely accepted ideas about posthumous interests. So, this calls for some explanation. There are at least three subjects in respect of which the concerns of living men extend beyond the boundaries of their lives (in such sort that also the law takes notice of them): wishes in how their corpse is to be disposed of; wishes regarding the posthumous disposition of their properties; and wishes concerning their good name and reputation. We regard it as a failure of respect for the (autonomy of the) person to disregard these wishes. My conception of autonomy mentioned above now seems to suggest that those wishes are unwarranted. And, I think, in a way they are, because the dead person exists only symbolically. Let me explain.

To begin with, there is a paradox: death is the moment at which the agent ceases to be, while events pursued after a person's death continue to relate to that agent as if he were alive. On close inspection, such events are incidents which cannot affect the person subjected to them. Note that I do not deal with religious aspects here. I am assuming that death does not mean transition or transformation, but – in all probability – cessation of the capacity to experience and annihilation of the person. On this assumption, it becomes difficult to see how the dead can be harmed or rewarded. How can one be harmed or rewarded when one does not experience the evil of harm or the rewarding good? How can the posthumous results of someone's actions change the value of those actions for *him*? Who is the subject of posthumous harm or reward? The dead does not exist as a person, let alone as the same person he was before death. How, then, can a dead person's interests be compelling?

David Sperling proposes the following solution (2008). First of all, to make after-life interests binding he conceptualises interests from an objective perspective: persons may have interests

even without their experiencing a specific outcome. For example, someone wants to be a famous performer. The idea is that the interests of this performer extend far beyond the period of his death. Let us suppose that Heath Ledger (1979-2008) wanted to be famous. Following Sperling's objectivist account of interests, the fact that Heath Ledger posthumously was nominated and won awards for his portrayal of the Joker in *The Dark Knight*, adds something to his interests as a performer. Sperling admits that the award-winning Ledger does no longer exist as a person. Therefore, he further claims that human existence can be conceptualised in different ways: one can not only exist as a person, but also as a 'Human Subject'. 'Human Subjects' do not exist in a physical or material way. Instead, their existence is logical (2008: 36). It is temporal and persistent. And 'Human Subjects' are holding four categories of interests: pre-birth interests, life interests, after-life interests, and far-lifelong interests (applying to one's life but also to episodes after one's death, for example the interest in being the object of affection or esteem of others). 'Human Subjects' hold these interests for the person; their interests are the latter's interests. Sperling believes that by dispensing with our habit of regarding and referring to ourselves exclusively as persons, the 'Human Subject'-model can protect human interests surviving death. He also thinks that the model finds some support in our use of language: when Ledger-fans say 'we love Heath Ledger', they presume in a way that Heath Ledger still exists to be loved, viz., that the 'Human Subject' bearing the interests of 'Heath Ledger' (the person who died in 2008) still exists and holds an interest in being loved and admired after death.

I am not convinced that, in order to solve the posthumous interest problem, we need to construct a new 'Human Subject' extending over the life of a person, whose interests they are. David Sperling's move strikes me as explaining *obscurum per obscurius*. I

think we can better face up to the fact that death makes us property of humanity. Take reputation as an example. During his lifetime Johann Sebastian Bach (1685-1750) was known as a skilled organist and a moderate composer. Nowadays, people baptise him the greatest composer ever, or the Harp of the Lord, or the Fifth Evangelist. I cannot think of Bach's present-day reputation as affecting literally the interests of the person Johann Sebastian Bach. This reputation management is ours. Here is another example. The British mathematician, code breaker and founder of computer science Alan Turing (1912-1954) was prosecuted in 1952 for having a homosexual relation, – homosexual acts were illegal in the United Kingdom at that time. The scandal ruined his career and reputation. He died in 1954, several weeks before his 42nd birthday, from an apparently self-administered cyanide poisoning. On 10 September 2009, following an Internet campaign, British Prime Minister Gordon Brown made an official public apology on behalf of the British government for the way in which Turing was treated after the second world war. In a petition, thousands of British citizens made posthumous excuses for the way Turing was treated after the war. Again, the actual excuses do not literally add something to the interests of the person Alan Turing. They can be no more than a kind of symbolic repair. The subscribers of the petition make clear that *they* now find it horrifying that Turing was treated so inhumanely and that *they* now think differently about gay men. At most, they are acting *as if* they compensate Turing for his misery.

This is not to say that it is completely useless to care for people's symbolic existence. My sceptical point is that whatever the use of the recognition of posthumous symbolic existence, it cannot be that something can be added to or subtracted from the interests of the deceased person. Most of us (say they) do respect last wishes and the reputation of the deceased. I think that for

surviving relatives to carry out the last wishes of a beloved person and to respect his memory can be a (ritual, symbolic) comforting way to come to terms with the dreadful loss of the beloved. Saying that they are doing it for the deceased or that it adds something to the interests of the deceased, in the final analysis, can be no more than a *façon de parler*. A last wish is not a trump. To be sure, this does not make the person's last wishes or reputation arbitrary or without moral importance.¹⁴⁷ It is morally problematic to break a promise or to drag someone's reputation through the mire. In the world most of us prefer to live in, we expect that, if people make a promise, they should keep it; and that what is said about (dead) people is not contrary to the truth. Let us leave the subject of posthumous interests here.

After this digression on the distal binding powers of identification, I am not sure that my narrativist conception of autonomy, for all I have argued so far, can settle the moral question why interests and values prior to the onset of dementia do trump preferences and interests expressed by the severely demented individual. Maybe other approaches will do better. Let us move on to two other defenses of the moral authority of advance directives in the dementia case, one offered by Jeff McMahan and one by communitarian thinkers.

McMahan's practical conclusion – at the end of his *The Ethics*

¹⁴⁷ In *The sources of normativity* C. Korsgaard offers us the case of a lawyer who discovers that a recently deceased client has changed his will. The client has eventually decided to leave all his money to his careless nephew and not – as originally intended – to medical research. From my perspective, I can agree that, since it would be far better to give the inheritance to medical research, the lawyer has a moral reason to violate the deceased client's will.

of *Killing. Problems at the Margins of Life* – is to allow Vera to die without treatment. He rejects the claim that the present best interests of Vera are determined by what would be best for her life as a whole. Vera's wish – in her state of severe dementia – to continue to live is only weakly affected by its lack of coherence with the rest of her life, because her phase of dementia is isolated from the rest of her life by dramatically weakened unity relations. McMahan finds it appropriate to regard Vera's 'unified life up to the onset of dementia as forming something as a whole, with the period of dementia being rather an excrescence dangling at the end – a period alien to and sufficiently distinct from the earlier unified life to have its own independent good' (2002: 501). According to McMahan, when the significant unity within a life is weak, the life, considered as a whole, has less significance, and the different periods that together constitute the life have greater importance, independently of their relation to the whole.

Now, roughly speaking, the life of demented people divides in two major parts: a dominant deeply unified part and a small dangling part outside the unified part. The later segment, considered *independently* of its relation to the whole, will be better if it is extended. What is best for Vera as she is in her demented state is to be treated and to continue to live. But the later segment, considered *dependently* of its relation to the whole, McMahan continues, will mar the whole. The longer Vera continues to exist as a demented patient, the more the integrity of the whole is degraded. McMahan believes that 'the meaning and value of a certain period within a life can be affected by the nature of the life as a whole of which it is a part. So whether the Demented Patient continues to live can retroactively affect the meaning and value of her life prior to the onset of dementia' (2002: 502). For this reason McMahan concludes that physicians ought to give priority to the earlier part of Vera's life and

allow her to die. The good of the earlier, higher self is more significant than the good of the shallow demented self. The good of the dominant, longest unified part of the life coincides with what would be best for the life as a whole. The wishes of that longest unity have to count as being closest to the life as a whole and determining the good of the whole. 'So the Demented Patient's present good ought to be sacrificed for the greater good of her earlier self, which is also the greater good of her life as a whole' (2002: 503). There is little prospect of anything good for Vera as a patient. Her demented life is describable as a series of 'discrete and mutually irrelevant episodes' and as a life of 'bovine contentment': 'Just as the prospect of this sort of life does not normally make a strong claim on us in the case of an animal, so it does not offer a strong reason for sustaining the Demented Patient's life' (2002: 503). McMahan closes with a further observation that reinforces, he thinks, his conclusion: the reasons people have for composing an advance directive are only partly self-interested and very often strongly motivated by a desire not to become an emotional and financial burden on those they care about.

I see reason to resist McMahan's proposal. Amongst its problems are these. (In line with Derek Parfit) McMahan abandons identity as what matters in our practical concerns, but he does not explain why 'life as a whole' should continue to matter. And for whom? There is a second, related puzzle. Accepting that the later Vera is psychologically discontinuous with her earlier healthy self, how could Vera's demented episode spoil the value of earlier stages? At the time it is heavily 'snowing in Vera's head', Vera's unified period of life is simply over. Her earlier wishes can no longer affect her life as a patient suffering from Alzheimer's disease. And vice versa, I would say. How, then, could her life with dementia retroactively undermine and mar the value of that

earlier unified episode?¹⁴⁸ How is such a reverse causality supposed to change the facts? And who is supposed to suffer the damage, and when exactly? Finally, I also have difficulties with McMahan's taxation of animal life, a point that I will flesh out in due course.

What this shows, I believe, is that McMahan does not add new refreshing arguments that can resolve the puzzle about the moral power of advance directives in the dementia case. Let me put a third proposal on the table. There also is a communitarian approach to the question of the moral authority of precedent autonomy. It says that part of a person's autonomy is co-constituted by others. Autonomy is something *ascribed* and *contested*; it is a socially situated status grounded in relations of inter-subjective recognition (J. Anderson, 2008). And indeed, to behave autonomously on a desert island will be a hard break. According to such a social account, in our case of demented personhood the most coherent values and interests of the person in question survive in the broader community, and proxies will try to reflect those life-spanning values and interests in the choices they make on the demented person's behalf by making treatment decisions that cohere with the direction of the patient's life story. Vera's advance directive was a public act of self-determination, which

¹⁴⁸ This is also David Shoemaker's objection (2009: 184). McMahan tries to argue from an external standpoint of temporal neutrality. But that is science fiction. More true to life is that a person in an early stage of dementia may look back on her abrupt life projects and sadly conclude: 'It was no more than a run-up to nothing.' Or that friends of a person suffering from dementia assess: 'Look at her now. She once was one of our most brilliant intellectuals. What a tragedy.'

authoritatively changed the normative situation of others with respect to her. She foresaw that her new life would be radically out of sync with how it used to be. And that there would be a gap between her competent and her incompetent self that could not be bridged from either side, but only by those who are witnesses to both sides, i.e., those who knew Vera in better times and experience her now. Only those witnesses are able to keep the person together, so to speak, and giving them a say in what should happen to Vera can be a way of respecting her precedent autonomy. This is sometimes called ‘relational autonomy’ (C. Mackenzie & N. Stoljar, 2000) or ‘assisted autonomy’ (T. Nys, unpublished manuscript).¹⁴⁹ Should we accept such a dialogical view on autonomy?

It is to the credit of such a view that Vera’s precedent autonomy (and advance directive) can best be articulated and interpreted by significant others who are able to take a view over her whole life. Nevertheless, Vera’s case demonstrates a considerably derived sense of respect for autonomy because it is far removed from Vera’s subjective endorsement. In dealing with dementia her capacities for identification evaporated. So, the proposed inter-subjective approach strips autonomy of the very *self*-governance that lies at the heart of the concept of autonomy. It leaves out the constitutive role of the subject. An implication of such an inter-subjective view could be that autonomy even survives the death of the subject. That sounds a little odd. In the communitarian proposal interests or identity conferring concerns are in

149 ‘Assisted’ autonomy is a misleading qualification in the severe Alzheimer’s case, where the autonomy in fact is delegated. It is not like visiting a doctor and being assisted by a dependable friend. Assistance suggests help on demand.

danger not to survive as interests or concerns in the usual literal sense of the word – in which such interests and concerns must have a subject.

Like the McMahan proposal, the communitarian approach does not seem to be a passable route to underpin the moral authority of precedent autonomy better than my narrative approach.¹⁵⁰ Back to the main line.

Where are we so far? This was the matter we need to resolve here: can my narrative view withstand the sceptical objections against the authority of advance directives in decision making for incompetent dementia patients? Or do we have to conclude that we better give up our initial intuition that advance directives are authoritative in the case of Alzheimer, and demote advance directives from their position in the hierarchy of decision-making standards? The first objection raised was the charge of theoretical tenuousness. Different from the Ulysses contract case, in the event of Alzheimer's disease it is less obvious that the interests and values discounted in the whole-life narrative of the person have not changed as the years went by. Alzheimer patients are irretrievably incompetent and not able to reassess their earlier views at some point in the future. Yet, I did not regard the charge of theoretical tenuousness as a decisive draw-back.

¹⁵⁰ I do not discuss one totally different route of thinking here : the proposal of A. Jaworska and others to treat marginal agents like Alzheimer's patients as rudimentary competent and autonomous persons. The idea is that just as long as basic capacities are in place like the ability to value (Jaworska 1999), or the basic emotional capacity to care (Jaworska 2007), or a basic ability to control one's experiences (S.V. Shrifin in Burley 2004:195-217), we must treat Alzheimer's patients with 'full moral respect'. Such a claim renders advance directives in the end useless.

The second sceptical objection was about the lack of moral authority and the moral dangers of advance directives in the case of severe dementia, and it came in two parts. The second (postponed) point was the mainstream patient-centered perspective that dominates the debate on advance directives in bioethics orthodoxy, and I will come to it in a minute. The first point was the problem under discussion in the foregoing paragraphs and about the conflict between strategies enforcing precommitment in the name of autonomy and the responsibilities to protect vulnerable patients from harm. I found out that attempts to solve the problem in terms of best interests were canceling each other out at first. Further independent moral arguments for the authority of advance directives were needed. And my own narrativist account seemed to deliver such an argument: precedent autonomous identification with the whole of one's life story, including the foreseeable parts of life when one will be severely demented and incapable to identify, survives the incompetence and dynarrativia caused by Alzheimer's disease. Autonomous advance treatment choices carry greater moral significance than other moral considerations like beneficence or mercifulness, because they encompass authenticated interests and values expressing the person's design for her life as a whole. However, closer inspection of the distal binding powers of identification with a self-narrative raised doubts. Things are valuable or important to or for some sensate creature endorsing it. Things are not important to the universe, so to speak. I can only consider a person to be autonomous as long as he is capable, or will be capable at some point in the future, of renewing or revoking his identification as a particular person. For post-persons suffering from Alzheimer's disease, this is not an option. So, I face up to a problem here that makes the prominence of anti-paternalistic respect for precedent autonomy highly debatable in the case of dementia.

Suppose I am wrong about the distal binding powers of identification. How, then, to decide the question what to do with Vera's advance directive? Let me follow, for the sake of argument, David DeGrazia's conscientious advice. This is how he is reasoning. Let us assume that Vera was well informed about dementia and the implications of her advance directive. There is no reason to think that, had she known even more about dementia, she would have decided differently. She made her decision in the understanding that she might be contented in her state of dementia. Moreover, she had no change of heart before losing competence. Then, '(e)ven if it were in (her) current best interests to continue to live, (her) precedent autonomy makes (her) life as a whole a very significant consideration (though not unambiguously a best-interests consideration) – and the best option from (her) whole-lifetime, autonomy-infused perspective, I submit, is to be allowed to die in circumstances like the present ones' (195). So DeGrazia goes for a very careful: Yes, we have to make Vera comfortable, but when pneumonia – 'an old man's friend' – comes, withhold antibiotics and allow her to die. Suppose, we are persuaded. Then, there is one last sceptical counter-argument left: the problem of the mainstream patient-centered perspective that dominates the debate on advance directives in contemporary bioethics.

It is common decency to respect the autonomy of persons who decide to perform actions that are problematic because they reflect shallow values, or because they harm the chooser's own interests or dignity; actions that are irrational, imprudent and not *self*-beneficial (smoking for example). That is fine with me. But such cases are different from our dementia example where respecting precedent autonomy implies: to inflict harm to some *other* person for reasons this non-autonomous person once found self-beneficial but no longer understands. This is an issue I will settle in the next section. Can this argument be the straw that

breaks the camel's back in the moral debate on the authority of advanced directives for the severely demented? I think it can.

4.9 The patient's keeper

Most philosophical support for advanced directives is built on an analysis of the competent individual.¹⁵¹ Its major premise is that no other person bears the burden of the decision as much as the patient. Unfortunately, advocates of advance directives tend to take a partial view of the treatment situation emphasising individual freedom of choice and discounting other important moral dilemmas. DeGrazia's approach is a good example. DeGrazia frankly admits that he takes a patient-centered perspective and that staying close to the mainstream patient-centered ethic he morally oversimplifies the decision-making (201). Unfortunately, pointing at such a biasedness does not take away the very serious flaw in such a position. Overlooked is the role of the others who have to enforce the advance directive that seeks to impose on care providers and proxies a responsibility to act according

¹⁵¹ Legal arguments also place high value on *individual* control over future care – in terms of rights of bodily integrity, personal liberty, privacy and self-determination. Principles of equality and respect for all individuals and the value of human dignity require that demented individuals be given the same freedom of choice as competent individuals. Advance directives 'give a voice to people who can no longer speak for themselves'. In absence of an advance directive two other standards are playing a role: the substituted judgment standard and the best interests standard. The *substituted judgment standard* instructs to make the treatment decision the patient would make if miraculously lucid for an interval. Because

to the patient's individual instructions. This responsibility may conflict with another responsibility, to wit the duty to protect incompetent patients from harm (the principle of non-maleficence). The preferences expressed in the advance directive should be respected only if they fit within the accepted bounds of treatment. The duty to protect vulnerable patients from harm is such a convention. A directive should not bind physicians to practice 'passive euthanasia' (though it is not always called that), that is, the withholding or withdrawal of treatment that would otherwise prolong the life of a non-suffering patient. That would clearly harm incompetent patients. And it could also harm the physician. This is a delicate point. I am fully aware that not every physician withholding treatment and thereby referring to an advance directive runs into qualms about his way of behaving. I am also aware of the fact that in some cases of dementia medical complications like pneumonia are welcomed as a relief by the people most concerned and the physician. So, one has to take my claim here very literally: that in Vera-like cases a directive should not *bind* physicians to practice 'passive euthanasia'. This leaves open the possibility to act according to circumstances and to decide to withdraw or to withhold treatment.

[*continuation of note 15*] there is no advance directive, the patient's general values, beliefs and attitudes toward medical care form the basis for treatment decision. At a certain point this standard begins to resemble the best interests standard: when evidence of the patient's past values and preferences is ambiguous or non-existent. *The best interests standard* seeks the decision most consistent with the patient's 'real' interests. It rests on community norms and embodies an objective, 'reasonable person' approach to treatment, in contrast to the subjective, more individualised approach embodied in advance directives and the substituted judgment. It endorses a standard of best care for the person.

Yet, advance directives may go one step further and require callous and even physically compulsory conduct by clinicians. This is at least what the Belgian professor in ethics Etienne Vermeersch is expecting from his friend, a physician, who promised to help him, when it turns out that Vermeersch in a supposed future state of severe dementia does no longer recognise his family for about two weeks. At that point, his medical friend will practice active euthanasia, following his earlier promise. Professor Vermeersch was asked in an interview what his friend would do in the event that the demented professor resists. He answered: ‘This friend will carry it out. Euthanising a suffering animal one will continue, when the animal resists.’¹⁵² Following Vermeersch, the suffering animal is a comparable case. His own future suffering, as he sees it, is that he (possibly) will do things in a demented state he normally would be ashamed of. Vermeersch mentions his demented mother, always a prudish woman, who could not care less about her being half-naked at the end of her life.¹⁵³ According to Vermeersch, to be satisfied with a wasting disease is not an option. It is essential for him ‘to remain himself’, that is, to remain gifted with reason. Beyond reason, life has no longer meaning for him. Vermeersch refuses ‘the contented life of a well-fed pig’.

152 P. de Koning, ‘Zonder de rede is het leven zinloos’ in *NRC Handelsblad* 29-03-2008. Note that human suffering cannot simply be equated with being humanly disgraceful.

153 What Vermeersch is fearing are things he can now imagine but would most likely not then experience; it is even unclear whether there will be a self to whom those ‘terrible’ experiences will happen. In this sense, the fears of Vermeersch are unrealistic. If having a sense of self is a prerequisite for suffering, the potential for suffering decreases as the illness advances.

I can imagine that such a planning may impress as an intensely personal act in the face of decay and shame, but it leaves out the impact of the killing act on others. The same holds for the Vera case of ‘passive euthanasia’: the impact of providing maintenance and support but not life-saving medical treatment, and so allowing Vera to die, when potentially fatal conditions arise. This is not about temporarily coercive measures, like in the case of preventing the harm of manic disturbance. It is about letting die or killing a human being in the name of an advance directive that in most cases is not recognised by the patient.¹⁵⁴ Letting die or killing a non-suffering human being is not just another medical intervention that serves the interests of the patient. It turns out to be emotionally, morally and professionally burdening and disrupting for a doctor. Physicians are not executioners. For them, it is a matter of sleepless nights. So, my point is that we have to show mercy not only for the patient but also for the vulnerability of the caregivers and health professionals. We cannot expect doctors to expose themselves to such psychological and moral risks. The moral costs of letting die another person who feels happy and comfortable and does not suffer¹⁵⁵, and who

¹⁵⁴ What is essentially lacking in severe dementia is, of course, the possibility of a shared understanding. Cees Hertogh (2009) argues that, next to autonomy and beneficence, reciprocity is a *conditio sine qua non* for euthanasia. Such an ultimate form of help is possible only in the context of a relation of trust and mutual understanding up until the moment the assistance is given. An advance directive can never replace this.

¹⁵⁵ See e.g. De Boer, Hertogh, Droë, et al (2007). ‘Suffering from dementia – the patient’s perspective: a review of the literature’ in *International Psychogeriatrics* 19, 1021–39. The authors conclude that despite the fact that dementia is accompanied by a lot of negative feelings, the literature on

can easily ('without heroic measures') be saved are too high for care providers. The moral costs are even so high, in my view, that they put the command envisioned by advance directives beyond reach. I think this scruple is ultimately fatal for the whole enterprise of advance directives in cases of severe dementia, and that the conclusion we have to draw is that in the case of Alzheimer's disease advance directives are an ineffective method of orchestrating one's future care. The recorded wishes of the person whose death is at issue cannot be determinative. Future care instructions can only and should only play a partial role in decisions concerning patients suffering from Alzheimer's disease. The idea that advance directives can do most of the hard work implies that those at the bedside can avoid hard choices (and conflicts) about which intervention their beloved should receive. But such difficult quality-of-life and end-of-life choices can rarely be avoided. Given the unforeseeable (!) complexities of illness and end-of-life decisions, only proxies and care providers are in a position to decide what is the best thing to do for the demented patient. This does not mean that future patients with Alzheimer's disease are totally helpless to prevent what they fear, however. Accepting their future vulnerability instead leads to two strategies that stand as alternatives to advance directives. But before I turn to that, I want to say a few words about the supposed worthlessness of a life with severe dementia.

the perspective of the patient offers no solid support to the widespread assumption that dementia is necessarily a state of dreadful suffering. Moreover, there is reason to assume that even patients with severe dementia, as a consequence of psychological coping strategies and progressive neurologically based unawareness, do not experience the suffering they formerly feared at the time they drew up their advance directive (Hertogh, 2009).

Envision once again the case of professor Etienne Vermeersch. He rejects a life of 'bovine contentment', because such a life is beyond reason and meaningless. Why continue a life that is meaningless? What would such a wretched epilogue of dementia add to his life? This is, I grant, a precarious point. Taking a personal stance, it turns out that episodes of severe dementia add very little to a person's life. At the time dementia rules, the person is gone. For reasons I have already touched on, we have to accept that a human being suffering from severe dementia is no longer a person. What is left from the person is settled on the body (dispositions, habits, signs of interest, responsiveness to others and resulting behavior) and scattered over the social environment. In fact, the demented patient is kept in personhood by others and surviving in those others. He can no longer put himself in a time structure. How then could he even want to continue his life? His behavior is largely automatised. His life is broken into separate events just like a sentence falling apart in single words. He is totally absorbed by the now here. His life only consists of a momentaneous bundle of experiences; often pleasant, positive experiences, as far as we know (disagreement here with Vermeersch's parallel of the *suffering* animal). Is it a meaningful life?

Taking Susan Wolf's proposal for what it is to live a meaningful life – a meaningful life is one that is actively and at least somewhat successfully engaged in a project (or projects) of positive value (2007: 65) –, we have to admit that the life of a demented human being is pointless. Yet, from the perspective of the severely demented, just as from the perspective of a cat, meaning is not an issue. 'Integrity value' is not an issue. 'Experiential interests' are the only things that matter to the severely demented. But here, I think, Vermeersch and I reach a fork in the road. The consideration I want to emphasize here, contra Vermeersch, is

that the meaninglessness of the life of a human being does not give us any positive reason to end it.¹⁵⁶

The equation with animal life is instructive now. We should not think that demented people like animals are less important to themselves as individuals because to us they seem to be less person or less rational than we are – for that is nothing from their point of view. Each of them has subjectivity that is still uniquely her own, however similar its contents are to that of others. Animals and Alzheimer’s patients are important to themselves by their own standards. What is important to them is to lead keen lives of appetite and sensation, to eat and drink, to be integrated with their bodies and to inhabit them joyfully, and to live in the moment. To be sure, the death of a person is tragic in a way different from the death of an animal or a demented post-person. But it is important to remember that persons, human beings and non-human animals who lose their lives all lose anything they have and that is important to them.¹⁵⁷ Persons are not more important than demented human beings or other animals. Such cosmic rankings make no sense. There is something imponderable about the comparison. They are just very different.

To put the point in a more clear-thinking, Korsgaardian way: as rational animals human beings have rational and natural

¹⁵⁶ Admittedly, it is true that Vermeersch does not put the point as a general claim; his remark is meant for people who claim the right to see their own lives as pointless.

¹⁵⁷ In a contra-intuitive, goading essay (*Why I have no future*, 2009) Galen Strawson defends a No Ownership of the Future View. It says that from a first person perspective the only life people can lose is the one they are living now in the present. One cannot lose the future, for no one can take from one what one does not have.

concerns. They place value on both parts of their nature. Many of their natural concerns – the desire to avoid pain is an obvious example – spring from their animal nature, not from their rational nature. So, as rational animals human beings have to treat their animal part with respect and compassion for its own sake. And the same holds for the treatment of the animal life of others, including the life of dementia patients. Like all animals, and more than animals, demented people are our fellow creatures, fellow travellers, and fellow sufferers. They make a strong moral claim on us. We expect civilised people not to kill their pet just like that or because the lives of pets are ‘without meaning’.¹⁵⁸

Not enough is said here, but even if one remains unconvinced and still wants to eliminate a non-suffering demented human being, one has to come up with positive reasons for doing that. Here is a much disputed candidate: we can admit that demented people have an interest in leading happy lives as long as they are living, but they do not have an interest in the prolongation of their happy lives. Only persons can have an interest in continuing their lives, whereas demented post-persons are not able to entertain reasons to wish extra lifetime. Life ‘is no longer calling them’. They have become indifferent to continuation. This is certainly true, although I do not think that this gives us a

¹⁵⁸ I am fully aware that human moral attitudes and practices with respect to the other animals exhibit a great instability. People often seem to accept without question that animals do not make a strong moral claim on us. This is, I think, a widespread moral shortcoming. Here I take a radical, revisionist stance on the heels of Christine Korsgaard. Recall that only two hundred years ago white Americans had no scruples about mutilating or killing black people. They were wrong. Nowadays humans are not justified in casually destroying animal life.

good reason to end such a life untimely. To end it untimely is to deprive the patient from a valuable life (this line of reasoning is often called the *deprivation argument*). The life of a demented human being may be of no additional *personal* value, it certainly adds *some* value. It may have instrumental value insofar as it contributes to the interests or well-being of others. And it has value to the extent that a demented life is often good and worth living for the human being who is experiencing it.

I can now turn, as promised, to the two strategies for future Alzheimer's patients that stand as alternatives for drawing up an advance directive. One possibility for diagnosed Alzheimer's patients is to forget about autonomy in their future incompetent state. Remarks about the future state of incompetence constitute a limited attempt to gain power over events that are often unmanageable. In Dresser's arresting phrase: 'In anticipating our possible futures as incompetent patients, we should accept that we cannot exercise firm control over our later medical care – that we will, for the most part, be dependent on others to protect us from harm and indignities' (2002-3: 1845). Accepting that, we may write in advance a durable power of attorney and name a proxy as surrogate, believing that we have a trustworthy family and a supportive circle of friends and preferring to endure the outcome if they err in predicting our preferences or if they choose to ignore our preferences, rather than to remove from them the opportunity and the burden of making the choices. Let our caregivers use their own judgment to reach their decisions on what will be in our best interests.¹⁵⁹ We can draw up an advance

¹⁵⁹ Note that no cut-and-dried solutions are offered in this scenario. In some cases it will be difficult to find a devoted proxy. In most cases proxy and care providers will have to make difficult decisions, 'passive euthanasia' (though it is not always called that) being one of them.

directive *ad libitum*, but this is not a necessary means. It can be a bit of encouragement. To appoint a future surrogate decision-maker is a matter of trust. Such a person is familiar with our life story. We value her judgment and the way she perceives and respects us. And we even deeply care about how *she* will feel about our future situation.

In this proposal, to give the last say to a surrogate decision-maker is our last act of autonomy-at-its-best. When dementia is setting in, we are at the mercy of others. Others decide. Most doctors are not willing to let die a happily senile patient with pneumonia *exclusively* acting on information received from an advance directive. They will do what is possible to maintain and supply medical and supportive treatment until the end of the patient's natural life. And most proxies will try to keep their happily senile beloved as long as possible in their midst. They will hold their beloved in personhood, maintaining her identity for her (as they did all along, by the way). H. Dupuis about her demented life partner in a recent interview¹⁶⁰: 'I care for my husband. He is like a child now. My motive is the supposed reciprocity. He would have done the same for me. It remains to

160 In *Filosofie Magazine* 2007 (8). Another well-known example is what Iris Murdoch's husband tried to realise for his wife and described in John Bayley (1998), *Iris, A memoir of Iris Murdoch*. London: Gerald Duckworth & Co. Ltd. Faced with the choice to decide how and when their demented loved one should die, proxies will be keen on signs that the patient himself initiates the train of events that leads to death. Advance directives cannot replace this action of the patient. Sometimes demented patients, who always rebelled against the loss of autonomy and control, die very soon and unexpectedly, as if they still 'choose' the time of their own death and put their signature on their lives. But, in general, nature imposes severe limits on our ability to control how and when we die.

be seen if it is possible to promote the well-being of other people. Anyhow, we must try not to harm them. Good care? That is paying as much attention as possible to the person one was and is. To pull one's individuality its weight.'

It is my belief that this strategy is the one we have to prefer. As a player affirms in Bert Keizer's brain-opera *Alzheimer*: 'Dementia can be seen as a disaster area that the patient has to traverse together with his loved ones. That there is someone at your side who knows the way through the jungle is invaluable, even without any means to alter that jungle somehow or other.'¹⁶¹ But my hunch is that the first strategy will not satisfy the advocates of respect for autonomy who may want to skip the whole sorrowful experience of severe dementia.

Even then, there is a second alternative for advance directives in the case of Alzheimer's disease. One can use his 'right to die' and commit suicide before becoming incompetent. This is very much a last resort position, but for people who want to avoid at all costs the dependence and humiliation of dementia and remain 'in charge', it offers a way out. No difficult decisions to be taken by doctors and friends. After Alzheimer's disease has been diagnosed or in the comparatively early stages of the disease, when one is still able to form and act on plan and is likely to have access to means of suicide, one can autonomously decide: 'I don't want to go through this illness', and take measures to die just before the illness sets in.¹⁶² Nature has provided us with the

¹⁶¹ Taken from Keizer B. (2003-5). *Alzheimer. Tragikomische Hersenopera in Twee Acten*, Amsterdam, 26 (unpublished libretto).

¹⁶² People considering suicide, just like people considering an advance directive, may misjudge the impact of the experience of severe dementia. Notice that stoic suicide is inconsistent with my preferred strategy in that it denies the deprivation argument.

key of the backdoor, as Montaigne sets it in *Coustume de l'Isle de Cea*. Escapist suicide puts us in hot water, of course. There are many questions. Why is the killing of another person morally insuperable and self-termination a viable option? Is suicide a failure to treasure the gift of life? Is suicide a bad or mad thing to do? It is my belief that it is better to think it is not a bad thing. More precisely: not in all cases. The rest of my argument will be an attempt at clarifying this claim.

4.10 Pre-emptive suicide as the triumph of autonomy?

In the standard psychiatric view, persons who threaten or commit suicide are seen as irrational or mentally ill. Thomas Szasz, an influential (anti)psychiatrist strongly holds the belief that there is no such connection between suicide and mental illness: '(T)here is neither philosophical nor empirical support for viewing suicide as different, in principle, from other acts such as getting married or divorced, working on the Sabbath, eating shrimp, or smoking tobacco. These and countless other things people do are the result of personal decisions...' ¹⁶³ Although it is

¹⁶³ Quoted in Jamison 2000: 254. Kay Redfield Jamison herself emphasises in her book on suicide that suicide in most cases is a rare act of a troubled or depressed person. According to Szasz, suicide is a matter of personal choice, not of sin, crime or illness. Self-killing is not even medical matter and physicians have no place in it. Death control is a moral matter and should be under the control of the individual. And if the day comes when you need to be in control of how your life ends, and you cannot find a doctor who will help, Peter Singer advises *Final Exit* by Derek Humphrey as the book to get. In Dutch Boudewijn Chabot's guide *Uitweg* (2010) is an option.

impossible to disregard the psychopathological causes of a lot of suicidal behavior, I think Szasz has a point. A person can commit a so-called balance-suicide on rational grounds. Historical examples of self-inflicted death are legion. There is 'altruistic' self-sacrifice among the elderly and the sick in nomadic societies to save commonly held resources of food. In pre-Christian times suicide was a matter of honor, of action taken to avoid falling into the hands of a military enemy, of atonement for wrongdoing, or a way to uphold a religious or philosophical principle. In ancient Greece the Stoics and the Epicureans defended the individual's right to choose the means and the time of his death. And at the end of his essay *On Suicide* David Hume wrote, convincingly, to my mind, that 'prudence and courage should engage us to rid ourselves at once of existence when it becomes a burden'. I do not want to idealise self-delivery here 'as the privilege of the human' (J. Améry). But my intuitive belief is that there is a lot that speaks for balance-suicide within the context of severe dementia. What I have in mind with that is what Boudewijn Chabot named 'self-euthanasia'.¹⁶⁴

Yet, suicide is morally suspect. In a diary note Ludwig Wittgenstein called suicide, without further argument, 'the elementary sin'.¹⁶⁵ Stronger arguments against escapist suicide can be

164 'Self-euthanasia' in Chabot's definition is ending one's own life by taking a combination of drugs or by knowingly stopping taking food and drinks. He lists four constraints (2010: 29): it is not a solitary act; there is no pain and no mutilation; it is effective; and the agent is causing his own death (contrary to assisted suicide).

165 'If suicide is allowed then everything is allowed. If anything is not allowed then suicide is not allowed. This throws a light on the nature of ethics, for suicide is, so to speak, the elementary sin' in *Notebooks 1914-16* (eds. G.H. von Wright & G.E.M. Anscombe), Oxford: Blackwell, 1961, 91c.

found in the deontological camp. Kant, for example, argued that cutting our life short is immoral when committed for the purpose of escaping from unhappiness. Suicide is opposed to the principle of humanity as an end in itself, to put it in a Kantian vernacular, because it throws away and degrades humanity in oneself. The issue is pressed with insight by David Velleman in his 1999 *A Right of Self-Termination?*¹⁶⁶ Velleman offers a clear and subtle reconstruction of Kant's argument by presenting a particular interpretation of Kant's formula of humanity, which says that a human person has a value that makes him an end in himself. Let us take a closer look at Velleman's account.

Defenders of escapist suicide claim that they have a moral right to live and die in the light of their own convictions about why their life is valuable and where its value lies. Following Velleman, this principle is based on two presumptions. The first is that a person has the right to make his own life shorter in order to make it better on the whole for him. The second is that we have to defer to the person's judgment on the subject of his own good (607). Velleman can accept the second presumption. The person living a life is the best judge of the value that its continuation would afford him; we have to defer to his considered judgment even when we have reason to regard his judgment about whether it would be good for him to go on living is mistaken. But Velleman rejects the first presumption, the idea that a life sometimes is no longer worth living and that a person has the right to end

166 And reconsidered in his 2008 'Beyond Price', *Ethics* 118, 191-212. In this paper Velleman supplements respect for persons with 'love' for persons. Note that the scope of the right to suicide is broader than the right to freedom from unwanted medical treatment or a right to 'passive euthanasia', as in the Vera case.

his life solely on the grounds of the benefits he will thereby obtain or the harms he will avoid. Velleman therefore invites us to think about the concept of interest-relative value.

What is good for a person? The best answer, to Velleman's knowledge, is Stephen Darwall's proposal: what's good for a person is what's rational to want *for his sake*, i.e., the person's good is what would be rational to want out of concern for that person. By the same token, Darwall argues that also the person himself is rationally obliged to care about his good insofar as he cares about the person whose good it is – that is, himself: 'What's good for a purpose is worth caring about only out of concern for the purpose, and hence only insofar as the purpose is worth caring about. Similarly, what's good for a person is worth caring about only out of concern for the person, and hence only insofar as he is worth caring about. A person's good has only hypothetical or conditional value, which depends on the value of the person himself (...). (E)very person already matters for his own sake, because of embodying an interest-independent value' (6II). Kant's term for this interest-independent value is 'dignity'. Kant attributes dignity to all persons in virtue of their rational nature. And what morality requires of us, in Kant's opinion, is that we respect the dignity (that is, the rational nature) of persons. This value is not just the affair of the person in question. The dignity of a person is something larger. It is a value that a person possesses by virtue of being 'one of us'. Dignity is a 'self-existent' value, according to Kant. It exists already. It is to be valued as it is. And it commands respect. Dignity makes a claim on us. It is a value that we must 'live up to'. Period. The value of a person is unconditional. The Kantian notion of dignity is that of valuing a person in a special way, not to a special degree – indeed in a way that does not admit of degrees. It cannot be balanced against the person's interests. Dignity is not something

the person can accept or decline. It is a value *in* him – not *for* him – which he can only violate or respect. With that clear for the moment, let us consider how dignity functions within Velleman's argument against escapist suicide.

According to Velleman, committing suicide on the grounds that life is not worth living is trading one's person in exchange for relief from harm. Such an act entails a practical irrationality analogous to that of sacrificing an end for the sake of the means to it. This is immoral in Velleman's view. His objection to suicide is not that it destroys something of value but that it denigrates the person's value by trading the person for interest-relative goods, namely escaping from decay. It weighs the person in the balance against interest-relative goods, thereby denigrating the person's value as a person – and, by implication, the value of all persons. But Kantianism leaves open the possibility that a person's dignity may justify suicide in other contexts, if suicide constitutes an appropriate expression of respect for the dignity of the person. Or so it seems to Velleman, controversially. But how so?

The idea is that a person's chosen death can be morally acceptable if the person can no longer live with dignity. Dignity can require the destruction of what is losing it, if the loss is irretrievable, like in cases of unbearable pain or terminal illness. Unbearable pain can tyrannise persons in a fashion that undermines their rational agency. Unbearable pain 'is somehow to fall apart in the face of it, to disintegrate as a person' (618). The person is falling apart synchronically and diachronically. Alzheimer's disease is another example of suffering that threatens personhood. Dementia is the distressing perception of actual or threatened disintegration in the self. It does not necessarily accompany pain, but 'it does necessarily touch one's dignity – the value that one has by virtue of being a person and that is jeopardized when

one falls apart' (627). This is not to say that Velleman defends a fundamental right to choose between life and death. There might be a moral justification for suicide in some cases, but it does not rest on a right of self-(de)termination. The moral justification is that it will spare the patient from degradation. Velleman also points at the fact that to permit oneself the choice of suicide for the sake of one's autonomy leads to a paradox. Suicide may be an exercise of autonomy and a violation of autonomy at the same time. It is 'to treat a single exercise of that autonomy as worth the sacrifice of one's autonomous self. It is to treat oneself like a stick of dynamite' (625).

I think this is an interesting result. Velleman's approach fits very well into a narrative theory of self-constitution and autonomy. And I take it that Velleman's reconstruction of the Kantian argument gives us all we need to defend escapist suicide in the case of threatening dementia, even along Kantian lines. I for one find Velleman's modification still unsatisfying for at least two reasons. My first worry is about the claim that human nature as a rational, autonomous person is of 'incomparable worth' and 'above all price' and admits no equivalent. This seems an unfounded assertion to me. It sounds like a secular version of religious talk about the sanctity of human life. Agreed, rationality is special but why should this faculty be inviolable and decisive? And what does it exactly mean that the value of personhood is 'self-existent' other than that the speaker is putting forward an a priori argument? To postulate the value of personhood as 'an end in itself' is to immunize it. It is a conversation stopper. To be sure, respect for persons and a ban on killing (other) persons are both wise and defensible moral guidelines. But we better take personhood (that is, our rational nature) as rooted in an *ideal* conception of human life, that is, as a construct we might equate with worthiness of respect. It is based on the ideal of the kind

of person we want to be and the kind of personal life we want to lead. Ideally it can be valued for its own sake. But it does not hold unconditionally and above all else. No more than that rejecting the 'self-existent' or noumenal, in Kant's rendering, status of personhood will reduce all other persons' interest-relative goods to things that no longer *really* matter to them.¹⁶⁷

In line with David Velleman, Thomas Hill, another neo-Kantian, argues that in three sorts of cases to end one's life before its natural end implies not necessarily an insult to the dignity of life (1991: 90-1): Firstly, the case of suicide when human life is no longer possible ('I prefer not to continue to live as a subhuman being'). Secondly, suicide to end gross irremediable pain (i.e., the Velleman cases). And thirdly, the case of suicide based on self-regarding moral beliefs ('I couldn't live with myself if I did that, e.g. becoming a slave, a prostitute, or a cannibal, because it violates the deepest values concerning how one should live'). Hill stresses that suicide in these three cases may be dignity-of-life-affirming and that one should treat dignity as a special value, independent of but not always overriding considerations of the person's good.¹⁶⁸ Moreover, according to Hill, in all three cases cutting short the time one could live as a rational, autonomous agent can be seen as a manifestation of autonomy, an ultimate

¹⁶⁷ When we bring up for discussion the sacrosanct, intrinsic dignity of personhood based on the reason-trafficking faculties of persons, it becomes less clear why persons exclusively deserve 'full moral standing' and sentient human beings or animals (maybe) less.

¹⁶⁸ This leaves us with the worry what to make of the values that constitute a person's integrity. Any value that a person chooses to live by? Only values that contribute to the person's good? How to define a person's good? These are not matters we need to resolve here.

decision of the author of a biography to conclude with a powerful expression of values one chose to live by. This leads me to my second and for present purposes more important dissatisfaction with Velleman's view.

David Velleman is reluctant to emphasise that asserting a 'right of self-termination' really is a matter of self-determination. I want to stress this relation. Taking my lead from Margaret Pabst Battin, I consider the right of suicide as a fundamental moral right, that is, a right 'to do certain sorts of things just because doing those things tends to be constitutive of human dignity' (1994: 280). And, like Battin and Hill, I consider pre-emptive suicide in the face of deteriorative mental conditions such as Alzheimer's disease as an autonomous act, based on an ideal self-conception, beneath which one is not willing to slip: 'I am what I have been, but cannot be anymore'. The suicide is not acting 'in the twilight of the patient's autonomy'. Looking out onto severe dementia, surcease or cessation suicide is anticipated and planned. The decision to suicide is a composed, rational, 'self-dignifying' affirmation of one's self-narrative. The person strongly identifies with his true self-narrative and decides that dignity requires that he has to end his life. This is, of course, fatal exercise of autonomy and elimination of a valuer. But it is not to adopt an attitude that denies or denigrates the valuer in question. It takes the person as an end in itself. Recall that deciding to give Alzheimer's disease free play also implies the destruction of autonomy and personhood. Pre-emptive suicide seems to me an ultimate act of self-determination or what Germans call *Freitod*: a profoundly individual voluntary choice to bring about one's own death. All in all, pre-emptive suicide – like all suicide – remains an act of aggression towards oneself and towards others. There is always an element of 'dynamite' in it that is self-

destructing and has damaging effects on life, loved-ones, care providers, and society as a whole. But pre-emptive suicide that is constituting human dignity and is anticipated and planned in a resigned and purposeful way will considerably temper the violence and collateral damage.¹⁶⁹ Let me end my discussion of pre-emptive self-delivery with an intriguing thought-experiment offered by Margaret Pabst Battin that provides insight in what the desiderata of autonomous suicide in the face of Alzheimer's disease are:

Suppose there were a simple medical device, based on triple technology of the time-release capsule, the subdermal contraceptive implant, and a painless, quick-acting euthanasia drug developed in the Netherlands. The device is called a "delayed-onset euthanatic implant". Anybody newly diagnosed with Alzheimer's disease (or any other long-term, degenerative, ultimately fatal disease), while still lucid and competent, can request one. Positioned painlessly and invisibly just below the surface of the skin inside the upper arm, the implant is designed to release its lethal drug instantaneously after a designated delay – say, two or three years, or whatever the patient requesting the implant stipulates. The implant can

169 There is of course some tension between the two strategies proposed in the last two sections. The first holds that we have to upgrade and respect our animal nature, especially the animal life of other people. The second allows the autonomous self-destruction of our animal-cum-personal nature as an exception. I am accepting some asymmetry here. I clearly consider pre-emptive suicide as a last resort strategy and submission as the preferable and most valid one. But I take it that both strategies are morally justifiable.

be easily removed and there are full guarantees, rigorously observed in practice, that a patient can have it removed at any time, for any reason, with no delay. If it is removed there are no after-effects. But if the implant is not removed, it will release the euthanaticum after the designated delay – without further warning, without pain or discomfort, and without requiring activation of any sort. It will just go off, and like an instantly fatal but pain free heart attack, that will be the end (1994: 160-1).

I am not sure Battin's device will be doomed to remain science fiction. Awaiting coming events, it will be clear that her thought-experiment with 'reversible euthanasia' succeeds in doing three important things: firstly, it permits the autonomously choosing early Alzheimer's patient who wishes to avoid late stage deterioration to bring about her own death. Secondly, it avoids having others kill the patient after she is no longer competent. And thirdly, it does not force the patient to waste 'good' life during the early phases of the disease, where there are still lucid intervals. Above all, the fictional implant permits direct autonomous choice. Advance directives, in contrast, displace actual decision-making onto persons other than the patient. They may masquerade as instruments of the patient's choice, but cannot fully serve this purpose.

4.11 Taking stock

My aim in this fourth chapter was to engage my accounts of personhood and autonomy with specific practical issues, in particular Ulysses contracts in mental health care. I have tried to demonstrate that my narrative approach to personhood and

autonomy has some advantages. It can considerably clarify the qualms about practical identity problems of episodically disordered psychiatric patients; and my account offers an improvement on our understanding of those patients' autonomy. I have argued for the following conclusions:

1. My narrative self-constitution view illuminates reported experiences of psychiatrists and patients with recurrent psychosis. Here, then, are the main results: in cases of severe psychosis the patient is temporarily a 'no-more-person'. In less severe cases the personal changes are experienced as deviant. Yet, there seem to be no practical reasons for supposing that the psychotic patient is someone else. The patient's psychotic experiences in some way or other take part in the life-spanning narrative self-constituting process. Nevertheless, frequent far-reaching changes of a person may raise the question of how a person's story can be designated as the authoritative one. After all, the self-narrative of the non-psychotic, the pre-psychotic and the psychotic will diverge considerably. The answer I propose is that coherence is of decisive importance here. The most coherent story is the authoritative one. In (pre-)psychotic stories the internal coherence often is seriously flawed and important parts of the non-psychotic life story sometimes disappear from sight, whereas the non-psychotic self-constituting narrative is sometimes more consistent but always keen on the integration of the (pre-)psychotic story into the self-constituting narrative. Of overriding importance, however, is that the (pre-)psychotic story as a rule lacks sense of reality ('external' coherence). To some extent it does not match with the facts of the world and with the way significant others interpret the (pre-)psychotic self-narrative.

2. My narrative outlook on autonomy can make sense of the notion of precedent autonomy and underpin a moral justification of the use of Ulysses contracts. Although psychosis causes narrative disruption of identity, persons are able to bridge the gap by making recurrent psychoses part of their self-narrative, and by prospectively identifying with their future psychotic state (patients composing a Ulysses contract experienced psychosis before). A core idea in my narrativist contribution to the discussion of advance directives is the person's identification with a set of whole-life, persistent and coherent identity conferring concerns (values and interests, as I also called them). Such entities can survive severe psychosis. In my view, autonomy serves those persistent and coherent values and interests and we best respect the person's autonomy overall when we give priority to the autonomy exercised by a person when most competent to exercise it. This person can decide what set of policies do best to protect the person's autonomy overall. I have my doubts, however, about the stability of what Ronald Dworkin coined the 'integrity-value' of a person's life. Contra Dworkin I take into consideration that the criteria for evaluating what to count valuable and what trivial often change within a human biography. The good news is that in the Ulysses contract case the lifelong persistence of values and interests is not of overriding importance. Psychosis is an episodic phenomenon and misjudgments can be rectified afterwards.

The upshot for the moral assessment of Ulysses contracts is this: accepting the primacy of respect for autonomy as a starting point and taking autonomy as a diachronic matter, the enforcement of Ulysses contracts can be morally justified. First of all, enforcing a Ulysses contract we do respect the precedent autonomous consent of the patient. In addition, in many cases it will be possible to respect the patient's subsequent autonomous consent too. Nevertheless, the moral costs are substantial. Advancing

identity conferring concerns the person no longer cares about and treating a relatively competent protesting person under coercion may be harmful as it may be experienced by patients and professionals as traumatic and as an attack on their identity. Moreover, whether coercive therapy succeeds is unpredictable and partly a matter of luck. But expectations are that a short violation of autonomy plus a long period in desirable circumstances after the compulsory intervention will, on balance, produce optimal self-governance.

3. Although the idea of precommitment may impress as a seductive strategy, my thought is that the use of advance directives is limited. This becomes clear, when we consider advance directives in cases of severe dementia of the Alzheimer's type. Again, my narrative account of personhood and personal autonomy turns out to be helpful to illuminate the highly problematic authority of advance directives in cases of severe dementia. Comparable with the recurrent psychosis case, the times at which there is no narrator can be bound up in a self-constituting narrative, and so those stages can be attributed to the person who narrates and identifies in advance. But the case of dementia is different, of course. Firstly, the narrative disruption of identity now is permanent and runs deeper. The person disappears. Secondly, identification with the incompetent state of dementia is more difficult than identification with a period of recurrent psychosis, because the person cannot show own experience with dementia. Thirdly, allowing for possible changes of heart is not an option.

The moral justification of advance directives as a vehicle of precedent autonomy also faces new problems in the Alzheimer's case. I considered advance directives requesting that the severely demented patients not receive life-sustaining treatment should they become life-threatening ill. The first problem is that, in

my view, at this stage of dementia the distal binding powers of precedent autonomy have expired. I take a person to be exercising autonomy as long as he is capable, or will be capable at some point in the future of identifying with his self-narrative. As a consequence, an advance directive enforcing precommitment choices now competes with the moral responsibilities to protect actual vulnerable incompetent patients from harm. My narrativist account of autonomy cannot deliver a decisive moral justification for trumping the actual interests of the severely demented post-person with the interests of the person prior to the onset of dementia. This gave me reason to look elsewhere for an extra argument. To no avail.

The second problem is that the moral debate on advance directives is dominated by the patient-centered perspective. This strikes me as a fatal flaw. We are in the business of life and death now, not of precarious compulsory treatment. Others have to enforce the directive, that is, to withhold treatment and to let the patient die. The moral costs of killing another person who feels happy and comfortable and does not suffer and who can easily be saved, are high and tip the scale. I argued that we have to show mercy not only for the incompetent patient but also for the vulnerability of the care providers. We cannot expect doctors to expose themselves to such moral risks. The moral costs are even so high, in my view, that they put the command envisioned by advance directives beyond reach. In the case of Alzheimer's disease advance directives are an ineffective method of orchestrating one's future care. They can be no more than a bit of support.

4. This leaves persons diagnosed with Alzheimer's disease and confronted with a catalog of horrors with two strategies that stand as alternatives to advance directives. One possibility is that they accept that autonomy in their future incompetent state of

dementia is not an issue and that they will be dependent on others to protect them from harm and indignities. Surely, they can draw up an advance directive, but this is not a necessary means. When severe dementia is setting in, others decide. The second possibility is that they commit pre-emptive suicide after Alzheimer's disease has been diagnosed. This is not claiming that it is better to be dead than to be demented. But for some people autonomy, including autonomy over future treatment, is a treasured value that cannot be abandoned and is more valuable than life itself. They want to remain in control and avoid what they see as a troubling, lost existence. For those patients *Romana mors* in the face of Alzheimer is a morally justifiable autonomous act.

SUMMARY

This dissertation is a two-stage project. In the second part (chapter four) I apply an account of personhood and autonomy to specific practical issues in mental health care. In the first part (the chapters one, two and three) I try to provide that account in terms of agency. Like other animals human beings act. But there is something distinctive about the nature of human action. In virtue of a certain type of self-consciousness humans are conscious of the grounds of their actions. They act for reasons and can make choices. This distinctive feature of human agency, reason, brings with it two capacities. First, human agents can take control of their actions and have to work out how to do that. They have a capacity for normative *self-government*. Second, whenever human agents take control over their actions, they are making themselves the authors of their actions. They are faced with the task of deciding who to be. As rational agents, human beings also have a capacity for normative *self-constitution*, i.e., for constructing norm-governed or practical forms of identity. It is this practical form of identity that I am after in my dissertation.

In the first chapter I take the ongoing debate on human agency as an entry point, especially the self-constitution view of Christine Korsgaard, who offers a picture of reflective agency and goes on to derive a very elaborate account of personal identity in terms of deliberative action. Humans are born for action, she states, and agents are moving through a 'space of reasons'. They have the capacity to deliberate, that is, to choose for acting on reasons about what they would do if their reasons were to survive a process of reflective scrutiny. When an agent reflectively does choose which reasons he is to act on, and, when he acts on these reasons, he does act in ways that are constitutive of his self,

viz., of his identity. So, on the Korsgaardian account, identity is practical identity and is bound up with practical reasoning. The *substantia nigra* of self-constitution, according to Korsgaard, is the normative force of practical reason. It is the normative consistency of reasons that is constitutive for personhood and for the unity of personhood. I agree, but I do however put some doubts on the table. I think Korsgaard overestimates the unifying power of the human rational abilities. Her idea that human beings by nature have rational capacities that ought to achieve overall rational unity within themselves is unwarranted. All that can follow from Korsgaard's account is that human beings *can* achieve overall rational unity. The rational capacities that human nature supplies do not by themselves dictate to what extent human beings have to reach rational unity and how to do it. It is true that Korsgaard, taking her lead from Aristotle, adds extra resources to support her unification argument. In spite of this, the unifying role of practical reason, c.q. the Korsgaardian principle of autonomy (universalisability), turns out to be overblown. We have to move to a less principled, instrumental role for reasons that are constitutive for personhood and the unification of personhood. Believing we need unifying reasons in a strongly Korsgaardian, formal sense to be persons is a mistake: we can be persons with far less. By implication, we have to weaken the unification requirement. Persons are constituted by a self-reweaving network of reasons, that is, a self-correcting enterprise which can put any claim in jeopardy, but not all at once. We can think here of Neurath's famous boat. All personal reasons are experimental, fallible, and replaceable like the planks of Neurath's ship. This is not to say that persons are not committed to unity. For the most part persons do not want to be like wrecked ships. But to lead the more or less unified lives of more or less unified persons, all these persons need is a Humean reading of instrumental reasoning

instead of Korsgaardian ‘principled reasoning’. And in addition to means-end reasoning they can adopt a Humean conception of ‘reflective rationality’, some steady and general point of view in order to correct their idiosyncrasies. In a word, I take it that Korsgaard’s self-constitution view is leaving us unsatisfied and only can be a reasonable start for a viable theory of persons in terms of agency. And my proposal is to develop the *reason view* of personhood into a *narrative view* of the person. We better take reasons as an incipient story about us, that is, as a ‘rationale’ that makes our actions intelligible as coherent developments in a life story. I have at least three arguments for such a move. First, stories in all probability are the easiest and most natural way to put in order our experiences and our knowledge. They seem to be a natural place to look for an adequate understanding of personhood. Second, a coherent self-constituting story will do more justice to the multiplexity, the discontinuities, and the changes of the person than unifying reasons can. In a narrative view on persons the unification constraint can be replaced by a demand for coherence, and reason can play a less normatively potent and more instrumental role. A third argument for developing the reason view on persons into a narrative view is that a narrative self-constitution view can be linked up better with the human body. It can repair a rather underexposed aspect of Korsgaard’s self-constitution view: the lack of an elaborated theory on the embodiment of persons. Reasons alone cannot be supposed to do the self-constituting and unifying work. Also emotions, desires and the human body do their share.

In chapter two I try to develop a narrative self-constitution view on persons following Marya Schechtman. A person’s identity now is to be found in narrative self-understanding, in the capacity to keep a particular narrative going. Such a self-narrative organises and integrates disparate elements of human experience

and action synchronically and diachronically into a more-or-less coherent storied whole and in this way constitutes a person. Internal coherence, being articulated to some extent, being in step with the facts of the world, and consensus with interpretations of others are important standards for self-narratives. A Schechtmanian narrative self-constituting view avoids the Korsgaardian overestimation of the formal, unifying powers of reason and settles for the more substantive, vernacular criterion of coherence of self-narrative. My narrative self-constituting view also takes self-constituting narratives as embodied and embedded structures (Schechtman under-describes the embodiment of self-narratives). Persons are incarnated and situated beings. To unpack the idea of the embodiment of self-narratives, I take an empirical line. Antonio Damasio demonstrates that self-constituting narratives have a bodily, neuro-physiological basis. They build, *inter alia*, on the presence of a core-self. Moreover, Damasio's picture clears the way for emotive constraints of self-narrative that keep self-narrative on track. Damasio makes plausible that autobiographical narratives are intertwined with emotion and feeling. Persons are partly and importantly constituted by how they feel, by how it feels to be them. Schechtman for her part is aware of the significance of affect and the role empathic access plays in the making of self-narratives. Antonio Damasio also is fully conscious – as is Marya Schechtman – of the embeddedness of persons. A full explanation of self-constitution needs to go not only beyond the brain to the rest of the body, but also into the world. Persons are permanently interacting with their environments. And how one feels about or thinks about one's self is partly constituted by the (social) world. As they are socialised into human culture, human beings are taught to operate with a background conception of themselves as continuing individuals leading the lives of persons. Moreover, self-narratives

are confined by observational and interpretive facts. In a deep way they are co-determined by the social world and constructed collaboratively. The individual person does not have the final say, so to speak.

I think a Schechtmanian narrative self-constitution view can do serious work in our understanding of the constitution of persons. All the same, Schechtman's view has to be slightly adjusted. Schechtman pays too little attention to the embodiment of narratives. And the articulation constraint has to be nuanced. There is a suggestion in it that persons have to talk themselves into existence. This is also one of Galen Strawson's worries. Yet, the articulation requirement is not that, in order to become a person, one must have a fully worked-out and explicit account of why everything in one's life is as it is, but rather that a person must recognise a certain kind of explanatory obligation, and be able to meet it for the most part. What is more, although persons couch their self-descriptions in language, they are not exhaustively captured in what they can say about themselves. Self-conception and its operations are largely implicit and automatic. Self-narratives are bio-psychological constructs that are only partly expressible in verbal form.

As mentioned earlier, one of my concerns is with applying my view on practical identity to specific practical issues. I want to concentrate on the use of advance directives in mental health care. Considerable attention will be drawn to Ulysses contracts. These are written by an author who is competent to consent to medical treatment, stating how he wishes to be treated under coercion in the event of becoming episodically a psychiatric patient (think of manic or schizophrenic psychosis). Also a related biomedical procedure will be considered: advance directives in the case of Alzheimer's disease. Both procedures have in common that they bear upon pre-commitment. Both are cases of

an 'earlier self' being in command of a 'later self', that is, of an earlier self exercising authority and imposing his or her will on a later self. This raises the question of (respect for) autonomy. So, before I embark on the project of applying my view to psychiatric practice, I have to make room for the problems of self-government of the person. Developing a theory of autonomy that is related very closely to my theory of narrative identity is the task of the third chapter of my dissertation.

Schechtman's narrative self-constitution view says very little about autonomy but certainly can take on board some account of it. Building on the work of Harry Frankfurt and Michael Bratman on autonomy, this is the solution I propose: a person acts in a *self-governing* manner and is acting as a governing *self*, if he encompasses his autobiographical narrative, i.e., if he identifies with (parts of) his self-narrative. Identification is fundamental to autonomous agency and selfhood. In my view, human self-consciousness is characterised by a hierarchy of attitudes about attitudes. By hierarchy I mean that there is mental room for assessment and endorsement (or rejection) of a self-constituting narrative. Human beings are able to take a stance on their own attitudes. This can take place in different ways. Ongoing self-narratives incorporate strands of reasons, memories, intentions, emotions, carings and the like. These 'Lockean ties', to put the point in Bratman's terms, resist hard and fast distinctions. Yet, we can discriminate between at least three aspects of identification. Identification partly is an act of *reflective endorsement* of reasons for action; this was Korsgaard's line of approach. As we can learn from Harry Frankfurt, identification in part is also an act of *acceptance of carings*. Frankfurt has a keen interest in the conative side of autonomy and the limits of reasoned deliberate action. Autonomous persons are both rider and horse. Often a rider, if he is not to be parted from his horse, is obliged to guide it

where it wants to go. Finally, identification is also partially an act of *emotional approval*. Emotions function as relevance-prompts. There is no general intellectual strategy for relevance search. We may notice a lot of what is relevant in our self-narrative by its 'feel'. And we may approve of that feel. Autonomy has to do with felt intensity, with experiencing oneself (more) intensely.

A further point I argue for is that the autonomy of an agent is increased by an action's springing from an attitude that coheres with certain other attitudes of the person's psychic economy. In Schechtman's narrative self-constitution view internal coherence (intelligibility) and 'external' coherence (with the facts of the world and the interpretations of others) were already crucial for successful self-constituting narratives. Yet, not only the person's ongoing self-narration has to be relatively coherent. Also the ongoing autonomous endorsement of it has to be relatively stable and reflecting the person's central, coherent values. Michael Bratman emphasises this point, and I am of his mind. Self-governing narrative-based agency involves norms of practical rationality: pressures of consistency, of means-end coherence, and of reasonable stability over time. Again, *coherence* turns out to be an important standard. Self-governing persons normally try to go beyond an incoherent stew of intentions, emotions and considerations. In the case of incoherent intentions, policies and plans, or in cases of an incoherent self-narrative (that is, plans writ large) the person does no longer know where he stands and how to govern his life as a person. Now, one could ask: Why make a choice for the norms of practical rationality listed above? I think the relevant reason for the coherence constraint is our reason to govern our own lives. My account of personal autonomy, like Bratman's, basically is a prudent one. Coherence contributes to the effectiveness of our agency and to the richness of our (social) lives. We are more likely to pursue our ends and not trip

over ourselves. In the picture of personal autonomy I offer, the person's choices and own values are central and exclude nothing. The person's autobiographical narrative and his endorsement of it are what count. I take a person to be exercising autonomy as long as he is capable, or will be capable at some point in the future, to endorse or reject his own choices and values, in one word, to identify with his self-narrative.

In the fourth chapter, I engage my narrative accounts of personhood and autonomy with pre-commitment cases in psychiatry, in particular Ulysses contracts and advance directives of persons suffering from Alzheimer's disease. The chapter demonstrates that a narrative theory on personhood and personal autonomy in terms of agency can make important contributions in the field of psychiatry. Here are the main results:

(i) My narrative self-constitution view can illuminate reported experiences of psychiatrists and patients with recurrent psychosis. In cases of severe psychosis the patient temporarily is a 'no-more-person'. In less severe cases the personal changes are experienced as deviant. Yet, there seem to be no practical reasons for supposing that the psychotic patient is someone else. The patient's psychotic experiences in some way or other take part in the life-spanning narrative self-constituting process. Nevertheless, frequent far-reaching changes of a person may raise the question of how a person's story can be designated as the authoritative one. After all, the self-narrative of the non-psychotic, the pre-psychotic and the psychotic will diverge considerably. The answer I propose is that coherence is of decisive importance here. The most coherent story is the authoritative one. In (pre-) psychotic stories the internal coherence often is seriously flawed and important parts of the non-psychotic life story sometimes disappear from sight, whereas the non-psychotic self-constituting

narrative is sometimes more consistent but always keen on the integration of the (pre)psychotic story into the self-constituting narrative. Of overriding importance, however, is that the (pre-) psychotic story as a rule lacks sense of reality ('external' coherence). To some extent it does not match with the facts of the world and with the way significant others interpret the (pre-) psychotic self-narrative.

(ii) My narrative outlook on autonomy can make sense of the notion of precedent autonomy and underpin a moral justification of the use of Ulysses contracts. Although psychosis causes narrative disruption of identity, persons are able to bridge the gap by making recurrent psychoses part of their self-narrative, and by prospectively identifying with their future psychotic state (patients composing a Ulysses contract can show experience with psychosis). A core idea in my narrativist contribution to the discussion of advance directives is the person's identification with a set of whole-life, persistent and coherent identity conferring concerns (values and interests, as I also called them). Such entities can survive severe psychosis. In my view, autonomy serves those persistent and coherent values and interests and we best respect the person's autonomy overall when we give priority to the autonomy exercised by a person when most competent to exercise it. This person can decide what set of policies do best to protect the person's autonomy overall. I have my doubts, however, about the stability of what Ronald Dworkin coined the 'integrity-value' of a person's life. Contra Dworkin I take into consideration that the criteria for evaluating what to count valuable and what trivial often change within a human biography. The good news is that in the Ulysses contract case the lifelong persistence of values and interests is not of overriding importance. Psychosis is an episodic phenomenon and misjudgments can be rectified afterwards.

The upshot for the moral assessment of Ulysses contracts is this: accepting the primacy of respect for autonomy as a starting point and taking autonomy as a diachronic matter, the enforcement of Ulysses contracts can be morally justified. First of all, enforcing a Ulysses contract we do respect the precedent autonomous consent of the patient. In addition, in many cases it will be possible to respect the patient's subsequent autonomous consent too. Nevertheless, the moral costs are substantial. Advancing identity conferring concerns the person no longer cares about and treating a relatively competent protesting person under coercion may be harmful as it may be experienced by patients and professionals as traumatic and as an attack on their identities. Moreover, whether coercive therapy succeeds is unpredictable and partly a matter of luck. But expectations are that a short violation of autonomy plus a long period in desirable circumstances after the compulsory intervention on balance will produce optimal self-governance.

(iii) Although the idea of pre-commitment may impress as a seductive strategy, my thought is that the use of advance directives is limited. This becomes clear, when we consider advance directives in cases of severe dementia of the Alzheimer's type. Again, my narrative account of personhood and personal autonomy turns out to be helpful to illuminate the highly problematic authority of advance directives in cases of severe dementia. Comparable with the recurrent psychosis case, the times at which there is no narrator can be bound up in a self-constituting narrative, and so those stages can be attributed to the person who narrates and identifies in advance. But the case of dementia is different, of course. Firstly, the narrative disruption of identity now is permanent and runs deeper. The person disappears. Secondly, identification with the incompetent state of dementia is more

difficult than identification with a period of recurrent psychosis, because the person cannot show own experience with dementia. Thirdly, allowing for possible changes of heart is not an option.

The moral justification of advance directives as a vehicle of precedent autonomy also faces new problems in the Alzheimer's case. I considered advance directives requesting that the severely demented patients not receive life-sustaining treatment should they become life-threatening ill. The first problem is that, in my view, at this stage of dementia the distal binding powers of precedent autonomy have expired. I take a person to be exercising autonomy as long as he is capable, or will be capable at some point in the future to identify with his self-narrative. As a consequence, an advance directive enforcing pre-commitment choices now competes with the moral responsibilities to protect actual vulnerable incompetent patients from harm. My narrativist account of autonomy cannot deliver a decisive moral justification for trumping the actual interests of the severely demented post-person with the interests of the person prior to the onset of dementia. This gave me reason to look elsewhere for an extra argument. To no avail.

The second problem is that the moral debate on advance directives is dominated by the patient-centred perspective. This strikes me as a fatal flaw. We are in the business of life and death now, not of precarious compulsory treatment. Others have to enforce the directive, that is, to withhold treatment and to let the patient die. The moral costs of killing another person who feels happy and comfortable and does not suffer and who can easily be saved, are high and tip the scale. I argued that we have to show mercy not only for the incompetent patient but also for the vulnerability of the care providers. We cannot expect doctors to expose themselves to such moral risks. The moral costs are even so high, in my view, that they put the command envisioned by

advance directives beyond reach. In the case of Alzheimer's disease advance directives are an ineffective method of orchestrating one's future care. They can be no more than a bit of support.

(iv) This leaves persons diagnosed with Alzheimer's disease and confronted with a catalogue of horrors with two strategies that stand as alternatives to advance directives. One possibility is that they accept that autonomy in their future incompetent state of dementia is not an issue and that they will be dependent on others to protect them from harm and indignities. Surely, they can draw up an advance directive, but this is not a necessary means. When severe dementia is setting in, others decide. The second possibility is that they commit pre-emptive suicide after Alzheimer's disease is diagnosed. This is not claiming that it is better to be dead than to be demented. But for some people autonomy, including autonomy over future treatment, is a treasured value that cannot be abandoned and is more valuable than life itself. They want to remain in control and avoid what they see as a troubling, lost existence. For those patients *Romana mors* in the face of Alzheimer is a morally justifiable autonomous act.

BIBLIOGRAPHY

- Anderson J. (2003). 'Autonomy and the Authority of Personal Commitments: From Internal Coherence to Social Normativity' in *Philosophical Explorations* VI (2), 90-108
- Anderson J. & J. Christman (eds.) (2005). *Autonomy and the Challenges to Liberalism. New essays*, Cambridge: Cambridge University Press
- Anderson J. (2008). 'Disputing Autonomy: Second-Order Desires and the Dynamics of Ascribing Autonomy' in *Nordic Journal of Philosophy* 9 (1), 7-26
- Atkins K. & C. Mackenzie (eds.) (2008). *Practical Identity and Narrative Agency*, New York: Routledge
- Baker L. Rudder (2000). *Persons and Bodies. A Constitution view*, Cambridge: Cambridge University Press
- Battin M. Pabst (1994). *The Least Worst Death. Essays in Bioethics on the End of Life*, Oxford: Oxford University Press
- Bennett M.R. & P.M.S. Hacker (2003). *Philosophical Foundations of Neuroscience*, Oxford: Blackwell Publishing
- Betzler M. & B. Guckes (eds.) (2001). *Freiheit und Selbstbestimmung*, Berlin: Akademie Verlag
- Blackburn S. (1998). *Ruling Passions*, Oxford: Clarendon Press
- Boer de M.E., C.M. P.M. Hertogh, R.M. Droës et al (2007). 'Suffering from dementia – the patient's perspective: a review of the literature' in *International Psychogeriatrics* 19, 1021-39
- Bransen J. (1999). *Drie modellen van het menselijk handelen*, Leuven: Peeters
- Bratman M. E. (1987). *Intention, Plans, and Practical Reason*, Cambridge, MA : Harvard University Press
- Bratman M. E. (1999). *Faces of Intention*, Cambridge: Cambridge University Press

- Bratman M. E. (2006). 'A Thoughtful and Reasonable Stability' in H. Frankfurt, *Taking Ourselves Seriously & Getting It Right* (ed. D. Satz), Stanford: Stanford University Press, 77-90
- Bratman M. E. (2007a). *Structures of Agency. Essays*, Oxford: Oxford University Press
- Bratman M. E. (2007b). 'Anchors for Deliberation' in C. Lumer & S. Nannini (eds.), *Intentionality, deliberation and autonomy: the action-theoretic basis of practical philosophy*, Farnham: Ashgate Publishing, 187-205
- Bratman M.E. (2009a). 'Intention, Practical Rationality, and Self-Governance' in *Ethics* 119, 411-43
- Bratman M.E. (2009b). 'Intention Rationality' in *Philosophical Explorations* 12 (3), 227-241
- Buchanan A. E. & D. W. Brock (1990). *Deciding for Others: The Ethics of Surrogate Decision Making*, Cambridge: Cambridge University Press
- Burge T. (1998). 'Reason and the First Person' in B.S.C. Wright and C. MacDonald (eds.), *Knowing Our Own Minds*, Oxford: Oxford University Press
- Burley J. (ed.) (2004). *Dworkin and his critics: with replies by Dworkin*, Oxford: Blackwell Publishing
- Buss S. & L. Overton (eds.) (2002). *Contours of Agency. Essays on Themes from Harry Frankfurt*, Cambridge, MA: The MIT Press
- Chabot B. & S. Braam (2010). *Uitweg. Een waardig levenseinde in eigen hand*, Amsterdam: Nijgh & Van Ditmar
- Christman J. (2004). 'Narrative Unity as a Condition of Personhood' in *Metaphilosophy* 35 (5), 695-713
- Damasio A. (1999). *The Feeling of What Happens. Body and Emotion in the Making of Consciousness*, New York: Harcourt Brace & Company
- Damasio A. (2000). *The Fabric of the Mind: A Neurobiological Perspective*, Den Haag/Rotterdam: NWO/NRC Handelsblad

- Damasio A. (2003). *Looking for Spinoza. Joy, Sorrow, and the Feeling Brain*, Orlando: Harcourt Inc
- Davis J.K. (2004). 'Precedent Autonomy and Subsequent Consent' in *Ethical Theory and Moral Practice* 7 (3), 267-91
- Davis J. K. (2007). 'Precedent Autonomy, Advance Directives, and End-of-Life Care' in B. Steinbock (ed.), *Oxford Handbook of Bioethics*, Oxford: Oxford University Press, 349-74
- Davis J. K. (2008). 'How to Justify Enforcing a Ulysses Contract When Ulysses is Competent to Refuse' in *Kennedy Institute of Ethics Journal* 18 (1), 87-106
- DeGrazia D. (2005). *Human Identity and Bioethics*, New York: Cambridge University Press
- Dresser R. (1982). 'Ulysses and the Psychiatrists: A Legal and Policy Analysis of the Voluntary Commitment Contract' in *Harvard Civil Rights-Civil Liberties Law Review* 16: 777-854
- Dresser R. (1994). 'Confronting the "near irrelevance" of advance directives' in *The Journal of Clinical Ethics* 5, 55
- Dresser R. (2002-3). 'Precommitment: A Misguided Strategy for Securing Death With Dignity' in *Texas Law Review* 81, 1823-47
- Dupuis H. (2008). 'Erger dan kanker? Een essay over dementie' in I. de Beaufort a.o. (eds.), *De Kwestie. Praktijkboek ethiek voor de gezondheidszorg*, Den Haag: Lemma, 256-60.
- Dworkin G. (1970). 'Acting Freely' in *Nous* 4, 367-83
- Dworkin G. (1988). *The Theory and Practice of Autonomy*, Cambridge: Cambridge University Press.
- Dworkin R. (1993). *Life's Dominion: An Argument About Abortion, Euthanasia and Individual Freedom*, New York: Alfred Knopf
- Ekstrom L.W. (2005). 'Alienation, Autonomy, and the Self' in *Midwest Studies in Philosophy* XXIX, 45-67
- Elliott C. (1999). *A Philosophical Disease: Bioethics, Culture and Identity*, New York: Routledge
- Elster J. (2003). 'Don't Burn Your Bridge Before You Come To It:

- Some Ambiguities and Complexities of Precommitment' in *University of Texas Law Review* 81, 1751-88
- Feinberg J. (1986). *Harm to Self*, New York: Oxford University Press
- Ferrero L. (2002). *Making Up One's Self*, Ph. D. Dissertation, Harvard University
- Flanagan O. (1996). *Self Expressions. Mind, Morals, and the Meaning of Life*, Oxford: Oxford University Press
- Flanagan O. (2002). *The Problem of the Soul. Two Visions of Mind and How to Reconcile Them*, New York: Basic Books
- Flanagan O., G. D. Fireman & T. E. McVay, jr. (eds.) (2003). *Narrative and Consciousness. Literature, Psychology, and the Brain*, Oxford: Oxford University Press
- Frankfurt H. G. (1988). *The Importance of What We Care About*, Cambridge: Cambridge University Press
- Frankfurt H. G. (1999). *Necessity, Volition, and Love*, Cambridge: Cambridge University Press
- Frankfurt H. G. (2000). 'Rationalism in Ethics' in Betzler M. & B. Guckes (eds.) *Autonomes Handeln. Beiträge zur Philosophie von Harry G. Frankfurt*, Berlin: Akademie Verlag
- Frankfurt H. G. (2001). *Some Mysteries of Love*, University of Kansas: Department of Philosophy
- Frankfurt H. G. (2004a). *The Reasons of Love*, Princeton: Princeton University Press
- Frankfurt H. G. (2004b). 'Disengaging Reason' in Wallace J.W., P. Pettit, S. Scheffler & M. Smith (eds.), *Reason and Value. Themes from the Moral Philosophy of Joseph Raz*, Oxford: Clarendon Press, 119-128
- Frankfurt H. G. (2006). *Taking Ourselves Seriously & Getting It Right* (ed. D. Satz), Stanford: Stanford University Press
- Fulford K.W.M. (2000). 'Disordered Minds, Diseased Brains and Real People' in C. Heginbotham (ed.) *Philosophy, Psychiatry and Psychopathy. Personal identity in mental disorder*, Aldershot: Ashgate, 47-73

- Fulford K.W.M., D. L. Dickenson & T. H. Murray (eds.) (2002). *Healthcare Ethics and Human Values. An Introductory Text with Readings and Case Studies*, Oxford: Blackwell Publishers
- Fulford K.W.M., T. Thornton & G. Graham (eds.) (2006). *Oxford Textbook of Philosophy and Psychiatry*, Oxford: Oxford University Press
- Gallagher S. (2000). 'Philosophical Conceptions of the Self: *Implications for Cognitive Science*' in *Trends in Cognitive Sciences* 4, 14-21
- Gallagher S. (2007). 'Pathologies in Narrative Structures' in D. Hutto (ed.), *Narrative and Understanding Persons*, Cambridge: Cambridge University Press, 203-224
- Gallagher S. & D. Zahavi (2008). *The Phenomenological Mind. An Introduction to Philosophy of Mind and Cognitive Science*, New York: Routledge
- Glover J. (1988). *I: The Philosophy and Psychology of Personal Identity*. London: Penguin
- Glover J. (2004). 'Towards Humanism in Psychiatry' in G.B. Peterson (ed.), *The Tanner Lectures on Human Values* 24, Utah: The University of Utah Press, 509-552
- Goldie P. (2000). *The Emotions: A Philosophical Exploration*, Oxford: Oxford University Press
- Goldie P. (2003). 'Narrative and Perspective; Values and Appropriate Emotions' in A. Hatzimoysis (ed.), *Philosophy and the Emotions*, Cambridge: Cambridge University Press, 201-220
- Goldie P. (2004). *On Personality*, London: Routledge
- Goldie P. (2009). 'Narrative Thinking, Emotion, and Planning' in N. Carroll (ed.), *The Poetics, Aesthetics, and Philosophy of Narrative*, Oxford: Wiley-Blackwell, 97-106
- Gowans C. W. (2002). 'Practical Identities and Autonomy: Korsgaard's Reformation of Kant's Moral Philosophy' in *Philosophy and Phenomenological Research* LXIV (3), 546-570
- Graham G. & G. Lynn Stephens (eds.) (1994). *Philosophical Psychopathology*, Cambridge: The MIT Press

- Hartogh G. den (2004). 'The Authority of Intention' in *Ethics* 115, 6-34
- Hatzimoysis A. (ed.) (2003). *Philosophy and the Emotions*, Cambridge: Cambridge University Press
- Herman D. (2009). 'Storied Minds. Narrative Scaffolding for Folk Psychology' in *Journal of Consciousness Studies* 16 (6-8), 40-68
- Hertogh C.M.P.M. (2009). 'The role of advance euthanasia directives as an aid to communication and shared decision-making in dementia' in *Journal of Medical Ethics* 35, 100-3
- Hill T. E. (1991). 'Self-regarding suicide: a modified Kantian view' in *Autonomy and self-respect*, Cambridge: Cambridge University Press, 85-103
- Hume D. (1978). *A Treatise of Human Nature* (ed. L.A. Selby-Bigge), Oxford: Clarendon Press
- Hutto D. (2007). 'The Narrative Practice Hypothesis: Origins and Applications of Folk Psychology' in D. Hutto (ed.), *Narrative and Understanding Persons*, Cambridge: Cambridge University Press, 43-68
- Hutto D. (2009). 'Folk Psychology as Narrative Practice' in *Journal of Consciousness Studies* 16 (6-8), 9-39
- James H. (ed.) (1920). *The Letters of William James*. Vol. 1, Boston: Atlantic Monthly Press
- Jamison K.R. (2000). *Night Falls Fast. Understanding Suicide*, London: Picador (first published 1999 by Alfred A. Knopf, New York)
- Jaspers K. (19739). *Allgemeine Psychopathologie*, Berlin/Heidelberg/ New York: Springer Verlag
- Jaworska A. (1999). 'Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value' in *Philosophy and Public Affairs* 28, 105-38
- Jaworska A. (2005). 'From Caring to Self-Governance: The Bare Bones of Autonomy and the Limits of Liberalism' published online in *APA Newsletter on Philosophy and Medicine* 5(1), 19-23

- Jaworska A. (2007a). 'Caring and Full Moral Standing' in *Ethics* 117, 460-97
- Jaworska A. (2007b). 'Caring and Internality' in *Philosophy and Phenomenological Research* 74 (3), 529-68
- Jaworska A. (2009). 'Caring, Minimal Autonomy, and the Limits of Liberalism' in H. Lindemann, M. Verkerk & M. Urban Walker (eds.), *Naturalized Bioethics*, Cambridge: Cambridge University Press, 80-105
- Kendler K.S. (2005). 'Toward a Philosophical Structure for Psychiatry' in *American Journal of Psychiatry* 162 (3), 433-440
- Kennett J. (2001). *Agency and Responsibility. A Common-sense Moral Psychology*, Oxford: Clarendon Press
- Korsgaard C. M. (1989). 'Personal Identity and the Unity of Agency: A Kantian Response to Parfit' in *Philosophy and Public Affairs* 18 (2), 101-132
- Korsgaard C.M. (1996). *The sources of normativity*, Cambridge: Cambridge University Press
- Korsgaard C.M. (1998). 'Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind' in *Ethics* 109, 49-66
- Korsgaard C.M. (1999). 'Self-constitution in the Ethics of Plato and Kant' in *The Journal of Ethics* 3, 1-29
- Korsgaard C.M. (2006). 'Morality and the Logic of Caring' in H. Frankfurt, *Taking Ourselves Seriously & Getting It Right* (ed. D. Satz), Stanford: Stanford University Press, 55-76
- Korsgaard C.M. (2008). *The Constitution of Agency*, Oxford: Oxford University Press
- Korsgaard C.M. (2009). *Self-Constitution. Agency, Identity, and Integrity*, Oxford: Oxford University Press
- Korsgaard C.M. (2010a). *Moral Animals. Three Lectures*, manuscript
- Korsgaard C.M. (2010b). *Comments on Jonathan Lear's Tanner Lectures 'Irony and Identity'*, Harvard University November 2009, manuscript

- Kuczewski M. (1994). 'Who's Will Is It Anyway? A Discussion of Advance Directives, Personal Identity, and Consensus in Medical Ethics' in *Bioethics* 8, 27-48
- Kusters W. (2004), *Pure waanzin. Een zoektocht naar de psychotische ervaring*, Amsterdam: Uitgeverij Nieuwezijds
- Laden A.S. (2009). 'The trouble with prudence' in *Philosophical Explorations* 12 (1), 19-40
- Lindemann Nelson H. (2001). *Damaged Identities. Narrative Repair*, Ithaca: Cornell University Press
- Locke J. (1979). *An Essay Concerning Human Understanding* (ed. P. Nidditch), Oxford: Clarendon Press
- McAdams D.P. (1999). 'Personal Narratives and the Life Story' in L.A. Pervin & O.P. John (eds.), *Handbook of Personality Theory and Research* (Second Edition), NY/London: The Guilford Press, 478-500
- MacIntyre A. (1981). *After Virtue*, London: Duckworth
- Mackenzie C. & N. Stoljar (eds.) (2000). *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, New York: Oxford University Press
- Mackenzie C. & J. Poltera (2009). *Narrative Integration, Fragmented Selves and Autonomy*, manuscript
- McMahan J. (2002). *The Ethics of Killing. Problems at the Margins of Life*, Oxford: Oxford University Press
- Meyers D. Tietjens (2004). 'Narrative and Moral Life' in C. Calhoun (ed.) *Setting the Moral Compass. Essays by Women Philosophers*, Oxford: Oxford University Press, 288-305
- Moran R. (2001). *Authority and Estrangement*, Princeton: Princeton University Press
- Moran R. (2002). 'Frankfurt on Identification: Ambiguities of Activity in Mental Life' in Buss S. & L. Overton (eds.), *Contours of Agency. Essays on Themes from Harry Frankfurt*, Cambridge, MA: The MIT Press, 189-217

- Noonan H. W. (2003). *Personal Identity*, London: Routledge
- Nussbaum M. C. & J. Glover (eds.) (1995). *Women, Culture and Development: A Study of Human Capabilities*, Oxford: Clarendon Press
- Olson E.T. (2003). 'Was Jekyll Hyde?' in *Philosophy and Phenomenological Research* LXVI (2), 328-448
- Parfit D. (1984). *Reasons and Persons*, Oxford: Clarendon Press
- Perry J. (1976). 'The importance of being identical' in A.O. Rorty, *The identities of persons*, Berkeley: University of California Press
- Pettit P. (2001). *A Theory of Freedom. From the Psychology to the Politics of Agency*, Oxford: Oxford University Press
- Quante M. (2002). *Personales Leben und menschlicher Tod. Personale Identität als Prinzip der biomedizinischen Ethik*, Frankfurt am Main: Suhrkamp
- Quante M. (2007). *Person*, Berlin: Walter de Gruyter
- Raz J. (1999). 'When We Are Ourselves: The Active and the Passive' in J. Raz, *Engaging Reason: On the Theory and Value of Action*, Oxford: Oxford University Press, 5-21
- Ricoeur P. (1992). *Oneself as Another, Chicago: The University of Chicago Press* (Originally published as *Soi-même comme un autre*, Paris: Editions du Seuil, 1990)
- Rorty R. (2007). 'Kant vs. Dewey. The current situation of moral philosophy' in R. Rorty, *Philosophy as cultural politics. Philosophical Papers* 4, Cambridge: Cambridge University Press, 184-202
- Rovane C. (1998). *The Bounds of Agency. An Essay in Revisionary Metaphysics*, Princeton: Princeton University Press
- Scanlon T. M. (1998). *What We Owe to Each Other*, Cambridge, MA: Harvard University Press
- Schechtman M. (1996). *The Constitution of Selves*, Ithaca: Cornell University Press
- Schechtman M. (2001). 'Empathic Access: The Missing Ingredient in Personal Identity' in *Philosophical Explorations* IV (2), 95-111

- Schechtman M. (2004). 'Self-expression and Self-control' in *Ratio* XVII (4), 409-427
- Schechtman M. (2005a). 'Personal Identity and the Past' in *Philosophy, Psychiatry & Psychology* 12 (1), 9-22
- Schechtman M. (2005b). 'Experience, Agency, and Personal Identity' in *Social Philosophy & Policy* (22) 2, 1-24
- Schechtman M. (2007). 'Stories, Lives, and Basic Survival: A Refinement and Defense of the Narrative View' in D. Hutto (ed.), *Narrative and Understanding Persons*, Cambridge: Cambridge University Press, 155-178
- Schechtman M. (2008). 'Diversity in unity: practical unity and personal boundaries' in *Synthese* 162, 405-423
- Schermer M. (2009). 'Wiens leven, welke belangen? Over dementie, wilsverklaringen en het "leven als geheel"' in *Als vuur. Opstellen voor Govert den Hartogh ter gelegenheid van zijn emeritaat*, Den Haag: Boom Juridische uitgevers, 109-25
- Searle J. R. (2001). *Rationality in Action*, Cambridge: MIT Press
- Shoemaker D.W. (1996). 'Theoretical Agents and Practical Agents' in *Philosophy and Public Affairs* 25 (4), 318-332
- Shoemaker D.W. (2003). 'Caring, Identification, and Agency' in *Ethics* 114 (1), 88-118
- Shoemaker D.W. (2009). *Personal Identity and Ethics. A Brief Introduction*, Ontario: Broadview Press
- Sie M.M.S.K., M. Slors & B. Van den Brink (eds.) (2004). *Reasons of One's Own*. Aldershot: Ashgate.
- Slors, M. (2001). *The Diachronic Mind. An Essay on Personal Identity, Psychological Continuity and the Mind-Body Problem*, Dordrecht: Kluwer Academic Publishers
- Slors M. (2004). 'Care for One's Own Future Experiences' in *Philosophical Explorations* 7 (2), 183-195
- Smith M. (1999). 'Search for the Source' in *The Philosophical Quarterly* 49 (196), 384-394

- Solomon R. (2004). 'Emotions, Thoughts, and Feelings. Emotions as Engagements with the World' in R. Solomon (ed.), *Thinking about Feeling. Contemporary Philosophers on Emotions*, Oxford: Oxford University Press, 76-88
- Spelley R. (2003). 'Reviving Ulysses Contracts' in *Kennedy Institute of Ethics Journal* 13 (4), 373-392
- Sperling D. (2008). *Posthumous Interests. Legal and Ethical Perspectives*, Cambridge: Cambridge University Press
- Strawson G. (2004). 'Against Narrativity' in *Ratio* XVII (4), 428-452
- Strawson G. (2009). 'Why I have no future' in *The Philosopher's Magazine* 38, October 12
- Tappolet C. (2006). 'Autonomy and the Emotions' in *European Journal of Analytic Philosophy* 2 (2), 45-59.
- Taylor C. (1989). *Sources of the Self: The Making of Modern Identity*, Cambridge: Cambridge University Press
- Taylor J. S. (2005). *Personal Autonomy. New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, New York: Cambridge University Press
- Underwood G. (ed.) (2001). *Oxford Guide to the Mind*, Oxford: Oxford University Press
- Velleman J. D. (1999). 'A Right of Self-Termination?' in *Ethics* 109, 606-628
- Velleman J. D. (2000). *The Possibility of Practical Reason*, Oxford: Clarendon Press
- Velleman J. D. (2003). 'Narrative Explanation' in *The Philosophical Review* 112 (1), 1-25
- Velleman J. D. (2006). *Self to Self*, New York: Cambridge University Press
- Velleman J. D. (2008). 'Beyond Price' in *Ethics* 118, 191-212
- Wallace R. J. (1999). 'Three Conceptions of Rational Agency' in *Ethical Theory and Moral Practice* 2, 217-242
- Wallace R. J. (2000). 'Caring, Reflexivity, and the Structure of

- Volition' in M. Betzler & B. Guckes (eds.), *Autonomes Handeln. Beiträge zur Philosophie von Harry G. Frankfurt*, Berlin: Akademie Verlag
- Wallace R. J. (2006). *Normativity & the Will. Selected Essays on Moral Psychology and Practical Reason*, Oxford: Oxford University Press
- Watson G. (1987). 'Free Action and Free Will' in *Mind* 96, 145-72
- Widdershoven G.A.M. & R.L.P. Berghmans (2001). 'Advance directives in psychiatric care: a narrative approach' in *Journal of Medical Ethics* 27, 92-7
- Widdershoven G.A.M. & R.L.P. Berghmans (2007). 'Coercion and pressure in psychiatry: lessons from Ulysses' in *Journal of Medical Ethics* 33, 560-563
- Wijtsbek H. (2010). "'To thine own self be true": on the loss of integrity as a kind of suffering' in *Bioethics* ISSN 0269-9702; 1467-8519 (online)
- Wilkes K.V. (1988). *Real People. Personal Identity without Thought Experiments*, Oxford: Clarendon Press
- Wilkes K.V. (1998). 'Gnothi Seauton (Know Thyself)' in *Journal of Consciousness Studies* 5, 153-65
- Williams B. (2002). *Truth and Truthfulness. An Essay in Genealogy*, Princeton: Princeton University Press
- Willigenburg T. van (ed.) (2002). *Zelfbinding in de psychiatrie. Een empirische, wijsgerige en juridische studie naar zelfbindingsverklaringen in de geestelijke gezondheidszorg*, Den Haag/Rotterdam: NWO/Erasmus Universiteit Rotterdam
- Willigenburg T. van & P.J. J. Delaere (2005). 'Protecting Autonomy as Authenticity Using Ulysses Contracts' in *Journal of Medicine and Philosophy* 30, 395-409
- Wolf S. (2007). 'The Meanings of Lives' in J. Perry, M. Bratman & J.M. Fischer (eds.), *Introduction to Philosophy. Classical and Contemporary Readings* (fourth edition), Oxford: Oxford University Press, 62-73

- Yolton J. W. (2004). *The Two Intellectual Worlds of John Locke. Man, Person, and Spirits in the "Essay"*, Ithaca: Cornell University Press
- Zhu J. & P. Thagard (2002). 'Emotion and action' in *Philosophical Psychology* 15 (1), 19-36

NEDERLANDSE SAMENVATTING

Mijn proefschrift gaat over psychiatrie en filosofie. Ik denk dat filosofen de psychiatrie een dienst kunnen bewijzen. Volgens de arts, psycholoog en filosoof William James (1842-1910) is filosofie ‘an unusually stubborn effort to think clearly’, een ongebruikelijk hardnekkige poging om helder te denken. En binnen de psychiatrie is de behoefte aan helder denken groot. Filosofen kunnen helpen om de betekenis te completeren van lastige concepten waarmee hulpverleners in de geestelijke gezondheidszorg hun patiënten proberen te begrijpen en bij te staan. Mijn proefschrift wil zo’n bijdrage zijn. Ik probeer er in te verhelderen wat persoon-zijn is.

Grosso modo zijn er drie filosofische manieren om de vraag ‘wat is een persoon?’ te benaderen. De eerste insteek is een *conceptuele* en gaat terug op inzichten van de *ordinary language* filosoof Ludwig Wittgenstein (1889-1951). Hier wordt als volgt geredeneerd: ons concept ‘persoon’ is bij uitstek toepasbaar op menselijke wezens – belichaamde sociale wezens die deel uitmaken van een morele gemeenschap –, en voldoet uitstekend in ons dagelijks spreken over onszelf en anderen. Maar Jekyll-en-Hyde fenomenen of psychotische mensen vormen een grensgeval, een afwijking van de norm. In dergelijke gevallen is ons gangbare persoonsbegrip niet langer inzetbaar. Daar is weinig aan te doen.

De tweede denkroute stelt de vraag centraal naar letterlijke identiteit. Ze is de inzet van een zeer ingewikkeld debat over *ontologische* identiteit. De precieze vraag luidt hier: is er een criterium te vinden voor persoonlijke identiteit dat de noodzakelijke en voldoende voorwaarden levert, op grond waarvan we kunnen volhouden dat een persoon op een later tijdstip en een persoon

op een eerder tijdstip fasen van een en dezelfde persoon zijn? Vinden we bijvoorbeeld dat we een en dezelfde persoon zijn omdat er sprake is van een bestendig aanwezig lichaam? Dan kiezen we voor een lichamelijk criterium voor persoonlijke identiteit. Of zijn we misschien een en dezelfde persoon omdat onze actuele mentale toestanden (intenties, overtuigingen, voorkeuren, herinneringen, waarden, karaktertrekken...) verbonden zijn met onze eerdere en latere mentale toestanden? In dat geval horen we bij het kamp van de voorstanders van een psychologisch criterium voor persoonlijke identiteit. Er zijn nog meer opties. Misschien denken we dat een persoon een tijdloze kern is, die een leven lang (en mogelijk ook voorbij de grenzen van geboorte en dood) identiek blijft. Dan hangen we een (overigens lastig verdedigbare) 'strikte identiteitstheorie' aan. De strijd over de ontologische vraag naar persoonlijke identiteit is nog onbeslecht.

Een aantal filosofen kiest daarom voor een *praktische* benadering van het raadsel van persoon-zijn. In ons dagelijks leven gaan we uit van een sterke eenheid van de persoon. Ik ben het die mij vandaag iets voorneem bijvoorbeeld, en ik ben het vervolgens weer die mij over vijf jaar gebonden weet door mijn eerdere voornemen. Zonder een sterk element van identiteit zou het zinloos zijn om mij iets voor te nemen. Ik ga ervan uit dat mijn leven en mijn persoon door de tijd heen een eenheid vormen; en dat niet een invaller maar ik zélf over vijf jaar mijn voornemen al of niet zal uitvoeren. In deze derde, praktische benadering staat de vraag centraal welke handelingen, ervaringen en psychologische karakteristieken aan een persoon toegeschreven dienen te worden om haar te maken tot de persoon die ze is.

Het is de vraag naar praktische identiteit die ik probeer te beantwoorden in voorliggend proefschrift, dat uit twee delen bestaat. In het tweede deel (hoofdstuk 4) pas ik mijn visie op persoon-

zijn en autonomie toe op zelfbindingscasus in de psychiatrie. In het eerste deel (de hoofdstukken 1, 2 en 3) ontwikkel ik die visie in termen van actorschap. Mensen zijn handelende wezens die beschikken over een zekere mate van zelfbewustzijn. Anders dan de meeste andere dieren zijn ze zich bewust van de redenen voor hun handelingen. Dat mensen zelfbewuste gevers en nemers van redenen zijn brengt twee vermogens met zich mee. Ten eerste stelt het hen in staat om controle uit te oefenen over hun handelen; ze kunnen manieren vinden om richting aan hun handelen te geven. Laten we dit het vermogen tot normatieve *zelfsturing* noemen. En ten tweede: telkens wanneer menselijke actoren door afwegingen en keuzes te maken de regie voeren over hun handelen maken ze zichzelf tot de auteurs van hun handelen. Ze beslissen tegelijk wie ze willen zijn. Als rationele actoren hebben mensen, naast het vermogen tot normatieve zelfsturing, ook een vermogen tot normatieve *zelf-constitutie*, d.w.z. tot het construeren van norm-gestuurde, praktische vormen van identiteit. Over dit laatste vermogen gaan hoofdstuk 1 en 2 van mijn proefschrift.

In het eerste hoofdstuk neem ik het lopend wijsgerig debat over actorschap als uitgangspunt, in het bijzonder de visie op zelf-constitutie van Christine Korsgaard (1952), die een beschrijving geeft van reflectief actorschap en daar een uitgewerkte kijk op persoonlijke identiteit uit afleidt in termen van deliberatief handelen. Mensen zijn geboren om te handelen, stelt ze, en actoren bewegen zich door een 'ruimte van redenen'. Ze hebben het vermogen om te delibereren, i.e. om te kiezen voor bepaalde handelingen op grond van redenen die eerst door een proces van nauwgezet rationeel onderzoek zijn gegaan. Wanneer een actor beredeneerd kiest op grond van welke reden zij tot handelen wil overgaan, en vervolgens ook handelt op grond van die reden, dan handelt zij op een manier die bepalend is voor wie zij is, met

andere woorden op een manier die haar identiteit constitueert. Volgens Korsgaard is persoonlijke identiteit praktische identiteit en verknoopt met praktisch redeneren. De *substantia nigra* van zelf-constitutie is volgens haar de normatieve kracht van praktisch redeneren. Het is ten diepste de normatieve consistentie van redenen die persoon-zijn constitueert en personen tot een eenheid maakt. Daar kan ik mij wel in vinden, maar niet zonder een paar twijfels op tafel te leggen.

Ik denk namelijk dat Korsgaard de unificerende kracht van de menselijke rationele vermogens overschat. Haar overtuiging dat menselijke wezens van nature zijn toegerust met rationele vermogens, die wel tot rationele eenheid van de persoon *moeten* leiden, is mijns inziens ongedekt. Al wat uit Korsgaards visie kan volgen is dat personen alomvattende rationele eenheid binnen zichzelf *kunnen* bereiken. De natuurlijke menselijke rationele vermogens dicteren niet uit zichzelf in welke mate mensen rationele eenheid moeten bereiken en langs welke wegen. Het is waar dat Korsgaard, zich baserend op Aristoteles, probeert extra argumenten toe te voegen ter ondersteuning van haar éénmakingsargument. Desalniettemin blijft haar nadruk op de unificerende rol van de praktische rede, i.c. het Korsgaardiaanse principe van autonomie (universaliseerbaarheid), naar mijn overtuiging overtrokken. We kunnen toe met een minder formele, minder principe-gestuurde, meer instrumentele rol voor de redenen die persoon-zijn constitueren en personen tot een eenheid smeden. Bij implicatie dient dan ook de unificatievereiste te worden gematigd. Personen worden geconstitueerd door een weefsel van zichzelf corrigerende redenen. Elke reden kan onder kritiek worden gesteld, zij het niet alle redenen tegelijk. We kunnen hier denken aan het fameuze schip van Neurath. Alle persoonlijke redenen zijn experimenteel, feilbaar en vervangbaar zoals de planken van de boot van Neurath. Dat wil niet zeggen dat personen helemaal niet uit zouden

zijn op eenheid. Meestal kiezen personen er niet voor om als een scheepswrak door het leven te gaan. Maar om de min of meer vereende levens te leiden van min of meer vereende personen is niet meer nodig dan een humeaanse instrumentele opvatting van redelijkheid. Personen kunnen zich behalve op doel-middel redeneren verder nog beroepen op een humeaanse conceptie van 'reflectieve rationaliteit': een duurzaam, algemeen standpunt vanwaaruit idiosyncrasieën kunnen worden gecorrigeerd. Kort en goed, Korsgaards 'reason view' is niet helemaal bevredigend en kan niet meer zijn dan een veelbelovende eerste stap op weg naar een persoonstheorie in termen van actorschap.

Mijn voorstel is vervolgens om de *redenen visie* op personen op te werken tot een *narratieve visie* op persoon-zijn. We kunnen redenen beter beschouwen als het begin van een verhaal over onszelf, d.w.z. als een 'rationale' die onze handelingen begrijpelijk maakt als coherente ontwikkelingen van een levensverhaal. Ik heb tenminste drie argumenten voor zo'n zet. Ten eerste: verhalen zijn naar alle waarschijnlijkheid de meest natuurlijke en eenvoudige manier om menselijke ervaringen en menselijke kennis te ordenen. Zij lijken een aangewezen vindplaats voor wie op zoek is naar een adequaat begrip van persoon-zijn. Ten tweede: een coherent zelf-constituerend verhaal doet meer recht aan de multiplexiteit, discontinuïteiten en veranderingen van personen dan unificerende redenen kunnen doen. In een narratieve visie op personen kan de unificatie-eis worden vervangen door een coherentie-eis en kan de rede een minder potente, meer instrumentele rol vervullen. Een derde argument om een visie op personen in termen van redenen uit te breiden naar een narratieve visie op persoon-zijn is dat narratieve zelf-constitutie beter kan worden verbonden met het menselijk lichaam. Zo kan een narratieve benadering ook nog een lacune in Korsgaards visie op zelf-constitutie opvullen: het ontbreken van uitgewerkte ideeën

over de belichaming van personen. Redenen alleen kunnen niet geacht worden al het zelf-constituerende en éénmakende werk te doen. Een persoon is meer dan haar rationele natuur. Ook emoties, verlangens en lichamelijkheid dragen bij aan persoon-zijn.

In het tweede hoofdstuk ontwikkel ik zo'n narratieve visie op zelf-constitutie gebaseerd op het werk van Marya Schechtman (1960). Persoonlijke identiteit bestaat nu uit narratief zelf-verstaan, uit het vermogen om een autobiografisch narratief gaande te houden. Zo'n zelf-narratief organiseert en integreert disparate ervaringen en handelingen synchroon en diachroon tot een min of meer coherent verhalend geheel en constitueert zo de persoon. Interne samenhang, een zekere mate van articuleerbaarheid, overeenstemming met de feiten van de wereld en consensus met de interpretaties van anderen zijn de belangrijkste standaards voor gelukke zelf-narratieven. Zo'n Schechtmanianse visie op zelf-constitutie vermijdt de Korsgaardiaanse overschatting van de formele, eenmakende krachten van de rede en neemt genoegen met het meer inhoudelijk, vernaculair criterium van coherentie voor zelf-narratieven. De door mij voorgestelde narratieve kijk op zelf-constitutie beschouwt zelf-narratieven bovendien als belichaamde en gesitueerde structuren (Schechtman zelf onderbeschrijft de rol van het lichaam).

Om het idee van de belichaming van zelf-narratieven te verduidelijken kies ik voor een empirische lijn. De neurowetenschapper Antonio Damasio (1944) toont aan dat zelf-constituerende narratieven een lichamelijke, neurofysiologische basis hebben. Ze bouwen, *inter alia*, voort op de aanwezigheid van een 'kern-zelf'. Damasio's bevindingen maken ook ruimte voor de affectieve vereisten waaraan zelf-narratieven moeten voldoen. Zijn onderzoek maakt aannemelijk dat autobiografische narratieven vervlochten zijn met emotie en gevoel. Persoon is men voor een

belangrijk deel door hoe het voelt om die persoon te zijn. En gevoelens houden zelf-narratieven op koers. Ook Schechtman is zich bewust van het belang van affect en de rol die empathische toegang speelt bij de totstandkoming van zelf-narratieven. Damasio wijst eveneens – net als Schechtman – op de gesitueerdheid van personen. Een volledig begrip van zelf-constitutie hoort niet alleen verder te gaan dan de hersenen en het hele menselijk lichaam te bestrijken, het dient ook de omringende wereld een aandeel te geven. Personen zijn voortdurend in interactie met hun omgevingen. Hoe iemand zich voelt of over zichzelf denkt wordt deels bepaald door diens (sociale) omgeving. Persoon kan men alleen worden wanneer men uitgaat van een sociaal-cultureel bepaalde achtergrondconceptie van zichzelf als een in de tijd voortbestaand individu, dat het leven van een persoon leidt. Bovendien worden zelf-narratieven meebepaald door de waarnemingen en interpretaties van anderen. De individuele persoon heeft om zo te zeggen niet het laatste woord.

Ik denk dat een van Schechtmans werk uitgaande visie op narratieve zelf-constitutie serieus kan bijdragen aan een beter begrip van de constitutie van personen. Toch moet haar positie enigszins worden aangepast. Niet alleen moet er meer aandacht zijn voor de rol van het lichaam, ook de articulatie-eis vraagt om enige nuancering. De filosoof Galen Strawson (1952) maakt bezwaar tegen de in deze eis vervatte suggestie dat personen zichzelf al pratend tot leven zouden moeten wekken. Echter, de articulatie-eis houdt niet in dat, om een persoon te zijn, iemand altijd een volledig uitgewerkte uiteenzetting paraat moet hebben over waarom alles in haar leven loopt zoals het loopt. De gedachte is veeleer dat personen een soort verplichting moeten voelen om desgevraagd tekst en uitleg te geven bij hun doen en laten, en daartoe ook tot op zekere hoogte in staat moeten zijn. Daar komt bij dat, hoewel zelfbeschrijvingen van personen in taal zijn

gevat, personen nooit volledig te vangen zijn onder hetgeen ze weten te articuleren over zichzelf. Zelf-concepties opereren ook impliciet en automatisch. Het zijn bio-psychologische constructen die zich maar gedeeltelijk laten uitdrukken in verbale vorm.

Zoals eerder aangekondigd wil ik mijn ideeën over praktische identiteit vruchtbaar maken voor de praktijk van zelfbinding in de psychiatrie. Zelfbindingsverklaringen in de psychiatrie zijn geschreven documenten waarin de auteur in goede tijden – op een moment in haar leven waarop zij vrij van symptomen is en competent om te beslissen over latere medische behandeling – aangeeft dat zij onder dwang behandeld wil worden in ‘slechte’ tijden, d.w.z. in een fase van haar leven waarin ze tijdelijk aan een psychiatrische stoornis lijdt (een manische psychose bijvoorbeeld of een periode van splitszinnigheid). Ook een hieraan verwante biomedische procedure wil ik tegen het licht houden: wilsverklaringen van alzheimerpatiënten. Bij beide strategieën lijkt er sprake van een ‘eerder zelf’ dat autoriteit uitoefent en haar wil oplegt aan een ‘later zelf’. Dit roept vragen op over (respect voor) autonomie. Alvorens ik mijn ideeën over praktische identiteit vruchtbaar kan maken voor de psychiatrische praktijk zal ik daarom eerst ook nog na moeten denken over het vraagstuk van de zelfsturing van personen. In hoofdstuk 3 ontwerp ik een theorie over autonomie van personen, die voortbouwt op mijn narratieve identiteitstheorie.

Schechtmans narratieve persoonstheorie zegt weinig over autonomie maar kan beslist worden aangevuld met een visie op persoonlijke autonomie. Voortbouwend op het werk van Harry Frankfurt en Michael Bratman over autonomie is dit de oplossing die ik voorstel: een persoon handelt autonoom, wanneer zij zich identificeert met (delen van) haar zelf-narratief. Identificatie acht ik fundamenteel voor autonoom actorschap en dito

persoon-zijn. Vanuit mijn oogpunt wordt het menselijk zelfbewustzijn gekarakteriseerd door een hiërarchie van attitudes over attitudes. Met hiërarchie bedoel ik dan dat er mentale ruimte is voor een inschatting en bekrachtiging (of verwerping) van een zelf-narratief. Menselijke wezens zijn in staat om zich tot hun mentale attitudes te verhouden. Dit kan verschillende vormen aannemen. Een zich ontwikkelend zelf-narratief bevat redenen, herinneringen, intenties, bekommernissen, emoties en dies meer. Deze zijn zo met elkaar verknoot dat het moeilijk is om ze scherp af te bakenen. Maar we kunnen onderscheid maken tussen tenminste drie aspecten van identificatie. Deels is identificatie een daad van reflectieve bekrachtiging van redenen tot handelen; dit was Korsgaards benaderwijze. Zoals we kunnen leren van Harry Frankfurt (1929) kan identificatie ook een daad zijn van acceptatie van *carings*; met deze laatste worden mensen of zaken bedoeld waar personen om geven, vaak zonder dat ze daar zelf veel aan kunnen doen. Frankfurt is vooral gespitst op de conatieve aspecten van autonomie en laat hiermee de grenzen zien van beredeneerd zelf-gestuurd handelen. Autonome personen zijn zowel ruiter als paard. Soms ziet een ruiter zich genoopt om, wil hij niet van zijn paard geworpen worden, zijn paard dan maar in de richting te sturen die het uit wil. Tenslotte is identificatie voor een deel ook een daad van emotionele instemming. Emoties geven aan wat relevant is in ons zelf-narratief. Die relevantie is niet altijd met intellectuele middelen vaststelbaar en blijkt vaak een gevoelskwestie. En dat gevoel kunnen we bevestigen. Autonomie is ook een kwestie van gevoelde intensiteit, van jezelf intenser ervaren.

Wat ik verder nog beargumenteer is dat de autonomie van een persoon toeneemt wanneer zij een attitude bekrachtigt die samenhangt met andere attitudes in de psychische huishouding van die persoon. Coherentie speelde al een belangrijke rol in

Schechtmans narratieve kijk op zelf-constitutie. Maar niet alleen het zelf-narratief dient aan de coherentievereiste te voldoen. Ook de voortdurende bekrachtiging ervan dient enigszins stabiel te zijn en een aantal centrale, samenhangende waarden van de betreffende persoon te weerspiegelen. Michael Bratman (1945) benadrukt dit, en ik volg hem hierin. Zelf-sturend op *narratio* gebaseerd actorschap bevat normen van praktische rationaliteit: een hang naar consistentie, naar doel-middel-coherentie en naar redelijke stabiliteit door de tijd heen. Opnieuw blijkt *coherentie* dus een belangrijke standaard te zijn. Normaliter proberen autonome personen uit te stijgen boven een onsamenhangende stooftop van intenties, emoties en overwegingen. In geval van onsamenhangende intenties, gedragslijnen en plannen of in gevallen van een onsamenhangend zelf-narratief weet de persoon immers niet langer waar hij staat en hoe hij zijn leven als persoon moet leiden. Nu zou men zich kunnen afvragen: waarom moeten we eigenlijk kiezen voor het genoemde lijstje normen van praktische rationaliteit? Ik denk dat de meest relevante reden voor de eis van coherentie is dat we reden hebben om sturing te geven aan onze eigen levens. Mijn opvatting van autonomie is, net als die van Bratman, ten diepste een prudente. Coherentie draagt bij aan de effectiviteit van ons actorschap en aan de rijkdom van onze (sociale) levens. Ze draagt ertoe bij dat we onze levensdoelen bereiken en onszelf niet voor de voeten lopen. In het beeld van autonomie dat ik schets staan de keuzes en waarden van de persoon centraal en wordt niets uitgesloten. Het autobiografisch narratief van de persoon en de bekrachtiging ervan zijn wat telt. Ik beschouw een persoon als autonoom zolang ze in staat is, of in staat zal zijn op enig moment in de toekomst, om haar eigen waarden en keuzes te bekrachtigen of verwerpen. In één woord: om zich te identificeren met haar zelf-narratief.

In het vierde hoofdstuk pas ik mijn narratieve kijk op persoon-zijn en autonomie toe op wilsverklaringen in de psychiatrie. Duidelijk wordt dat een narratieve benadering van persoon-zijn en autonomie in termen van actorschap een aantal problemen, die aan het gebruik van wilsverklaringen kleven, kan oplossen. Dit zijn mijn belangrijkste bevindingen:

(i) Mijn narratieve visie op zelf-constitutie verheldert de ervaringen van patiënten die lijden aan weerkerende psychoses, en van hun psychiaters. In geval van een ernstige psychose is de patiënt tijdelijk geen persoon meer. In minder ingrijpende gevallen wordt de psychotische persoon als afwijkend ervaren. Toch lijken er geen praktische redenen te zijn om te veronderstellen dat de psychotische persoon 'iemand anders' is. De psychotische ervaringen van de patiënt maken op een of andere manier deel uit van het levensomspannende proces van narratieve zelf-constitutie. Ondertussen rijst bij dergelijke ingrijpende persoonsveranderingen wel de vraag welk verhaal van die persoon kan worden aangeduid als het verhaal met de meeste autoriteit. Er zijn dan immers meerdere, uiteenlopende levensverhalen in omloop: dat van de niet-psychoticus, het verhaal van de pre-psychoticus en dat van de psychoticus. Het antwoord dat ik voorstel is om coherentie doorslaggevend te laten zijn. Het meest coherente verhaal heeft het grootste gezag. De interne samenhang van de (pre) psychotische verhalen is dikwijls serieus verzwakt en belangrijke delen van het niet-psychotische levensverhaal blijven daarin vaak geheel buiten beeld, terwijl het niet-psychotische zelf-narratief daarentegen soms een stuk consistentier is, maar altijd gespitst op het integreren van de (pre-)psychotische verhalen in het zelf-narratief. Maar bovenal is er bij (pre-)psychotische verhalen in de regel sprake van een gebrek aan realiteitszin (aan 'externe' samenhang). Tot op zekere hoogte stroken ze niet met de feiten

van de wereld en met de wijze waarop significante anderen het (pre-)psychotisch zelf-narratief interpreteren.

(ii) Mijn narratieve benadering van autonomie kan de notie 'precedente autonomie' invulling en betekenis geven en de morele rechtvaardiging van het gebruik van zelfbindingsverklaringen ondersteunen. Hoewel psychoses dikwijls een narratieve disruptie van identiteit tot gevolg hebben, zijn personen in staat om weerkerende psychoses te overbruggen. Dat doen ze door die psychoses deel te laten uitmaken van hun levensverhaal. En door zich prospectief te identificeren met hun toekomstige psychotische staat (patiënten die een zelfbindingsverklaring opstellen hebben de nodige ervaring opgedaan met het fenomeen psychose). Wat mijn narrativistische positie in het bijzonder relevant maakt voor het debat over wilsverklaringen is de aandacht voor identificatie door de persoon met een set levensomspannende, persistente en coherente 'identity conferring concerns' (waarmee waarden en belangen zijn bedoeld). Zulke entiteiten kunnen een ernstige psychose overleven. In mijn optiek staat autonomie ten dienste van dergelijke persisterende en coherente waarden en belangen; en respecteren we *in globo* het beste de autonomie van een persoon wanneer we voorrang geven aan het moment waarop die persoon het meest competent is tot het uitoefenen van haar autonomie. Op dat moment kan die persoon het beste bepalen welke gedragslijn haar autonomie beschermt over het geheel van haar leven bezien. Ik heb echter wel enige aarzeling over hetgeen Ronald Dworkin (1931) de 'integriteitswaarde' van iemands leven muntte. Meer dan Dworkin hou ik er rekening mee dat de criteria, waarmee wordt gewogen wat waardevol is en wat triviaal, nogal eens willen veranderen in de loop van een mensenleven. Het goede nieuws is dat in het geval van zelfbindingsverklaringen de levenslange houdbaarheid van waarden en

belangen niet van doorslaggevend belang is. Psychose is een episodisch fenomeen, en verkeerde inschattingen kunnen na afloop worden bijgesteld.

Voor wat betreft de morele beoordeling van zelfbindingsverklaringen kom ik tot de volgende slotsom: wanneer we het pri maat van respect voor andermans autonomie als startpunt kiezen en autonomie diachroon opvatten, dan kan het ten uitvoer leggen van zelfbindingsverklaringen moreel gerechtvaardigd worden. Eerst en vooral respecteren we daarbij de precedente autonome toestemming van de patiënt. Bovendien zal het in veel gevallen mogelijk zijn om daar respect voor de autonome instemming achteraf bij op te tellen. Niettemin zijn de morele kosten aanzienlijk. Voorrang geven aan ‘identiteit verlenende bekommernissen’ waar een patiënt op dat moment vaak niet meer om geeft, en op grond daarvan een relatief competente, protesterende persoon onder dwang behandelen kan beschadigend uitpakken; veel patiënten en hulpverleners ervaren zo’n interventie ook als traumatiserend en een aanslag op hun identiteit. Bovendien is het succes van dwangbehandeling onvoorspelbaar en deels een kwestie van geluk. Maar alles in aanmerking nemend mogen we verwachten dat korte schendingen van iemands autonomie door middel van dwangtoepassing, afgewisseld met lange periodes in gewenste condities na een ingreep onder dwang, per saldo de best mogelijke zelfsturing opleveren.

(iii) Hoewel het idee van zelfbinding een verleidelijke strategie mag lijken, denk ik dat het gebruik ervan begrensd is. Dit wordt duidelijk wanneer we wilsverklaringen van alzheimerpatiënten onder de loep nemen. Opnieuw blijkt mijn narratieve visie op personen en hun autonomie behulpzaam te zijn bij het verhelderen van de uiterst problematische autoriteit van wilsverklaringen in geval van ernstige dementie. Net als in gevallen van

wekerende psychose kan de episode van ernstige dementie, waarin geen eigen verhaal meer mogelijk is, al in een eerder levensstadium zijn opgenomen in het zelf-constituerend narratief van de persoon in kwestie, en kan de fase van dementie worden toegeschreven aan deze persoon die zich er vooruitziend over uitsprekt en mee identificeert. Maar gevallen van Alzheimer verschillen op een aantal punten ook grondig van gevallen van wekerende psychose. Primo, de narratieve ontwrichting van de identiteit is nu permanent en gaat veel dieper. De persoon verdwijnt. Secundo, identificatie vooraf met de incompetente staat van dementie is moeilijker dan identificatie met een periode van wekerende psychose, omdat de betreffende persoon geen eigen ervaring heeft met dementie. Tertio, er is geen gelegenheid tot bijstellen.

Ook bij de morele rechtvaardiging van wilsverklaringen als vehikel voor precedente autonomie loop ik tegen nieuwe vragen op in het geval van Alzheimer. Ik kies als voorbeeld wilsverklaringen waarin verzocht wordt om een ernstig demente patiënt niet te behandelen als hij een levensbedreigende maar behandelbare ziekte oploopt. Het eerste probleem is dat, vanuit mijn oogpunt, in de ernstige fase van Alzheimer de oorspronkelijke bindende kracht van precedente autonomie is uitgewerkt. Ik ga er immers van uit dat een persoon autonomie uitoefent zo lang hij in staat is, of in staat zal zijn op enig moment in de toekomst, om zich met zijn zelf-narratief te identificeren. Zo'n moment komt er, naar mijn weten, niet meer. Het gevolg daarvan is dat een wilsverklaring die vast wil houden aan eerder gemaakte bindende keuzes nu moet wedijveren met de morele verantwoordelijkheid om een op dat moment kwetsbare, incompetente patiënt geen schade toe te brengen. Aan een narrativistische uitleg van autonomie kan ik niet langer een beslissende morele rechtvaardiging ontleen voor het overklassen van de actuele belangen van de ernstig demente

post-persoon met de belangen van de persoon die aan de komst van Alzheimer voorafging. Het is reden geweest voor mij om elders te zoeken naar een extra argument. Tevergeefs.

Het tweede probleem is dat in het morele debat over wilsverklaringen steevast de individuele patiënt centraal wordt gesteld. Dat lijkt me een ernstig en uiteindelijk onoverkomelijk tekort. Het gaat hier niet langer om hachelijk ingrijpen met gebruik van dwang, maar over kwesties van leven en dood. Anderen moeten de wilsverklaring uitvoeren door af te zien van behandeling en de patiënt te laten sterven. De morele kosten van het doden van een medemens die zich redelijk behaaglijk voelt, niet lijdt en makkelijk kan worden gered, zijn hoog en doen wat mij betreft de balans doorslaan. Ik betoog dat we in zo'n geval mededogen moeten hebben, niet alleen met de incompetent patiënt maar ook met de kwetsbare zorgverlener. We kunnen niet van artsen verwachten dat ze zichzelf blootstellen aan zulke morele risico's. De morele kosten zijn, zover ik kan zien, in dit geval zo hoog dat ze het doel dat met de wilsverklaring wordt beoogd buiten bereik plaatsen. In geval iemand de diagnose Alzheimer meege-deeld krijgt is een wilsverklaring een ineffectieve methode om toekomstige zorg te organiseren. Voor wilsverklaringen is dan hooguit een ondersteunende rol weggelegd.

(iv) Voor personen met de diagnose Alzheimer, die zich geconfronteerd zien met een catalogus aan verschrikkingen, staan twee strategieën open als alternatieven voor een wilsverklaring. Eén mogelijkheid is dat ze accepteren dat autonomie niet langer een issue is in hun toekomstige incompetente toestand van ernstige dementie en dat ze afhankelijk zullen zijn van de hulp van anderen om beschermd te worden tegen letsel en krenking. Natuurlijk kunnen ze een wilsverklaring opstellen, maar nodig is het niet. Zodra de fase van ernstige dementie inzet beslissen an-

deren. De tweede mogelijkheid is dat ze zich preventief van het leven beroven nadat de diagnose Alzheimer is gesteld. Daarmee is niet gezegd dat het beter is om dood te zijn dan dement. Maar voor sommige mensen is autonomie, meester zijn over het eigen leven inclusief toekomstige verzorging, een gekoesterde waarde die onopgeefbaar is en waardevoller dan het leven zelf. Zij willen kost wat kost controle houden en voorkomen dat ze terechtkomen in wat ze beschouwen als een verward en verloren bestaan. Voor deze patiënten is *Romana mors*, sterven in eigen beheer in het zicht van Alzheimer, een moreel te rechtvaardigen daad van autonomie.

CURRICULUM VITAE

Patrick J. J. Delaere is geboren op 31 december 1953 te Izegem (B). In juli 1971 behaalde hij zijn Getuigschrift van Hoger Middelbaar Onderwijs (Afdeling Latijn-Grieks) aan het Sint-Jozefscollege te Izegem. Gedurende ongeveer 20 jaar was hij daarna in Nederland werkzaam in de antroposofische Jeugd-GGZ resp. als verpleegkundige, Afdelingshoofd, Hoofd Verzorging en Behandelingcoördinator. Naast werk en gezin studeerde hij van 1989 tot 1997 in deeltijd wijsbegeerte aan de Universiteit Utrecht. Vanaf 1999 is hij verbonden aan de Faculteit der Wijsbegeerte van de Erasmus Universiteit Rotterdam als onderzoeker en docent Praktische Filosofie, en sinds september 2010 ook als Opleidingsdirecteur.

© 2010 P.J.J. Delaere

Cover design Proefschriftmaken.nl || Printyourthesis.com

Printed by Proefschriftmaken.nl || Printyourthesis.com

Published by Uitgeverij BOXPress, Oisterwijk

ISBN 978-90-8891-215-3