Persons, Self-Conceptions *and* Self-Self Relations

Christiane E. Seidel

Persons, Self-Conceptions *and* Self-Self Relations

Christiane E. Seidel

ISBN-10: 90-9020349-4
ISBN-13: 978-90-9020349-2
Copyright© Christiane E. Seidel

No part of this book may be produced in any form, by print, photoprint, microfilm, or other means without permission of the author.

Printed by Haveka bv

Persons, Self-Conceptions and Self-Self Relations

Persoon-zijn, zelf-begrip en zelf-zelf-relaties

Thesis
to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus
Prof.dr. S.W.J. Lamberts
and in accordance with the decision of the Doctorate Board

The public defence shall be held on
Thursday March 23rd 2006 at 13.30 hrs

by

Christiane Elisabeth Seidel
born at Lübeck, Germany

Doctoral Committee

Promotors:
Prof. L.R. Baker, Ph.D.
Prof.dr. J. de Mul

Other members:
Prof.dr. L. van Bunge
Prof.dr. I.U. Mäki
Prof. D. Zahavi, ph.d., dr.phil.

Contents

Acknowledgements 9

Chapter 1: Introduction 11

1 About self-relations in a broad sense 12
2 How to investigate self-self relations 16
3 Other philosophers' views 19
4 Overview of the following chapters 22

Chapter 2: The relation of oneself to one's self 27

Part I – The notion of 'the self' 29

1 A first-person perspective – the fundamental element of one's self 29
2 A narrative self-conception – the structuring element of one's self 33
3 Making qualitative distinctions – the orientation-providing element of one's self 34

Part II – The concern with one's self, as a relation of oneself to one's self 39

4 Our concern with our rational agency – belonging to the fundamental element of one's self 40
5 Our concern with our autobiographical narrative – belonging to the structuring element of one's self 42
6 Our concern with our values – belonging to the orientation-providing element of one's self 44
7 The need for unity of one's self 47
8 Conclusion 48

Chapter 3: Practical irrationality from a self-related perspective 49

Part I – Conceiving of one's own actions (or of oneself) as irrational 50

1 Agents and persons 50
2 Two sorts of evaluative reasoning 53
3 Actions and action structures 55
4 Our evaluative agenda 57
5 Different ways of conceiving of one's own actions, regarding aspects of rational agency 60
6 Acting irrationally from a self-related perspective 63
7 Being irrational from a self-related perspective 66

Part II – Why acting irrationally or being irrational matters to us 68

8 Acting irrationally matters to us – the concern with the soundness of our will 69
9 Being irrational matters to us – the concern with coherence of our evaluative agenda 72
10 Concluding summary 76

Chapter 4: Wrongness from a self-related perspective 79

Part I – Conceiving of one's own actions (or of oneself) as wrong 80

1 Prerequisites of conceiving of one's own action as wrong 81
2 A person's purposes and a person's values 84
3 A first-personal conception of acting wrongly 90
4 Critical analysis of my view on conceiving of one's own action as wrong 96
5 A first-personal conception of one's own wrongness 99

Part II – Why wrongness of (actions of) ourselves matters to us 101

6 The overriding nature of the self-related conception of wrongness of action 102
7 Acting wrongly matters to us – the desire to be loyal to oneself 107
8 Wrongness of oneself matters to us – the desire to be worthy of self-esteem 112
9 Concluding summary 114

Chapter 5: Selfhood: Unity in changeability 117

1 One's sense of selfhood 118
2 One's sense of one's continued selfhood 121
3 One's self-conception as a rational agent 126
4 One's awareness of the changeability of one's self 129
5 Embedded self-reflection and the desire for unity of one's self 133
6 Explicit self-reflection and the desire for unity of one's self 138
7 Meta-reflection about oneself and the desire for unity of one's self 142
8 Conclusion 147

Chapter 6: Self-conception and the limits to dissociation 149

1 The extreme case of dissociation 150
2 A wide notion of dissociation 154
3 Self-conception 160
4 Dissociation of oneself from oneself: possibilities and impossibilities 174

Chapter 7: Conclusion: Personhood, models of self-conception, and regions of mattering 179

1 About the project, objections to the proposed account, and the rest of the chapter 179
2 Personhood, selfhood, and the first-person perspective 182
3 Self-self relations 185
4 Concealed first-person self-conceptions: the basis for mattering-relations to oneself qua self 186
5 New questions 192

References 199

Samenvatting (Summary in Dutch) 203

Curriculum vitae 211

Acknowledgements

I am grateful to Lynne Rudder Baker and Jos de Mul who read and commented on the manuscript. Particular thanks go to Theo van Willigenburg for hours of helpful discussion and his detailed comments on the material from which this book has sprung. I also thank Gerrit Glas for discussing with me the ideas for this project.

Papers based on earlier versions of various chapters of this book were presented at a number of conferences, and I am grateful to many people who have commented on those papers. The following list provides the titles of those papers and the conferences in question:

- 'Practical irrationality from a self-related perspective', presented at the Fifth International Conference of the German Society for Analytic Philosophy, held at the University of Bielefeld, Germany, 22-26 September 2003.
- 'Wrongness from a self-related perspective', presented at the symposium 'The Constitution View of Persons and Artifacts', held at the Technical University of Delft, The Netherlands, 20 October 2003.
- 'Conceiving of oneself as oneself', presented at the '2004 Joint Session of the Mind Association and the Aristotelian Society', held at the University of Kent at Canterbury, 9-12 July 2004.
- 'Conceiving of oneself as oneself', presented at the international conference 'Dimensions of Personhood', held at the University of Jyväskylä, Finland, 13-15 August 2004; published in: Heikki Ikäheimo, Jussi Kotkavirta, Arto Laitinen & Pessi Lyyra (eds.), *Personhood. Workshop Papers of the Conference 'Dimensions of Personhood'*, Publications in Philosophy, 68. Jyväskylä: University of Jyväskylä, 2004.
- 'Conceiving of oneself as oneself', presented at the '7th International Conference on Philosophy, Psychiatry and Psychology', held at the University of Heidelberg, 23-26 September 2004. An extended version of this paper has been accepted for publication in: Thomas Fuchs and Christoph Mundt (eds.), *Time, Memory and the Self*. Basel: Karger Publishers, 2006 (forthcoming).
- 'Believing to be a multiple – a case of self-knowledge?', presented at the postgraduate conference 'Rational Believing and Knowledge', held at the Free University of Amsterdam, 2-4 February 2005.
- 'We cannot but conceive of ourselves as ourselves', presented at the '2005 Joint Session of the Mind Association and the Aristotelian Society', held at the University of Manchester, 8-11 July 2005.

- ‘Knowing what you are doing but not knowing why – a dissociative experience of non-rational agency’, presented at the Fifth European Congress for Analytical Philosophy, held at the University of Lisbon, 27-31 August 2005.

Christiane E. Seidel
Zeist, The Netherlands
October 2005

Chapter 1
Introduction

It generally bothers us if we are not motivated by the motives that we want to be motivated by or if we cannot conceive of ourselves as the source of our actions or thoughts. And it characteristically matters to us if we are confronted with fragments of our life which we cannot piece together in a self-narrative, or if we, due to some of our actions, cannot justifiably conceive of ourselves anymore as the kind of person that we thought we were and still want to be. In the last two situations, creatures like us may even feel dissatisfied or deeply unhappy with themselves.

Some people may deny that all this matters to them. They may believe that they do not care whether they act rationally, whether their life unfolds as a narrative, or what kind of person they are. They may even claim to admire irrationality and an incoherent way of life. Such people cannot but entertain a different notion of rational action than I do – probably equating rational action with well-considered action or with action void of passion and emotion. They may moreover think that such 'mattering-relations' (as I will call them) only obtain when people have well-defined life-plans and regularly engage in critical self-evaluation. However, apart from exceptional (sometimes pathological) cases, people's self-regarding reactive attitudes, the structure of their reflections on their actions and motives for actions, and their explanations of their behaviour, goals and plans testify to the existence of these mattering-relations to themselves qua self.

These mattering-relations seem to be inextricably bound up with particular capabilities that are characteristic of persons and which we continually exercise – at least if these capabilities are not diminished by say extensive alcohol consumption, or temporarily or permanently lost due to brain damage. However, if those latter conditions obtain, a person lacks ways of thinking and feeling about herself that are characteristic of fully developed persons.

One of the capabilities characteristic of persons, namely the capability to conceive of oneself as oneself, is a prerequisite for all kinds of self-regarding thoughts and feelings that a person (qua person) may entertain and experience. My account therefore relies on a view of personhood which takes the capability to conceive of oneself as oneself as the defining characteristic of all persons. And it is this concept of person which forms the starting point of my analysis of our relations to ourselves qua self. My strategy, however, does not imply that I restrict myself to the analysis of highly reflective

phenomena of self-awareness. Rather, it will turn out that reflective acts of self-awareness are not intelligible without presupposing a pre-reflective sense of selfhood.

My book touches areas that are well known to philosophers of, for example, action or mind. But by consistently taking the first-person perspective as the thread running through my analysis, I hope to contribute to the philosophical discussion on personhood in a unique way.

In this introductory chapter, I first provide a distinction between reflective self-self relations – the main emphasis of this book – and other kinds of self-relations. I then clarify my approach to the subject and (in the third section) list those 'building blocks' of my account the inspiration for which I owe to other philosophers. Finally, I provide an overview of the rest of the chapters.

1 About self-relations in a broad sense

We are creatures that are capable of standing in various self-relations: we can reflect on our motives for our actions, entertain self-images, and form judgements regarding our actions or personality; we can experience negative or positive feelings regarding ourselves, find ourselves in various self-regarding attitudes of mind, exhibit reactive attitudes towards ourselves, and have self-regarding concerns. Some of these self-relations are highly reflective; they are established by ourselves, so to speak. Others are experienced as given and seem to be outside our direct control. Both categories of self-relations have in common, however, that we are reflectively aware of taking or experiencing a particular self-regarding position: whether we evaluate our former actions and, say, regard them as irrational, whether we engage in self-evaluation and come to realize that we care about things that are not worth caring about, or whether we feel sad or disappointed about ourselves, or even feel desperate about ourselves.

Actively taking a certain position towards oneself as well as (consciously) experiencing feelings or moods about oneself obviously requires self-conception, namely conception of oneself as oneself. Conception of oneself as oneself is in itself a self-relation. It is the fundamental relation of reflective self-awareness, since all other acts of reflective self-awareness presuppose that fundamental reflective self-relation. After all, if we did not conceive of ourselves as ourselves, we could not think about the kind of life we want to live or ask ourselves what kind of person we really are, and we could not blame ourselves for what we have done or feel dissatisfied with what we have accomplished. This book is concerned with these kinds of self-relations, that is to say, with self-relations that require self-conception. More precisely, this book is devoted to self-relations that we qua self-comprehending creatures entertain with ourselves qua self. I label them reflective self-self relations.

The self is generally taken to be an ambiguous notion, which some

philosophers may even want to get rid of.[1] So, what do I mean by relations to ourselves qua *self*? Well, roughly speaking, I take those 'things' that we – when establishing, entertaining or experiencing a certain self-relation – do not attribute to parts of our body while still attributing them to ourselves to be aspects of ourselves qua self. For example, when regarding our actions as wrong or mistaken, we do not attribute those actions to our hands, muscles, or brain states, but we attribute them to ourselves qua person. The same holds when we, for instance, regard our goals or plans as mistaken, or our motives as wrong; we do not attribute those goals, plans and motives to certain brain states, but we attribute them to ourselves in our entirety, so to speak. Take, as a last example, our promises. When we feel sad or disappointed about ourselves because we recurrently do not keep our promises, we do not attribute our promises to whatever part of our organism might be appropriate, but we still attribute our promises to ourselves. By contrast, when someone is dissatisfied with, for instance, the shape of her ears or the colour of her hair, that person, although she conceives of her ears as her ears and of her hair as her hair, attributes the shape of her ears or the colour of her hair not to herself qua person in her entirety but to herself qua body. Therefore, according to my classification, being dissatisfied with the shape of one's ears or the colour of one's hair are self-relations that that person qua self-comprehending creature entertains with herself qua body.[2] And these kinds of self-relations are not my concern here. My emphasis is on our relations to aspects of ourselves qua self – in the sense described above. The more general question about the notion of the self, which I introduced at the beginning of this paragraph, is obviously not answered by my

[1] See, for example, Eric T. Olson's essay 'There is no problem of the self' (Olson 1999).

[2] One reader objected that I seem to confuse the body-mind distinction with the part-whole distinction. In case another reader might fall prey to the same misreading (the rest of the readers may skip this note), I want to clarify that point and provide some background theory. My distinction is none of the two aforementioned distinctions, but the *body-person* distinction. That distinction has its basis in Lynne Rudder Baker's constitution view of persons (Baker 2000). According to that view, living human organisms (in short, bodies) that have the capability of a first-person perspective constitute persons. A person and her body (that is, the living human organism that constitutes that person) are one 'thing', although they are not identical. A person has certain properties owing to the fact that she is a person, others owing to the fact that she is constituted by a particular body. When attributing properties to oneself, one can either attribute them to oneself qua person or to oneself qua body – depending on whether one has those properties owing to one's personhood or owing to one's body. When we attribute certain 'things' to ourselves qua body, we typically distinguish between our body in its entirety (for example, feeling tired or full of energy) and a certain part of it (for example, toothache or colour-blindness). However, when attributing 'things' to ourselves qua person (for example, spontaneity, sensitivity to the feelings of others, a certain ambition, or belief in God), we always attribute them to ourselves in our entirety. Let me now turn to the body-mind distinction. I take the mind to be the set of one's psychological states. The notion of the mind is hence different from the notion of a person; after all, the person is the whole embodied thing, not just the set of her psychological states. To sum up, the body-person distinction is neither a body-mind nor a part-whole distinction.

explications so far. I will anyhow not discuss the numerous competing accounts of the concept of the self. My view on the self will emerge in due course.

I made clear that not all self-relations that we establish, entertain or experience are relations to ourselves qua self, but I have not yet clearly separated reflective self-self relations from *all* other types of self-relations. Reflective self-self relations, to recapitulate, are relations of oneself qua self-comprehending creature to (aspects of) oneself qua self – in contrast to oneself qua body. This statement gives rise to the question whether we can stand in self-relations that are *not* maintained by ourselves qua self-comprehending creature. I believe that we indeed (can) have positive or negative attitudes towards ourselves that we are not reflectively aware of (as Sigmund Freud famously argued). We may, for example, have repressed feelings about ourselves. Such feelings can reveal themselves in our actions, reactions to certain situations or interactions with others. And we might become aware of them by observing and reflecting on our own behaviour, or by means of psychotherapy or psychoanalysis. Having positive or negative attitudes towards oneself, say, feelings due to an inferiority complex, presupposes that a person is a self-comprehending creature, even if she is not aware of those beliefs and feelings. Only a self-comprehending creature – that is, a creature that has the capability to conceive of itself as itself – can form and entertain beliefs about itself, whether those self-regarding beliefs are readily available to that creature or not. I call self-relations that we are not reflectively aware of but which nonetheless require our ability of self-comprehension pre-reflective self-relations. A further distinction between pre-reflective relations to oneself qua self and to oneself qua body (a distinction that I made regarding reflective self-relations) does not make sense. After all, only if a pre-reflective self-relation became reflective – in other words, only if the person became reflectively aware of the (for example, repressed) beliefs and feelings about (aspects of) herself – could she attribute the involved aspect(s) of herself either to herself qua self or to herself qua body. We might therefore label that category of self-relations pre-reflective 'self-self or self-body' relations.

The three categories of self-relations discussed so far – reflective self-self relations, reflective self-body relations, and pre-reflective 'self-self or self-body' relations – are established, maintained or experienced by creatures that are self-comprehending (that is to say, that have the capability to conceive of themselves as themselves), although they may not be reflectively aware of those self-relations (at that moment). Not all self-relations, however, require a creature's capability of reflective self-awareness. We might also talk about self-relatedness on the level of the body, that is to say, on the level of the living human organism.[3] More specifically, some kinds of self-relatedness are not only *not* established or maintained by the human being qua self-comprehending creature but do not even require that the creature should have the capability of

[3] By the way, not only human organisms establish or entertain such self-relatedness.

reflective awareness. Think, for example, of a creature's natural instinct to save its own life, of our body's compensatory movements if the body threatens to get out of balance[4], or of an organism's activation of its recuperative powers if harmful substances penetrated into that organism. Let us label that type of self-relations body-body relations.

Table 1.1 gives an overview of the various types of self-relations and provides an indication regarding their discussion in the context of this book.

Table 1.1: Overview of the various types of self-relations

Reflective self-self relation

Characterization: Relation of oneself qua self-comprehending creature to (aspects of) oneself qua self:

- Person establishes, entertains, or experiences a self-relation – for example, she thinks or feels about (an aspect of) herself in a certain way – and she is reflectively aware of that attitude towards (that aspect of) herself.
- The person attributes the aspect of herself which her thinking or feeling relates to, to herself qua person in her entirety.

Discussion in this book: Main emphasis of this book.

Reflective self-body relation

Characterization: Relation of oneself qua self-comprehending creature to (aspects of) oneself qua body:

- Person establishes, entertains, or experiences a self-relation – for example, she thinks or feels about (an aspect of) herself in a certain way – and she is reflectively aware of that attitude towards (that aspect of) herself.
- The person attributes the aspect of herself which her thinking or feeling relates to, to herself qua body or organism.

Discussion in this book: Outside the scope of this book.

Pre-reflective 'self-self or self-body' relation

Characterization: Relation of oneself qua self-experiencing creature to (aspects of) oneself:

- Person entertains a certain self-relation – for example, she has certain self-regarding feelings or beliefs – but she is not reflectively aware of the involved self-regarding feelings or beliefs.

Discussion in this book: Chapters 5 and 6 touch on this topic.

Body-body relation

Characterization: Relation of oneself qua organism to oneself:

- Human organism entertains or experiences a self-relation.

Discussion in this book: Outside the scope of this book.

---

2 How to investigate self-self relations

2.1 Analytical, structural, descriptive – not historical or empirical

The approach I take in trying to understand self-self relations – the relations of a person qua self-comprehending creature to herself qua self – is to consider the structure of those relations and the dispositions and desires that contribute to those relations. Understanding what those relations are differs from understanding how people came to be that way, either as an entity or as an individual. So, my considerations are primarily analytical and structural rather than historical.[5]

To explore self-self relations, I typically dig into the subject by considering commonly held pre-philosophical notions and by analysing phenomena that are our shared experience and all of which exhibit a self-self relation. To name some of those notions and phenomena: our regarding ourselves qua self as changeable, our conception of ourselves qua self as something constant, our regarding ourselves as irrational or wrong in certain situations, our confusion if we do not know why we are doing what we are doing, or our distancing ourselves from certain of our desires or actions. I then try to come to grips with the phenomenon or pre-philosophical notion in question by a thorough analysis that takes the concept of person that I favour as a starting point and applies that concept and its implications.[6] Not only does this approach illuminate certain phenomena and pre-philosophical notions that exhibit various kinds of self-self relations, it also reveals the structures and dispositions that constitute those self-self relations. At the same time, the considerations I put forward on these matters contribute to a deeper understanding of the concept of personhood.

My explorations are often descriptive without being empirical. I am not asking what kind of self-self relations people at large have. That question is an empirical question which I as a philosopher have no special competence to answer. My point is conceptual: people have certain self-self relations owing to capabilities by virtue of which they count as (fully developed[7]) persons. The structures and dispositions that constitute the various self-self relations are inextricably bound up with those capabilities.

---

[4] "Humans [but not only humans – C.S.] have a "sixth sense" called proprioception, which is the sensory feedback they constantly receive from their muscles, joints, and skin, signalling the position of their bodies and limbs. Without knowing it, we [qua human creature – C.S.] constantly monitor this feedback and make adjustments to our bodies; for example, when we lift our left arm, we subtly shift some weight to the right side of our bodies to maintain our balance. If we didn't, we would list dangerously to one side." (Wilson 2002, p. 19)

[5] Hence, I am *not* exploring the developmental stages of personhood – neither from an evolutionary nor from a psychological perspective.

[6] For my view on personhood, see especially Chapter 2 (The relation of oneself to one's self).

[7] Some capabilities are characteristic of fully developed persons although not being strictly required for a person to count as a person; see Chapter 2 (The relation of oneself to one's self).

I should furthermore emphasize that I am not concerned with the ethical question of what kind of self-self relations people ought to have or what about themselves qua self ought to matter to them. My concern is the self-self relations that people cannot but have, given particular characteristic capabilities.

2.2 Taking the first-person perspective seriously

My analysis of our self-self relations will take the first-person perspective seriously, for I believe that we can illuminate and reveal the structures and dispositions inherent to our self-self relations by analysing the various ways in which we conceive of, for instance, our actions, purposes, goals, motives, intentions, beliefs, values, character traits, life, or biography. We feel, for example, slightly confused if we find ourselves doing something without knowing why we are doing what we are doing. This phenomenon will, for instance, turn out to be evidence of our disposition to comprehend ourselves as rational agents, that is to say, as agents that act based on reasons.[8]

By taking the first-person perspective seriously I neither ignore nor deny people's capability to take an objectifying view on themselves. An objectifying view is not in conflict with the notion of a first-person perspective that I adhere to.[9] Having a first-person perspective is not just *having* a perspective but being reflectively aware of having *one's own* perspective and of the fact that one's own perspective differs from the perspective of others. It is therefore characteristic of creatures that have the capability of a first-person perspective – and it is necessarily entailed by that capability – that such creatures can take a detached stance towards themselves and reflect on, for example, their motives for action or their personality. And taking an objectifying view on oneself is, after all, nothing more than that: taking a detached stance towards oneself, reflecting on a certain aspect of oneself, thereby taking certain standards or the perspective of others into account, and eventually taking a position towards that aspect of oneself. In short, an objectifying view on oneself still is a self-conception, and every self-conception is evidence of the capability of a first-person perspective.

Some readers may feel inclined to raise a general objection to the project of this book, namely that our relation to ourselves qua self cannot be analysed in

---

[8] See Chapter 6 (Self-conception and the limits to dissociation).

[9] Having a first-person perspective, in this sense, is synonymous with conceiving of oneself as oneself. "To be able to conceive of oneself as oneself is to be able to conceive of oneself independently of a name, or description, or third-person demonstrative. It is to be able to conceptualize the distinction between itself and everything else there is. It is not just to have thoughts expressible by means of 'I', but also to conceive of oneself as the bearer of those thoughts. […] [M]erely having a perspective, or a subjective point of view, is not enough […]. One must also be able to conceive of oneself as having a perspective, or a subjective point of view." (Baker 2000, p. 64)

isolation from our relation to others and to the world. They may argue that the physical, emotional, and cognitive interaction with others plays a major role in the development of the self[10] and that, even in a mature stage, our self and the ways we think and feel about ourselves are still shaped by the feedback of our colleagues, friends and loved ones. I will not challenge that view. My concern with our self-self relations does not imply that I propose a solipsistic view on personhood. We could, for example, not blame ourselves for not keeping certain promises or for not being sensitive to the feelings of others if we had never had interactions with our fellow human beings. In that case, how could we know the concepts of promise and sensitivity? And I could never entertain the thought "I wonder if this book will ever be published" if I had never had interactions with the 'outside' world; for I would lack the concepts of book and publishing.[11] I thus do not propose a view that diminishes the importance of other- and world-relatedness and that ignores the connections and dependencies between self-relatedness, other-relatedness, and world-relatedness. Yet, to analyse the structure, dispositions and desires that are inextricably bound up with our self-self relations, I have to place these relations in the foreground and analyse the corresponding phenomena.

2.3 The quest to find an adequate account of personhood

Persons, obviously, think and feel in different ways about themselves qua self. I will argue that the ways in which we think and feel are not merely contingent; that is to say, they not only depend on, for example, our actions and experiences, our character and biography, or the reaction of others to our own behaviour. Our self-regarding thinking and feeling also has certain a priori structures and involves particular self-regarding dispositions, inclinations or desires. Here are two examples. Firstly, we cannot but think of ourselves as generally acting based on reasons. We are confused if we find ourselves doing something without knowing why we are doing what we are doing. Secondly, we cannot but think that *we* (qua person) – not forces *in* us, so to speak – generally decide what we are doing. It would bother us if we regularly or frequently did not want to be motivated by a certain desire to act in a certain way and precisely performed that action.

If I am right, certain things concerning our *self* characteristically matter to us. To avoid misunderstanding, I am not thinking of the fact that we have self-interested concerns – for example, the self-interested concern to survive. The concerns that I have in mind are more concealed, so to speak. They exhibit themselves in 'models' of conceiving of ourselves which, for instance, become explicit in regarding certain of our actions as non-rational or irrational, or as

---

[10] See, for example, Butterworth 1999 and Butterworth 2000.

[11] Lynne Rudder Baker uses a similar argumentation to defend her theory against the objection that her account is Cartesian. See Baker 2004, p. 133.

not in tune with our character. An adequate theory of personhood should therefore not only show which capabilities are characteristic of personhood but also account for the self-regarding dispositions, desires, etcetera that persons characteristically have and which constitute particular mattering-relations to themselves qua self.

3 Other philosophers' views

3.1 The areas concerned are well-known to philosophers

My investigations of self-self relations bear upon a diversity of areas that are familiar to philosophers with various specializations – whether they are concerned with personhood, the self, personal identity, or rationality, whether they are specialized in the philosophy of action, mind, psychology, or psychiatry, or whether they are ethicists. My approach, however, illuminates connections that otherwise threaten to be overlooked or underemphasized.

The phenomenon of dissociation, for instance, which is primarily discussed in the philosophy of psychiatry[12], turns out to have non-pathological forms that are our shared experience. Think of the phenomenon of distancing oneself from certain of one's desires or actions, and think of one's lack of first-person access to one's perceptions and actions when driving on autopilot. By analysing dissociation in the light of a general account of self-conceptions – instead of, for example, in the context of a theory about multiple selves or in the context of an account of free action – specific questions appear in the spot-lights. Answering these questions can illuminate our relation to ourselves qua self: Which aspects of herself can be subject to a person's dissociation? Which dissociative phenomena can count as dissociative acts of a person qua person? Can a person dissociate herself from herself qua self?[13]

Take as a second example the different ways in which we sometimes judge our actions – for instance, as rational, irrational, non-rational, or wrong – and the different feelings that can precede, accompany or follow, and inform our judgements – like satisfaction, dissatisfaction, regret, confusion, or contrition. Proceeding from the concept of person and the various ways in which people may regard their actions raises other questions than does discussing these phenomena in the context of theories of action. Think of the following 'mattering-questions': Why does acting irrationally bother us? Why does acting non-rationally bother us? Why does acting wrongly bother us? Being able to answer these questions is indispensable for an account of our relation to ourselves qua self.[14]

---

[12] See, for example, Braude 2004, Radden 1999, or Stephens and Graham 2000.
[13] See Chapter 6 (Self-conception and the limits to dissociation).
[14] See Chapter 3 (Practical irrationality from a self-related perspective) and Chapter 4 (Wrongness from a self-related perspective).

I should mention that, although not being a psychiatrist, I sometimes refer to pathological cases of, for example, Manic Depressive Illness or Dissociative Identity Disorder. But I am not interested in the mental illness qua illness, I am interested in what it reveals about the structure, processes, dispositions and desires that are bound up with our relation to ourselves qua self.

3.2 Inspiration drawn from other philosophers' views

Questions concerning our relation to ourselves necessarily take the concept of person as starting point. After all, we are persons. How could one ever propose an account of our relation to ourselves without clarifying the concept of person or presupposing a view on the kind of entities persons are? The concept of person, however, is subject to unremitting philosophical debate. Different contemporary philosophers have ended up with rival theories.[15]

The theories differ, for instance, in identifying 'person' with either 'body', 'human being', or 'self'. And they also differ (partly because of their diverging answers to the preceding question) on subsequent questions, for example: firstly, whether personal identity over time consists in physical or psychological continuity, or whether it has to be thought of in terms of characterization instead of re-identification; secondly, which capabilities are characteristic of personhood; or, thirdly, whether selfhood requires reflective self-awareness. Furthermore, there are debates in other areas of philosophy – for instance, regarding rational agency, values, self-knowledge, or self-consciousness – that intertwine with a clarification of the concept of person.

My project is *not* to compare and categorize rival theories of the concepts of person and self, or to defend one or some against the others. None of the existing theories of personhood or selfhood, if I am right, cover the scope of my project of a person's relation to herself qua self and consistently take the first-person perspective as the thread running through the analysis. I will therefore unfold my account of personhood and self-self relations. That account draws its inspiration from theories of various contemporary philosophers of person, mind or action, from works of philosophers of psychiatry, and from works of some psychologists and psychiatrists.[16] In some

---

[15] There are exceptions. For example, Eric T. Olson argues that such a debate is unnecessary: "The word 'person' is well enough understood for there to be philosophical problems about people." See Olson 1999, p. 53.

[16] I mean the following philosophers of person, mind, or action: Lynne Rudder Baker, Michael E. Bratman, Owen J. Flanagan, Manfred Frank, Harry Frankfurt, Shaun Gallagher, Christine M. Korsgaard, Raymond Martin, Thomas Nagel, Julian Nida-Rümelin, Eric T. Olson, Derek Parfit, Philip Pettit, Paul Ricoeur, T. M. Scanlon, Sydney Shoemaker, Charles Taylor, Marya Schechtman, Michael Smith, Galen Strawson, Ernst Tugendhat, J. David Velleman and Dan Zahavi; the following philosophers of psychiatry: Owen J. Flanagan, George Graham, Jennifer Radden and G. Lynn Stephens; the psychologists John Barresi, George Butterworth and Timothy D. Wilson; and the psychiatrists Peter D. Kramer and Kay Redfield Jamison.

cases, my account owes the inspiration for certain key notions or 'building blocks' to those philosophers. By explicitly mentioning those items and their source (in Table 1.2), I aim to give readers who are familiar with those areas of philosophy a first impression of the shape of my account.

Table 1.2 shows a list of those 'building blocks' of my account the inspiration for which is drawn from the views of other philosophers; it refers to the main source of each of those 'building blocks' and lists the chapters in which those 'building blocks' appear.

Table 1.2: Building blocks of my account

First-person perspective as defining characteristic of persons
Main source: Lynne Rudder Baker's theory of persons (Baker 2000)
Chapters in this book: All chapters, especially Chapter 2

Distinction between weak and strong evaluation
Main source: Charles Taylor's view on persons as strong evaluators (Taylor 1985)
Chapters in this book: Chapters 2, 3, 4, and 5

Narrative self-conception
Main source: Marya Schechtman's narrative self-constitution view (Schechtman 1996)
Chapters in this book: Chapters 2, 3, and 5

Notion of caring (caring about being a certain kind of person, caring about one's values)
Main source: Harry Frankfurt's view on caring as being not under a person's immediate voluntary control (Frankfurt 1982)
Chapters in this book: Chapters 4 and 5

Distinction between agency and rational agency
Main source: Lynne Rudder Baker's view on (rational) agency (Baker 2000)
Chapters in this book: Chapters 2, 3, and 4

Rational agency; rational action (taken as acting based on reasons)
Main source: Primarily Lynne Rudder Baker's, secondarily J. David Velleman's view on rational action (Baker 2000, Velleman 1992)
Chapters in this book: Chapters 2, 3, 4, 5, and 6

Notion of structural reasoning
Main source: Julian Nida-Rümelin's distinction between punctual optimalization and structural rationality (Nida-Rümelin 2001)
Chapters in this book: Chapter 3

Notion of commitment to one's purposes and values
Main source: Christine M. Korsgaard's view on committing oneself to take the means to one's ends (Korsgaard 1997)
Chapters in this book: Chapters 3 and 4

Pre-reflective self-awareness
Main source: Dan Zahavi and Josef Parnas's view on the 'givenness' of the (experiential) self (Zahavi and Parnas 1999)
Chapters in this book: Chapters 5 and 6

Pre-reflective sense of identity
Main source: Manfred Frank's reference to the notion of 'Gefühl von Identität' (Frank 2002)
Chapters in this book: Chapter 5

Pre-reflective sense of (continued) selfhood
Main source: Shaun Gallagher and Anthony J. Marcel's view on ecological self-awareness (Gallagher and Marcel 1999)
Chapters in this book: Chapters 5 and 6

Distinction between first-person and third-person self-conceptions
Main source: Ernst Tugendhat's view on unmediated and mediated self-knowledge (Tugendhat 1979)
Chapters in this book: Chapter 6

Distinction between experiencing oneself as subject and as source of, for instance, one's thoughts
Main source: G. Lynn Stephen and George Graham's, and Jennifer Radden's view on ego-alien experiences (Stephens and Graham 2000, Radden 1999)
Chapters in this book: Chapter 6

I refrain here from presenting a second table listing all other key notions that make up my account. That would merely serve matters of completeness without being informative for the reader, as those notions require explanation and argumentation.

4 Overview of the following chapters

4.1 Developmental stages of the presented account

My ideas concerning an adequate view of person- and selfhood – that is to say, a view that accounts for our relation to ourselves qua self and reveals its structures and concealed dispositions, inclinations or desires – have

developed gradually. Retrospectively, I can distinguish five stages in the development of my account. These developmental stages are still recognizable in the structure of this book and knowledge of them might help readers to orientate themselves within the presented framework of concepts and distinctions.

I first develop a view on the structure of person- and selfhood, addressing the *reflective aspects* of person- and selfhood. This account of person- and selfhood identifies three properties that are constitutive of fully developed person- and selfhood: the property of a first-person perspective and the (non-fundamental) properties of a narrative self-conception and of making qualitative distinctions and applying them as self-regarding standards. The three properties turn out to be the basis of characteristic self-self relations. As the self-regarding reactive attitudes in question are more or less directly connected to our actions, I, in a second stage, provide an analysis of our feelings and thoughts concerning our actions. The resulting account describes the different functions of purposes and values in our evaluative reasoning, and identifies a set of characteristic attitudes concerning our actions and ourselves. Some of those self-regarding attitudes belong to the scope of the property of a first-person perspective; others also require the (non-fundamental) properties of a narrative self-conception or of applying self-regarding qualitative standards.

After having investigated the *reflective* aspects of person- and selfhood in considerable depth, it becomes clear that the account developed so far does not account for the 'givenness' of the self. I therefore then examine the notion of a *pre-reflective* sense of selfhood. The proposed account distinguishes between fundamental self-awareness and a pre-reflective sense of (continued) selfhood. The concept of a pre-reflective sense of (continued) selfhood turns out not only to account for the 'givenness' of the self, but probably to provide the ultimate ground for a person's implicit desire for unity of her self, which expresses itself in the various self-regarding reactive attitudes. The results of this third developmental stage of my account reinforce my thesis that a person cannot but conceive of herself as herself. That leads to an investigation of the limits to a person's dissociation from herself qua self. My account of person- and selfhood, which by now covers reflective as well as pre-reflective components, now deals with pathological self-conceptions and experiences, like the conception of being a 'multiple' or of having *alters* (alternate selves), or the experience of not being the source of one's actions or thoughts. The analysis of dissociations turns out to require a distinction between first- and third-person self-conceptions.[17] And it becomes clear that there are various kinds of non-pathological dissociative acts and experiences that are our shared experience. In the fifth and last stage of the development of my account, I finally pull together the various

[17] Both, first- and third-person self-conceptions, are obviously (after all, they are *self*-conceptions), entertained from a first-personal point of view.

conclusions and draw further implications.

Next to knowledge of the developmental stages of my account, it might be helpful to know that some aspects of my account will be addressed in more detail than others. A first example is the property of having a narrative self-conception. It is less fully discussed than the property of applying self-regarding qualitative standards and the typical reflexive thoughts and feelings that belong to the domain of that property. A second example is the constitution view of persons. That view is only briefly explained. Take as a final example the property of a first-person perspective. The *concept* of a first-person perspective – although it runs through all of the chapters – is less thoroughly discussed than are the *implications* of the property of having a first-person perspective.

There are two main reasons for the different degrees of elaboration of the various aspects of my account. Firstly, a particular aspect may be simply presupposed – for example, Lynne Rudder Baker's constitution view of persons. Secondly, I may – rightly or wrongly – believe that a certain aspect needs less elaboration, for instance, because I take it to be sufficiently discussed by the philosopher to whom I owe the inspiration for it – for example, Marya Schechtman's elaboration of the notion of a narrative self-conception. And, needless to say, an in-depth argument for all aspects of my account of personhood, selfhood, and self-self relations would not require one book but a whole series.

4.2 How to read this book

This book lends itself for two reading strategies. It can, of course, be read in its entirety, in the sequence of the chapters. Reading from cover to cover, obviously, is the most appropriate way to get a *complete* overview of my account of personhood and our relation to our 'self'. This book can, however, also be read selectively: one can read one or some of its chapters, in whatever sequence one prefers. That approach is possible, since the various chapters – with the exception of this introductory chapter and the concluding chapter – were written as independent papers. The reader may hence discover some redundancy. But because the context is different each time, I believe that the redundancies will have an illuminating effect. And, of course, I hope that the selective reader will become interested to such an extent that she will eventually read the whole book.

4.3 The central claims and assumptions of the following chapters

Brief synopses of the subsequent chapters are given in Table 1.3.

Table 1.3: Descriptions of the subsequent chapters

Chapter 2 ***The relation of oneself to one's self*** This chapter presents a structural account of the self and considers the mattering-relation that a person has to her self. The account distinguishes three components of that mattering-relation and identifies them as separate regions of mattering with different structures. The chapter rests on the following central assumption: Three properties jointly constitute a (mature) self: having a first-person perspective (which is the fundamental property), having a narrative self-conception, and making qualitative distinctions.
Chapter 3 ***Practical irrationality from a self-related perspective*** This chapter studies two 'mattering-phenomena': regarding one's action as irrational and regarding oneself as irrational. The claim is that our own irrationalities not only bother us because irrationality hinders the accomplishment of our aims, but also disturb us more intrinsically. The chapter rests on the following central assumption: Actions are self-conceived as rational, irrational, rational-and-irrational, non-rational, mistaken, or wrong because of how the action relates to a possible process of evaluative reasoning.
Chapter 4 ***Wrongness from a self-related perspective*** This chapter studies two 'mattering-phenomena': regarding one's action as wrong and regarding oneself as wrong. The claim is that conceiving of one's action as wrong overrides other possible ways of conceiving of one's action. The chapter rests on the following central assumption: Purposes and values – by relating to two different ways of reflecting on one's actions – play essentially different roles in our evaluative reasoning.
Chapter 5 ***Selfhood: Unity in changeability*** This chapter considers the tension between one's conceiving of oneself as one and the same self over time and one's reflective awareness of the changeability and instability of one's self. The claim is that certain characteristic ways of thinking and feeling about ourselves reveal our desire for unity of our self. The chapter rests on the following central assumption: Personhood – by requiring reflective self-awareness – presupposes a pre-reflective sense of selfhood, or rather a pre-reflective sense of continued selfhood.

Chapter 6 ***Self-conception and the limits to dissociation***

This chapter analyses various phenomena of dissociation in the light of an account of self-conceptions. The claim is that a person's dissociation from her *self* is impossible on conceptual grounds.

The chapter rests on the following central assumption: We have to distinguish between first-person and third-person self-conceptions; the latter, in contrast to the former, require conceptualization and symbolization of oneself qua self.

Chapter 7 ***Conclusion: Personhood, models of self-conception, and regions of mattering***

This chapter draws connections among the preceding chapters; it puts the emphasis on those aspects of the proposed view of person- and selfhood that can be expected to be most controversial. The claim is that the three properties that are constitutive of personhood (in its fullest sense) constitute separate regions of mattering with different structures and dispositions; in this way they are the basis for certain self-relations that are characteristic of persons.

The chapter rests on the following central assumption: The three properties that are constitutive of personhood (in its fullest sense) include certain 'models of self-conception'.

I do not pretend to have analysed all phenomena of self-self relations. But I hope that my analysis uncovers structures, inclinations, dispositions or desires that are concealed in certain characteristic ways of thinking and feeling about ourselves. In that case, this book should contribute to a deeper understanding of person- and selfhood. Needless to say, my account is still open to new perspectives.

Chapter 2
The relation of oneself to one's self

In everyday language, we find expressions which show that 'our self' matters to us. Such expressions can, for example, be found in the autobiographical book *An Unquiet Mind*, by Kay Redfield Jamison[1], describing her life with manic-depressive illness.

> […] I was confused and frightened and terribly shattered in all of my notions of myself; […]
> […] I tend to compare my current self with the best I have been […]
> After each of my violent psychotic episodes, I had to try and reconcile my notion of myself […]
> […] I despaired of ever returning to my normal self.[2]

My question is whether the pretheoretically presupposed mattering-relation of oneself to one's self, which is expressed in utterances like the above, is supported by a philosophical notion. And if so, how can we understand that relation?

In my view, the mattering-relation of oneself to oneself qua self belongs inherently to the phenomenon of 'the self'. I will split the concept of the self into three, and I will argue that when we understand the distinctions involved, we recognize three self-regarding concerns: concern with[3] one's rational agency, with one's autobiographical narrative, and one's values; and all three concerns exhibit a need for unity of one's self.

The ingredients of the account of the self presented here have their origins in theories of Lynne Rudder Baker, Marya Schechtman and Charles Taylor, but, if I am right, the proposed combination of certain elements of their theories produces a notion of the self that elucidates the full structure of selfhood. By focusing on structure and on the functional role of elements with respect to the whole, the notion of the self loses its vagueness. My structural account makes intelligible that one's self necessarily matters to oneself. To avoid misunderstanding, I should stress that such a structural account does not

---

[1] Jamison 1997.

[2] Jamison 1997, pp. 85, 92, 121, and 146.

[3] I deliberately use the expression 'concern with' instead of 'concern for', since the three self-regarding concerns are not – or rather, not primarily – *conscious* worries about one's rational agency, autobiographical narrative or values. The three 'things' simply matter to us, whether we are explicitly concerned about them or not.

presuppose or entail claims about the specific biological or social requirements for the development and sustainability of (certain elements of) the self. Neither does it presuppose or entail claims about the dependence of a particular person's 'content of the self' on physical, psychological or social factors. My interest is strictly conceptual: I try to make explicit what is implicit in our use of the concept of a self and in our idea that 'our self' matters to us. In trying to make explicit what is implicit, I will focus on revealing the distinctions, connections and relations that provide for conceptual structure.

My view can be summarized in the following five statements:

(S1) One's self is jointly constituted by three properties:
- having a first-person perspective;
- having a narrative self-conception;
- making qualitative distinctions.

(S2) The three properties differ in their role with respect to the self:
- a first-person perspective is fundamental for being a self;
- a narrative self-conception and the property of making qualitative distinctions are not necessary for being a self, but characteristic of *our* selves.

(S3) Each of the three properties constitutes an identity of one's self:
- a first-person perspective constitutes one's neutral identity;
- a narrative self-conception constitutes one's narrative identity;
- the property of making qualitative distinctions constitutes one's ethical identity.

(S4) Each of the three properties is the basis for a specific concern with one's self:
- a first-person perspective is the basis for concern with one's rational agency;
- a narrative self-conception is the basis for concern with one's autobiographical narrative;
- the property of making qualitative distinctions is the basis for concern with one's values.

(S5) The three concerns exhibit – by means of the normative criterion that we apply in the respective concern – our need for unity of our self:
- our concern with our rational agency exhibits our need for consistency of our self;
- our concern with our autobiographical narrative exhibits our need for coherence of our self;
- our concern with our values exhibits our need for integrity of our self.

I will begin by offering a characterization of the notion of the self (Part I), which will also lead to some reflection about the notions of agent, rational agency, person and personal identity, and then take that characterization as the basis for an account of the relation of oneself to oneself as a relation of concern with one's self (Part II).

Part I – The notion of 'the self'

INTRODUCTION

One's self is the 'product' of three properties: having a first-person perspective, having a narrative self-conception, and making qualitative distinctions. Because of a first-person perspective we are selves; because of a narrative self-conception we conceive of our life as a narrative and of our self as the protagonist of that narrative; and because of the property of making qualitative distinctions we have values and conceive of ourselves as having these values. Our self is the product of these three properties, and at the same time, because of these properties, we ourselves are the producers of our self.

1 A first-person perspective – the fundamental element of one's self

ONLY BEINGS WITH A FIRST-PERSON PERSPECTIVE ARE SELVES

To have a first-person perspective means much more than merely *having* a perspective. A being with a first-person perspective not only acts from its own perspective, but it is aware of the fact that it has its own perspective and that this perspective differs from the perspective of others. Hence, such a being thinks of itself as the subject of its behaviour, thoughts, feelings and experiences; in other words, it thinks of itself as itself. And, because having a first-person perspective means conceiving of oneself as oneself, a being with a first-person perspective can think of itself qua self in the future and in other circumstances. In short, beings with a first-person perspective can conceptualize their own perspective.[4]

A being with a first-person perspective conceives of itself as itself; it conceives of itself as a subject that is aware of itself. Such a being *is* a self for itself. Or, in other words, such a being exists as a self for itself. Conversely, one can say: A being that exists as a self for itself, conceives of itself as itself. A being that conceives of itself as itself conceptualizes its own perspective. And a being which conceptualizes its own perspective is a being with a first-person perspective.

I call a property which the existence of something is dependent on fundamental for the thing in question.[5] And because our existence as self is dependent on having a first-person perspective, a first-person perspective is fundamental for being a self.

---

[4] I owe my notion of a first-person perspective to Lynne Rudder Baker; see Baker 2000, especially, pp. 21 and 60-64.

[5] Baker does not use the term 'fundamental property' but the term 'essential property', see Baker 2000, pp. 35-39.

WE ARE HUMAN BEINGS WITH A FIRST-PERSON PERSPECTIVE

The question is: Who are the 'we' we are talking about? – Biologically we are human beings, animals of the species *homo sapiens*; but not all human beings have a first-person perspective. Infants, for example, do not (yet) think of themselves as themselves, and patients in a late stage of Alzheimer's disease do not conceive of themselves as themselves (anymore). – We are human beings with a first-person perspective, and because we have a first-person perspective, we are selves.

Let us define beings with a first-person perspective, in other words, beings that are selves, as *persons*. Using this definition we can say: we are human persons. By the way, if there were beings of a non-human species with a first-person perspective, then these beings would be persons, and everything that I say about beings with a first-person perspective would apply to them as well, at least as far as it is dependent on a first-person perspective only.

A FIRST-PERSON PERSPECTIVE CONSTITUTES AN IDENTITY OF ONE'S SELF, WHICH IS NOTHING MORE THAN 'I AM I' IN THE MOST ABSTRACT SENSE

Our existence as self implies our certainty of 'being-self'. Because of our first-person perspective we never doubt our own existence. We cannot doubt that we are the ones that we conceive of as ourselves. After all, we do not identify ourselves before we conceive of ourselves as ourselves; indeed, we do not identify (or mis-identify) ourselves at all.[6] One simply *is* for oneself as self. Our 'being-self' is a certainty for ourselves. This certainty can only disappear together with our 'being-self', because 'being-self' implies the certainty of 'being-self'.

Because of a first-person perspective one's self appears to oneself as the certainty of 'being-self', and by this as something constant that cannot be specified any further. I call this identity our *neutral identity*[7], because a notion of the self as 'being-self' is the notion of a 'content-free' self, so to speak.

BEINGS WITH A FIRST-PERSON PERSPECTIVE ARE CAPABLE OF (WEAK) RATIONAL AGENCY

What are rational agents? Rational agents are first of all *agents*, in other words, beings that do things. 'Doing things'[8] means that a subject has desires, beliefs and intentions, and that certain events would not have happened if the subject in question had not had certain desires, beliefs and intentions.

---

[6] See Baker 2000, pp. 136-137.

[7] The term 'neutral identity' is derived from Charles Taylor. Taylor writes in Taylor 1989, p. 49: "This is what I call the 'punctual' or 'neutral' self, – 'punctual', because the self is defined in abstraction from any constitutive concerns […] Its only constitutive property is self-awareness."

[8] This notion of 'doing things' is borrowed from Lynne Rudder Baker; see Baker 2000, pp. 149-150.

What sorts of beings 'do things' in this sense? Beings that 'do things' in this sense are beings which seem to be reasoning from their own points of view. Think, for example, of chimpanzees or dogs which are (or seem to be) capable of (more or less complex) means-end reasoning and to which we attribute motives for the things they do. Having motives presupposes that the being in question has its own perspective. But does this imply that the being conceives of its motives as its own motives? Does it imply that the being conceives of itself as itself? The answer is 'no'.

Only beings with a first-person perspective conceive of themselves as themselves, conceive of their desires, beliefs, intentions as desires, beliefs, intentions of themselves, and of their actions as actions of themselves, and can, through that, take a position regarding their motives for their actions. Therefore, only beings with a first-person perspective have the capability to reflect on their motives for their actions; and having the capability to reflect on one's motives for one's actions is what *rational agency* means.

We are beings with a first-person perspective, and therefore, because of our first-person perspective, we are beings that are capable of rational agency. Does this mean that we *act rationally* in every concrete situation? Of course not. Firstly, we sometimes act without knowing the reason for our action; we find ourselves performing a certain action, so to speak. Secondly, we often act knowing why we do what we do, but without having reflected on our motives for action. Thirdly, sometimes, after having reflected on our motives for action, we are not moved by the reason we want to be motivated by. We know why we do what we do, but we do not want to be motivated by that reason. In none of these situations do we act rationally. In short, we are rational agents, but we are not acting rationally in every concrete situation.

Rational agents *will* act rationally, in certain situations. A rational agent has the capability to reflect on her reasons for action, which (could) motivate her to act in a certain way. Therefore, at least in situations of conflicting reasons, a rational agent will reflect on her reasons. A rational agent will evaluate, for example, whether some of her conflicting desires could be fulfilled later, whether some of her desires are less strong than others, whether a certain desire can be fulfilled with less effort than the others, whether she could think of an action the performance of which would fulfil all (or some of the) conflicting desires; and she will then prefer to be motivated by a certain reason and not by others.[9] If the agent actually acts based on the reason she wants to be motivated by, then she acts rationally in the situation in question.[10]

---

[9] Preferring to be motivated by certain (of one's) desires and not by others is a sort of identification with certain (of one's) desires; therefore, this view could be called a Frankfurtian view.

[10] The fact that, in situations of conflicting reasons, a rational agent will prefer to be motivated by certain reasons and not by others, does not mean that the person in question actually will *act* based on the reasons she wants to be motivated by. I will come back to such situations in Part II of this chapter.

Evaluating desires based on preferences and rejecting desires because of some contingent conflict with other goals is a kind of deliberation that we may call (with Charles Taylor) *weak evaluation*[11]. The agent in question is a 'simple weigher of alternatives'[12]. Beings with a first-person perspective are, because of a first-person perspective, capable of this kind of rational agency, that is, they are capable of weighing alternatives in the sense of weak evaluation. I call this kind of rational agency *weak rational agency*. (We have to qualify the kind of rational agency, because, as I will argue later, there is a stronger form of rational agency which requires, on top of a first-person perspective, another property of the being in question.)

A FIRST-PERSON PERSPECTIVE IS NOT SUFFICIENT TO EXPLAIN OUR SELF

Persons characteristically have certain properties that cannot be explained with a first-person perspective only, for example: we make life plans; we ask ourselves, in certain situations, what kind of person we would be if we were to pursue a certain goal; we ask ourselves what kind of life we want to live. Why is a first-person perspective not sufficient to explain these properties? To make life plans requires that one think of one's life as a unity; and this way of conceiving of one's experiences and actions is not an implication of a first-person perspective, but it requires that the person in question have a narrative self-conception.

Deliberation about the kind of person one would be if one were to pursue a certain goal in a certain situation requires that one make value judgements; and the capability to make value judgements is not an implication of a first-person perspective. Making value judgements is a 'deeper' form of evaluation than weighing alternatives; it requires that the person in question have the property of making qualitative distinctions[13].

To ask ourselves what kind of life we want to live even requires all three properties: (1) a first-person perspective, which is necessary for conceiving of oneself as oneself and thinking of oneself qua self in the future and in other circumstances; (2) a narrative self-conception, which is necessary for conceiving of one's experiences and actions as interrelated within the unity of 'living a life'; and (3) the property of making qualitative distinctions, which is necessary for judging that a certain kind of life is of higher value than another.

---

[11] See Taylor 1985.

[12] Charles Taylor's term, see Taylor 1985, p. 23.

[13] This statement obviously calls for explanation. I elaborate on the property of making qualitative distinctions in the forthcoming section 'Qualitative distinctions – the orientation-providing element of one's self'.

2 A narrative self-conception – the structuring element of one's self

A PERSON WITH A NARRATIVE SELF-CONCEPTION CONCEIVES OF HERSELF AS THE PROTAGONIST OF HER OWN AUTOBIOGRAPHICAL NARRATIVE

We think of our life as an autobiographical narrative.[14] That means we interpret our experiences, actions, traits, beliefs and desires, we connect them and position them in relation with each other, and as a result of that, we 'produce' the autobiographical narrative that we conceive of as our life. I call the property of a narrative self-conception *the structuring element of the self*, because, owing to our narrative self-conception, we give a certain structure to our experiences, actions, traits, beliefs and desires. This does not mean that we (have to) remember all incidents and episodes of our life. Narrative self-conception means that a person sees her experiences, actions, traits, beliefs and desires through the glasses of her life story, so to speak. Depending on the person's self-interpretation, some experiences, actions, traits, beliefs and desires get a central place in the autobiographical narrative whereas others acquire peripheral positions.

Our self-interpretation is not a one-off activity, but a *continuous process* of interpretation and re-interpretation. We interpret our current attitudes and actions within the context of our autobiographical narrative, we re-interpret past actions and attitudes in the light of present ones, and our current interpretation of our life, which includes our plans and goals, is the one which we will confront ourselves with in the future.

A NARRATIVE SELF-CONCEPTION CONSTITUTES THE IDENTITY OF A SELF AS CONTINUOUSLY EVOLVING SELF-INTERPRETATION

On the basis of the structuring element of our self, we characterize ourselves by means of an autobiographical narrative, in which our self appears as the protagonist. One could say that we try to answer the question 'Who am I?' – a question that only beings with a first-person perspective are able to think. And we seem to get an answer to that question, because we conceive of the character that is the protagonist of our autobiographical narrative as our self. But one's answer to the question 'Who am I?' is the answer of a certain moment. The answer will change with one's continuously evolving self-interpretation. Hence, the character that is the protagonist of our autobiographical narrative and that we conceive of as our self will change with the development of the narrative that we consider our life.

In short, a narrative self-conception constitutes an identity with certain content, but this content changes with one's self-interpretation. Therefore, one's *narrative identity* is imbued with uncertainty about one's self.

---

[14] With respect to this aspect of my notion of the self, I was inspired by the theory of Marya Schechtman, as expounded in Schechtman 1996.

A NARRATIVE SELF-CONCEPTION IS CHARACTERISTIC OF BUT NOT FUNDAMENTAL FOR BEING A SELF

Although a narrative self-conception is one of our characteristics, this property is not fundamental for being a self. In other words, there are human *persons* – human creatures with a first-person perspective – that lack a narrative self-conception.

Firstly, the property of a narrative self-conception may not be (fully) developed (yet). Think of young children: At a certain age, they conceive of themselves as themselves, but, even though they already have a first-person perspective, their world is less connected and integrated than it is for normally developed adults. Secondly, narrative self-conception may diminish. A patient with Alzheimer's disease, for example, seems to lose his life story, he seems to lose the capacity to connect current desires, beliefs, actions and experiences to past ones; and so, with progressing disease, this person comes to live more and more in the present only.[15] Thirdly, a person's self-conception may be fragmented. A long-term drug user, for instance, may have lost access to long periods of his life (story); he is no longer able to connect actual desires, beliefs, actions and experiences to his past.

Furthermore, the property of a narrative self-conception may be developed to a greater or lesser degree. Think of a person who makes life plans, but who does not try to interpret and position her attitudes and actions in relation to her life plans. Despite making life plans, this person's narrative self-conception is underdeveloped, otherwise she would interpret and position her attitudes and actions in relation to her life plans.

In sum: The property of a narrative self-conception may not be (fully) developed (yet) or completely or partially lost. In such cases, the person in question completely or partially lacks the structuring element of the self.

3 Making qualitative distinctions – the orientation-providing element of one's self

ONLY PERSONS THAT MAKE QUALITATIVE DISTINCTIONS HAVE VALUES

A rational agent who, in a concrete situation, makes qualitative distinctions[16] does not evaluate the alternatives based on preferences (only), but she assesses the alternatives in a 'deeper' way (also). That is to say, she assesses alternatives using a language of qualitative contrast, judging, for example, certain desires to be nobler, or certain actions to be more honourable than

[15] The human being in question, although still alive, will probably even lose the property of having a first-person perspective. If this were to happen, the human being would not exist as self anymore.

[16] The notion of 'qualitative distinctions', as I use it, is derived from Charles Taylor, as set forth in Taylor 1982a and in Taylor 1985.

others. Such a person typically asks herself what kind of person she would be if she were motivated by a certain desire (in the current situation), if she strived to realize a certain goal (in the current situation), or if she performed a certain action (in the current situation). "Motivations or desires do not only count in virtue of the attraction of the consummations, but also in virtue of the kind of life and kind of subject that these desires properly belong to."[17] A rational agent who evaluates in this 'deep' sense 'has things' which truly matter to her and which she therefore wholeheartedly wants to protect or to realize, and which she associates with the kind of person she wants to be. Such a being has values.

Could we have values without making qualitative distinctions? The answer is 'no'. The property of making qualitative distinctions is constitutive of having values.[18] One cannot sensibly say one has values if one does not make qualitative distinctions. By making qualitative distinctions, we 'discover' what really matters to us; and these values play a role when we assess our options for action, in future situations. If we were not making qualitative distinctions, we wouldn't have values.

STRONG RATIONAL AGENCY REQUIRES THE PROPERTY OF MAKING QUALITATIVE DISTINCTIONS

A rational agent who, in a concrete situation, evaluates the alternatives in the 'deep' sense described above is a 'strong evaluator' (Charles Taylor's term). The 'trigger' for strong evaluation is not the contingent incompatibility of the various alternatives (as it is for 'weak evaluation'), but the agent's sensitivity to the fact that her values are at stake. The following example, which I sketch in two scenarios, illustrates the crucial difference between weak and strong evaluation.

A woman who has restricted herself to a diet gives a birthday party. She has prepared a cake, because, for her and her family, a birthday cake is part of the celebration. Obviously, the cake does not fit into her diet and she deliberates what to do. In the first plot, her thoughts are: My birthday is a special day that I want to enjoy; it will not make much difference to the result of the diet whether I eat a slice of cake or not; I can easily compensate this deviation from my diet by reducing calories tomorrow. Eventually, she decides to forget about her diet for this special day. In the second plot, the woman has thoughts like this: Although I feel tempted to eat a slice of cake, I should be able to resist that temptation; if I were to eat cake, I would not have control over myself; I do not want to be a person with such a weak character. So in the end she decides not to take a slice of cake.

Let me analyse the two kinds of deliberation. In the second plot, the

---

[17] See Taylor 1985, p. 25.

[18] This view is neutral with respect to the question whether values exist. My point is that we only *have* values if we make qualitative distinctions in which we 'apply' these values.

woman experiences the situation as her value of 'rational control of oneself' being at stake. She asks herself what kind of person she would be if she were to eat a slice of cake. In answering that question, she makes a qualitative distinction, which belongs to the domain of the value in question, between persons with a strong character, that is to say, those who succeed in having rational control of themselves, and persons with a weak character, that is to say, those who fail to achieve rational control of themselves. The woman in the first plot, however, does *not* experience the situation as calling for strong evaluation; she is weighing alternatives. On the one hand, she wants to enjoy her birthday including the cake, on the other, she wants to lose weight, and, in her deliberation, both motives "count in virtue of the attraction of the consummations"[19].

Strong evaluation, to recapitulate, requires that the rational agent count "motivations or desires in virtue of the kind of life and kind of subject that these desires properly belong to"[20], thereby assessing her motivations and alternative actions in qualitative terms, judging certain motivations and certain alternative actions to be qualitatively 'higher' than others. This is not to say that a person who has the property of making qualitative distinctions will apply strong evaluation in every process of deliberation (the woman from the second plot may seem somewhat over-sensitive regarding one of her values), but that such a person is *capable* of strong rational agency. I call our property of making qualitative distinctions *the orientation-providing element of the self*, because, owing to this property, we have values, are sensitive to situations where our values are at stake, and are inclined to assess our alternatives by applying strong evaluation (instead of simply weighing alternatives), in these situations.

Making qualitative distinctions constitutes an identity of oneself, which is one's ethical orientation[21]

Our values provide orientation in two ways: they provide orientation by making us sensitive to situations where strong evaluation would be

---

[19] This clause belongs to a citation that I referred to in the first paragraph of this section: "Motivations or desires do not only count in virtue of the attraction of the consummations, but also in virtue of the kind of life and kind of subject that these desires properly belong to." (Taylor 1985, p. 25)

[20] See the preceding note.

[21] Because the substance of my notion of 'ethical orientation' has its origins in the view of Charles Taylor, as expounded in Part I of *Sources of the Self* (see Taylor 1989), and Taylor elaborates in these chapters on our *moral* thinking, I should explain why I designate the orientation (and identity) in question as being *ethical* instead of moral. Taylor uses 'moral' in a broad sense. He identifies three axes of moral thinking (or moral life): (1) respect for and obligations to others; (2) our understandings of what makes a full life; and (3) dignity (our sense of ourselves as commanding respect); see pp. 14-15. Other mainstream philosophers in the domain of ethics would probably regard Taylor's first axis as pertaining to moral thinking, whereas they would regard the other two axes as ethical thinking. I take it that our qualitative distinctions, which we make when we apply strong evaluation and which concern the three axes described by Taylor, exhibit an orientation to 'the good' – which is an ethical notion – and that this orientation to the good is the basis for moral thinking in a narrow sense too.

appropriate, because our values could be at stake in these situations, and they provide orientation when we evaluate our desires and (alternative) actions, by functioning as standards. This orientation, which is based on our property of making qualitative distinctions, is an *ethical orientation*, because it means to "think, feel, judge [...] with the sense that some action, or mode of life, or mode of feeling is incomparably higher than the others which are made readily available to us"[22]. Our ethical orientation is characteristic of who we are, because it sketches the frameworks of the 'ethical space'[23] that we operate in as thinking and acting beings. In short, making qualitative distinctions is constitutive of one's ethical orientation. And because our ethical orientation is characteristic of who we are, we may call our ethical orientation our *ethical identity*.

Could we answer our question 'Who am I?' by listing our values? – No. One's ethical identity is imbued with uncertainty about one's self, because one cannot be sure about one's values. Why is that? Firstly, our 'deepest' values may be values that were never at stake in the concrete situations we found ourselves in. And, because we only conceive of a certain value as a value of ourselves when we take (or have taken) this value into account in concrete situations of strong evaluation, we may not know that certain values are of utmost importance to us. Secondly, our values can change during our existence (as selves). In concrete situations of strong evaluation, it may turn out that we do not take certain values into account anymore, and that other values prevail.

TO MAKE QUALITATIVE DISTINCTIONS IS CHARACTERISTIC OF BUT NOT FUNDAMENTAL FOR BEING A SELF

When I say that it is characteristic of us to make qualitative distinctions, I do not mean that we, in every concrete situation, assess our alternatives by asking ourselves what kind of person we would be if we were to perform a certain action. What it means is this: it is characteristic of ourselves to be sensitive to situations where our values are at stake, and to be inclined to apply strong evaluation, instead of weighing alternatives based on preferences, in such situations.

Although it is characteristic of us to make qualitative distinctions, this property is not fundamental for being a self. There are human beings with a first-person perspective that lack the property of making qualitative distinctions. Here are two different cases. Firstly, the property of making qualitative distinctions may not be (fully) developed (yet). Young children that already conceive of themselves as themselves and who, because of their first-person perspective, are already able to evaluate their desires based on

[22] Taylor 1989, p. 19.

[23] The space metaphor is borrowed from Charles Taylor, see Taylor 1989, p. 25. It has to be remarked that Taylor does not use the term 'ethical space' but the term 'moral space'.

preferences, may not yet assess their (alternative) actions as 'right' or 'wrong'. For example, when a young child wants to play with a certain toy, which another child is playing with, it just tries to take it. It will take a period of time, before the same child will (probably) reject a desire to take the toy that another child is playing with, because it judges the action to take the toy away from the other child to be 'wrong'.

Secondly, one can (temporarily) partially lose the property of making qualitative distinctions. Eugen Kogon has written about nazi concentration camps that prisoners, after a period of time, developed a kind of personality of which "seelische Primitivierung" (Kogon's term, meaning 'degradation of the soul') is a characteristic feature. "The soul developed a protecting crust, a kind of armour, preventing that every impression would come through. [...] One became hard, many prisoners became deadened [insensitive to harm, sorrow, compassion, dread, grief]."[24] If I am right, what Kogon calls "seelische Primitivierung" is – presumably together with other changes – the partial loss of the property of making qualitative distinctions. Such a person is not sensitive anymore to whether certain values, which did matter to her long ago, are 'sacrificed' by acting the way she acts. She does not ask herself what kind of person she would be if she were to perform a certain action. The person in question has, (at least) temporarily[25], partially lost the property of making qualitative distinctions.

WHY CAN WE NOT GET RID OF THE QUESTION 'WHO AM I'?

Because of our first-person perspective, we are able to ask 'Who am I?'. But we cannot answer that question based on a first-person perspective only, because the identity of our self, which is constituted by a first-person perspective, is a content-free identity. It is nothing more than the certainty of 'being-self'. Nevertheless, we want to give content to that certainty. Provided that we have the required properties, we try to answer our question by characterizing our self as the protagonist of our autobiographical narrative or by sketching the 'ethical space' that we operate in as thinking and acting beings. Both attempts provide answers to our question 'Who am I?'. But these answers do not satisfy us, because they are provisional and uncertain.

If both options to answer our question 'Who am I?' are provisional and uncertain and we therefore cannot give definite content to our certainty of 'being-self', why do we not just lay that question aside? I claim that we cannot get rid of that question, because we are, *each of us*, concerned with our self.

---

[24] Kogon 1974, p. 369, freely translated. The explanation between square brackets is added by me, but derived from the context of the citation.

[25] Afterwards, (some of) the few survivors did (and still do) not only suffer from the incredible terror they were exposed to, but also from things they themselves did in order to survive, the latter showing, in my view, that the property of making qualitative distinctions has been recovered (and, thus, that it can recover).

Part II – The concern with one's self, as a relation of oneself to one's self

INTRODUCTION

Because of the properties that constitute our self, persons – or rather, persons that do not lack the two characteristic but not fundamental properties – are concerned with their self in three specific ways. Firstly, because of a first-person perspective, persons have a mattering-relation with their rational agency, that is to say, it matters to them whether they can justifiably conceive of themselves as rational agents. Secondly, because of a narrative self-conception persons have a mattering-relation with their autobiographical narrative, in other words, it matters to them whether they live a life that they can conceive of as a narrative. Thirdly, because of the property of making qualitative distinctions persons have a mattering-relation with their values, that is to say, it matters to them whether they can justifiably conceive of themselves as persons who have (those) values.

Taking into consideration that 'the self' one is concerned with is object *and* subject of this concern, we can imagine that one's concern with one's self has a certain 'effect' on oneself. The question is: What is the 'effect' of one's concern with one's self on oneself? In my view, the 'effect' is threefold. Firstly, one's concern with one's self is partially *constitutive of one's self*. Let me explain this statement with respect to the structuring element of the self. We not only think of our life as a narrative, but we also strongly desire to live a life that we can conceive of as a narrative. After all, living a life that one can conceive of as a (coherent) narrative means to be able to conceive of one's self as the (well-defined) protagonist of one's self-narrative. And, because desires strive for fulfilment, one's desire to live a life that one can conceive of as a coherent narrative is partially – that is to say, together with the property of a narrative self-conception, which is the basis for this desire – constitutive of a self that one can conceive of as the well-defined protagonist of one's autobiographical narrative. So, keeping in mind that product and producer are one, we can say: one's concern with one's autobiographical narrative is partially constitutive of one's self.

Secondly, one's concern with one's self motivates oneself *to reflect on oneself* qua self. Let us again look at the structuring element of the self. The desire to be able to conceive of one's self as the well-defined protagonist of one's autobiographical narrative would have the greatest chance of being fulfilled if one continuously reflected upon whether one's actual desires and actions fit the protagonist of one's most recent autobiographical narrative. That is why one's concern with one's autobiographical narrative motivates oneself to reflect on oneself qua self.

Thirdly, one's concern with one's self is the basis for certain *emotions*[26] *towards oneself* qua self. If a person is not able to conceive of herself as the well-defined protagonist of her autobiographical narrative anymore, because

she lives a life that she can no longer conceive of as a (more or less) coherent narrative, then she can feel distressed or may have emotions which she interprets as loss of self-confidence. These emotions may result from the person's reflection on herself qua self, but they may also be a 'direct' effect – that is to say, not mediated by conscious reflection – of her unfulfilled desire to be able to conceive of herself as the well-defined protagonist of her autobiographical narrative. In the last situation, her emotions may motivate the person to reflect on herself qua self.

4 Our concern with our rational agency – belonging to the fundamental element of one's self

BEINGS THAT ARE SELVES HAVE THE DESIRE TO JUSTIFIABLY CONCEIVE OF THEMSELVES AS BEINGS THAT ARE LED BY REASONS OF THEIR OWN

Because of a first-person perspective, persons have the capability of rational agency and conceive of themselves as rational agents. However, this way of conceiving of ourselves is threatened if we, in concrete situations, do not act as rational agents. This fact can explain why the capability of (weak) rational agency, which is an implication of the property of having a first-person perspective, is not only the basis for conceiving of ourselves as rational agents but, by this, also for our concern with our rational agency. In short, conceiving of oneself as a rational agent involves concern with one's rational agency.

Being concerned with one's rational agency means having a strong desire to justifiably think of oneself as a rational agent, in other words, to justifiably think of oneself as a person who is led by reasons of her own, that is to say, reasons she wants to be motivated by. What does 'justifiably' mean in this context? When evaluating one's actions, one does not want to have to conclude that one acted as a 'wanton'[27], that is, without having used one's capability of rational agency, or that one was motivated by reasons one did not want to be motivated by.

A person *without* concern with her rational agency is not conceivable. Such a person would be a person to whom it would not matter, in any concrete situation, by which of her desires she were motivated. Even in

[26] Generally speaking, to have a certain emotion involves a conscious or unconscious judgement about a situation. The emotion in question is an expression of that judgement – whether in physiological occurrences, sensations or (voluntary or involuntary) actions. In the case of emotions connected with one's concern with one's self, a person's judgement may have its initial expression in a sensation of (comfort or) discomfort – for instance, distress or shame – and the person may then interpret that sensation as, for example, loss of self-respect or loss of self-esteem. The sensation of discomfort may also be secondary; it may appear during a person's explicit self-reflection, in which she, for example, comes to realize that she has lost her self-esteem.

[27] I borrow the term 'wanton' from Harry Frankfurt, see Frankfurt 1971, p. 16.

situations of conflicting desires, such a person would not evaluate her desires (which could motivate her to act in a certain way). But this would be in contradiction with the fact that a person is a being with a first-person perspective. A being with a first-person perspective is aware of her agency, has the ability to reflect upon her desires, and will, at least in situations with conflicting desires, evaluate her desires (which could motivate her to act in a certain way).[28] And if a person evaluates her desires and prefers to be motivated by a certain desire and not by others, it matters to her whether she is motivated by the 'chosen' desire or by one of the others which she does not want to be motivated by. Therefore, our concern with our rational agency belongs to our personhood.

IN OUR CONCERN WITH OUR RATIONAL AGENCY, WE APPLY THE NORMATIVE CRITERION OF CONSISTENCY

Most (if not all) of us know situations where we catch ourselves doing something without knowing why we are doing what we are doing. In such situations, we are doing things and we are aware of the fact that we are doing things, but our actions are not intelligible to us as actions based on reasons, let alone as actions based on reasons of our own. We also know situations where we act based on reasons – that is to say, a certain reason is intelligible to us as reason for our action – but these reasons are not the reasons we want(ed) to be motivated by (in the concrete situations in question). For example: A person feels like eating chocolate; however, she does not want to be motivated by this desire, because she is on a diet; nevertheless, she takes a piece of chocolate and eats it. In a case like this, there is no *consistency* between the reason a person wants to be motivated by and the reason she actually is motivated by.

If we find ourselves frequently and regularly in situations where we do not know why we are doing what we are doing, or in situations where we know the reasons for our actions but do not want to be motivated by those reasons, then we do not meet the normative criterion of consistency that we (because of our concern with our rational agency) implicitly set for ourselves. After all, you can only justifiably think of yourself as a rational agent if there is consistency, in concrete situations, between the reasons you want to be motivated by and the ones you are actually motivated by.

---

[28] Why would a being with a first-person perspective evaluate her conflicting desires? "If one tries to satisfy conflicting desires, the result is likely to be chaos that satisfies none of them. So, if one has the ability to evaluate and order desires, one will attempt to do so. And a first-person perspective entails the ability to evaluate and order desires." (Baker 2000, pp. 158-159)

NOT MEETING THE CRITERION OF CONSISTENCY CAN LEAD TO NEGATIVE EMOTIONS TOWARDS ONESELF

If we have to conclude that we are frequently and regularly motivated by reasons that we do not want to be motivated by, then we can be overcome by distress or shame, because we do not meet our criterion of consistency. Not meeting the criterion of consistency may even result in situations of mental crises. Think of a person who cannot become master of her food addiction; the person does not succeed in abandoning her desires to eat high-calorie food, and she suffers from the fact that she acts based on these desires.

If we realize (and this may be even more threatening than the aforementioned situation, although in another way) that we frequently and regularly perform actions without knowing why we are doing what we are doing, then we feel confused, because our actions are not even intelligible to us as actions based on reasons. Having frequently and regularly realized that the things we do just happen to us, we may become afraid for our future. We might be afraid to face a future in which we will not be conscious of our agency anymore. And, because being conscious of one's agency belongs to personhood, we in fact might be afraid to face a future in which we will not exist as self anymore, although the human organism that now constitutes us as a person will still be alive.[29]

Although I gave emphasis to negative emotions, it should be mentioned that the concern with one's rational agency does also become 'visible' in *positive* emotions towards oneself. We can, for example, feel proud of ourselves or feel satisfied with ourselves if we succeed in abandoning desires that we do not want to be motivated by.

5 Our concern with our autobiographical narrative – belonging to the structuring element of one's self

MOST BUT NOT ALL PERSONS HAVE A DESIRE TO LIVE A LIFE THAT THEY CAN CONCEIVE OF AS A NARRATIVE

Because of a narrative self-conception, persons conceive of their life as a narrative and of themselves qua self as the protagonist of their autobiographical narrative. But this way of conceiving of oneself is threatened if a person does not actually live a life that she can conceive of as a narrative. This fact can explain why a narrative self-conception is not only the basis for conceiving of one's life as a narrative (and of one's self as the protagonist of that narrative) but, by this, also for one's concern with one's autobiographical narrative. In short, conceiving of one's life as a narrative (and of one's self as the protagonist of that narrative) involves concern with one's autobiographical narrative.

[29] I use Baker's 'constitution view' of persons, see Baker 2000.

Being concerned with one's autobiographical narrative means having a strong desire to live a life that one can conceive of as a narrative and, by this, to be able to conceive of one's self as the well-defined protagonist of that autobiographical narrative. We do not want to have to conclude that our current goals and actions are by no means intelligible to us as evolving from our earlier beliefs, goals, traits, actions and experiences. Thus, our autobiographical narrative is an object of our concern with our self.

Does this mean that *all* persons have this concern with their self (in the same way)? The answer is 'no'. Why is that? Firstly, because a narrative self-conception is not fundamental to being a person; there are persons who do not have the desire to live a life that, for themselves, is conceivable as a narrative. In the case of these persons, the property of a narrative self-conception is either not developed (yet) or lost; and persons who lack the property of a narrative self-conception also lack the corresponding concern with their self. Secondly, because the property of a narrative self-conception is not an all-or-nothing property, but a property that comes in degrees, it may be developed to a large extent or to a limited extent, or it may be partially lost. And to the various possible degrees of narrative self-conception correspond various *degrees of concern* with one's autobiographical narrative. Therefore, persons can have a strong or a less strong desire to live a life that, for themselves, is conceivable as a narrative.

IN OUR CONCERN WITH OUR AUTOBIOGRAPHICAL NARRATIVE, WE APPLY THE NORMATIVE CRITERION OF COHERENCE

The criterion of our concern with our autobiographical narrative is *coherence*[30], that is to say, we want to be able to conceive of our self as the well-defined protagonist of the unfolding story of our life. Most of us have experienced situations of (radical) changes in their goals, plans and practical beliefs. In such situations we characteristically reinterpret our past to achieve some coherence in our self-narrative; in other words, we reinterpret our past to be able to conceive of the changes in our goals, plans and practical beliefs as developments of our self. And we can probably imagine that we would feel very disconcerted if the changes in our goals, plans and practical beliefs were by no means intelligible to us as developments of our self.

By saying that we want to achieve *some* coherence, I want to point to two

---

[30] In my view, the coherence criterion is *the* criterion of our concern with our self-narrative. One could object that we are also concerned with the *correctness* of our self-narrative. I claim, although I will here not argue for this point, that a criterion of correctness (for example, Schechtman's 'reality constraint', Schechtman 1996, pp. 119-130) is primarily a criterion for the interlocutors of the person in question and only secondarily for the person herself. Although a person, confronted with (minor) factual or interpretative inaccuracies, may be willing to revise her self-narrative, she then, because of conceiving of her self as the protagonist of her self-narrative, will be concerned with the *coherence* of that narrative.

things. Firstly, a narrative can be coherent in different ways; it has not necessarily to be linear. Secondly, because persons (depending on the degree of their narrative self-conception) can have a stronger or a weaker desire to live a life that, for themselves, is conceivable as a narrative, different persons can be satisfied with different *degrees of coherence*.

NOT MEETING THE CRITERION OF COHERENCE CAN LEAD TO NEGATIVE EMOTIONS TOWARDS ONESELF

If, to be able to relate current goals, plans, practical beliefs and actions to goals, plans, actions, practical beliefs, traits and experiences in former periods of our life, we (often) have to reinterpret our life (and thereby our self) dramatically, then we can be overcome by distress or by emotions that we interpret as a sense of inferiority, as a loss of self-respect or of self-confidence. On the other hand, we can feel satisfied or have emotions that we interpret as self-confidence if we succeed in living a life that unfolds as a coherent narrative.

Whether a certain person has the just mentioned emotions towards herself in the described situations depends on the strength of her desire for coherence. If a person (because of a high degree of narrative self-conception) has a strong desire to live a life that she can conceive of as a coherent narrative, and that person does *not* succeed in conceiving of herself as the well-defined protagonist of her self-narrative, she may suffer from her own fragmented self and may even end up in a mental crisis.

6 Our concern with our values – belonging to the orientation-providing element of one’s self

MOST BUT NOT ALL PERSONS HAVE A DESIRE TO JUSTIFIABLY CONCEIVE OF THEMSELVES AS PERSONS WHO HAVE VALUES

Because of the property of making qualitative distinctions, persons have values and conceive of themselves as having values. But this way of conceiving of oneself is threatened if a person does not actually act as a strong rational agent, that is to say, if a person, in concrete situations, when certain values of that person are at stake, does not assess the alternatives by applying strong evaluation. That person cannot justifiably think of herself as a person who has values. This fact can explain why the property of making qualitative distinctions is not only the basis for conceiving of ourselves as having values but, by this, also for our concern with our values. In short, conceiving of oneself as having values involves concern with one’s values.

Being concerned with one’s values means having a strong desire to justifiably conceive of oneself as a person who has values. Persons

characteristically have this desire. We do not want to have to conclude, in retrospect, that we 'sacrificed' our values by acting as we did.

Does this mean that all persons have this concern with their self (in the same way)? The answer is 'no'. Why is that? Firstly, because the property of making qualitative distinctions is not fundamental to being a person, there are persons who do not have the desire to justifiably conceive of themselves as persons who have values. The property of making qualitative distinctions may either not be developed (yet) or lost; and persons who lack the property of making qualitative distinctions also lack the corresponding concern with their self. Secondly, because the property of making qualitative distinctions is not an all-or-nothing property, but a property that comes in degrees, it may be developed to a large extent or to a limited extent, or it may be partially lost. And to the various possible degrees of that property correspond various *degrees of concern* with one's values. Therefore, persons can have a strong or a less strong desire to justifiably conceive of themselves as persons who have values.

What does it mean to say that persons may be concerned with their values to a greater or lesser extent? Think of a person who makes qualitative distinctions in a certain situation, but who, in another situation, is not sensitive to the fact that just these values are at stake. By making a qualitative distinction in the first situation, the person has certain values. By, in the second situation, not being sensitive to the fact that just these values are at stake and therefore not applying strong evaluation, the person fails to reinforce her values. If this scenario is not just an incident for the person in question and that person in no way suffers from 'sacrificing' her values, then we can say that she has limited concern with her values.

IN OUR CONCERN WITH OUR VALUES, WE APPLY THE NORMATIVE CRITERION OF INTEGRITY

The criterion of our concern with our values is *integrity*, that is to say, we want our actions 'to be in line with' (that is, to be not in conflict with, to protect or to exemplify) our values. If we were sensitive to *every* situation where our values are at stake, and if we applied strong evaluation in *each* such situation (and if the situation in question is *not* an ethical dilemma), and if we, based on our strong evaluation, in *each* such situation, were motivated by the reason we want to be motivated by, then we 'automatically' would meet the criterion of integrity. But, because we are not perfect (and can be confronted with an ethical dilemma), this is not what happens in reality.

Sometimes, when retrospectively reflecting on our actions, we must conclude that we actually acted in conflict with our values. Firstly, one may have acted as a wanton, without any reflection. Secondly, because of not being sensitive to the fact that one's values were at stake in a certain situation, one may have evaluated the alternatives by 'just' weighing them with respect

to their being conducive to the achievement of a certain goal. Thirdly, one may have been sensitive to the fact that one's values were at stake in a certain situation and, therefore, may have started deliberating about one's alternatives by applying strong evaluation, but then one may have switched to weak evaluation by weighing the alternatives with respect to a certain goal, and, given the strength of the desire to achieve that goal, one finally may have preferred to be motivated by that desire. Fourthly, one may have reflected on one's alternatives by applying strong evaluation, without, when performing the action, being motivated by the reason one wanted to be motivated by.

There are also situations that are experienced as *ethical dilemmas*. In such situations, according to a person's judgement, all her alternatives are in conflict with her 'deepest' values, and whatever she will do, she will fail to meet her own criterion of integrity. Although the person in question will probably (continue to) apply strong evaluation and perform the action she finally decides to do, the person does not really want to perform that action and may suffer from having acted in conflict with certain of her 'deepest' values.

When a person experiences or concludes that she actually acts or acted in conflict with her values, then she, characteristically, will be dissatisfied with herself, because she does not meet the criterion of integrity, which she applies to herself because of her concern with her values.

NOT MEETING THE CRITERION OF INTEGRITY CAN LEAD TO NEGATIVE EMOTIONS TOWARDS ONESELF

Depending on the degree of one's concern with one's values, a person who experiences or concludes that she acted in conflict with her values can be overcome by regret, remorse or emotions that she interprets as loss of self-esteem. Think of the person from the third example above, who neglected her initial value judgement. In retrospect, this person may be dissatisfied with herself and feel regret about the fact that she sacrificed (one or some of) her values to achieve a certain goal. The negative emotions towards herself may be more or less strong; these differences can be explained by the fact that not all persons have the same strong desire for integrity.

If a person (because of the property of making qualitative distinctions being developed to a high degree) has a strong desire for integrity – in other words, a high degree of concern with her values – and if that person knows that she acted in conflict with her 'deepest' values, then she may suffer from not having been loyal to her own values. We might, for example, think of the person who found herself in an ethical dilemma and could not do otherwise than perform an action that was in radical conflict with her 'deepest' values. This person may not be able to conceive of herself as a person who has values anymore. She may be afraid that she never will gain back her self-esteem. Such a person may even end up in a mental crisis.

7 The need for unity of one's self

OUR CONCERN WITH OUR SELF EXHIBITS A NEED FOR UNITY

All three concerns with one's self exhibit a need for unity. Firstly, we are concerned with our rational agency, and this concern, belonging to the fundamental element of one's self, exhibits *a need for consistency of one's self*. We want to justifiably conceive of ourselves as rational agents (that is, agents who act based on reasons of their own); and that desire will only be satisfied if there is consistency between the reasons we want to be motivated by and the actual reasons for our actions[31]. This *soundness of the will* is the specific unity that one's concern with one's rational agency aims at.

Secondly, we are concerned with our autobiographical narrative, and this concern, belonging to the structuring element of one's self, exhibits *a need for coherence of one's self*. We want our self to be the well-defined protagonist of our autobiographical narrative; we do not want to appear as a fragmented self or as a number of protagonists. This *well-defined character* is the specific unity that one's concern with one's autobiographical narrative aims at.

Thirdly, we are concerned with our values, and this concern, belonging to the orientation-providing element of one's self, exhibits *a need for integrity of one's self*. We want to justifiably conceive of ourselves as persons who have values; and that desire will only be satisfied if our actions are conceivable for ourselves as guided by our values and, in retrospect, as protecting or exemplifying our values. This *loyalty to oneself* is the specific unity that one's concern with one's values aims at.

THE UNITY OF ONE'S SELF IS THREATENED BY INCONSISTENCY, INCOHERENCE AND NON-INTEGRITY

As a result of self-reflection or because we have been overcome by certain negative emotions towards ourselves, we may be confronted with fragmentation of our self, and with divergence and conflict in our self. Firstly, if we are confronted with divergence between the reasons that we were motivated by and the reasons we want to be motivated by, the unity of our self is threatened by inconsistency. Secondly, if we are confronted with a fragmented self as the protagonist of our autobiographical narrative, the unity of our self is threatened by incoherence. Thirdly, if we are confronted with conflict between our actions and our values, the unity of our self is threatened by non-integrity.

Frustration of one's need for unity of one's self can lead to negative

[31] The actual reasons for our actions are the ones that we think of as being the motivating reasons for our actions, and these reasons need not necessarily be identical with the reasons we want to be motivated by. This leaves aside that a person may be mistaken about the actual motivating reasons for her action.

emotions towards oneself and sometimes even to situations of mental crises.[32] Whether and to what degree one suffers from inconsistency, incoherence and non-integrity of one's self depends on the strength of one's desire for consistency, coherence and integrity. This strength of the respective desires depends on the degree of the concern in question, and the degree of the concern in question depends on the degree to which the corresponding property is developed. In short, different persons may be satisfied with different levels of unity.

8 Conclusion

ONE'S CONCERN WITH ONE'S SELF CONTRIBUTES TO THE CONSTITUTION OF A SELF THAT IS CONCEIVABLE (FOR ONESELF) AS A UNITY

Keeping in mind

(a) that the degree of one's concern with one's rational agency, autobiographical narrative and values may differ from person to person,

(b) and that there may be persons who do not (yet) have the property of a narrative self-conception or the property of making qualitative distinctions, respectively, or who do not have (one of) these properties anymore, and who, due to these circumstances, do not have the concern that the property in question is constitutive of,

we can state that one's concern with one's self, which can be distinguished in:

(1) one's concern with one's rational agency, which is a desire to justifiably conceive of oneself as an agent who acts based on reasons of her own, and, through this, exhibits a need for consistency of one's self,

(2) one's concern with one's autobiographical narrative, which is a desire to live a life that one can conceive of as a narrative, and that, through this, exhibits a need for coherence of one's self, and

(3) one's concern with one's values, which is a desire to be a person that one can justifiably conceive of as a person who has values, and, through this, exhibits a need for integrity of one's self,

contributes to the constitution of a self that is conceivable (for oneself) as a unity.

[32] Probably our need for unity of our self may also be the motivation to overcome such a situation. (This statement is not based on empirical facts but is purely based on theory.)

Chapter 3
Practical irrationality from a self-related perspective

In certain situations, persons conceive of their actions as irrational. This phenomenon calls for further explanation, because questions like the following arise: When do persons conceive of their actions as irrational? What standards do persons apply (whether consciously or not) when conceiving of their actions as irrational? Does it matter to us whether we act irrationally, and if so, why? Before we submit the notion of conceiving of one's own action as irrational to conceptual analysis, let me illustrate the phenomenon with an example: When I want to lose weight and therefore put myself on a diet, it may happen that I feel like eating chocolate and that, violating the behavioural rules I have set for myself, I take a piece of chocolate and eat it. I am not involuntarily eating the piece of chocolate (for instance, because it is hidden within other food) or unconscious of violating my diet (for instance, because I do not know that chocolate is high-calorie food); rather, the reverse is true. After deliberation, I decide to eat chocolate and, before, during and after eating the piece of chocolate, I am conscious of violating my diet. On the one hand, I know that I myself decided to eat chocolate; on the other hand, I nevertheless conceive of my action of eating chocolate as an irrational action. After all, what I am doing is contrary to my own behavioural rules and hinders the achievement of the purpose behind these rules, which is to lose weight. What obviously bothers me is the fact that, by acting this way, it will take longer to lose weight, if I will succeed at all. However, my contention is that what bothers me too is the brute fact of my acting irrationally.

My claim is: Conceiving of our actions as irrational matters to us, even apart from the specific purpose the achievement of which is threatened by acting the way we do (or did). Before arguing for this claim (Part II), I first will elucidate the notion of first-personal practical irrationality, which in turn will also lead to some reflection about the notions of agent and person, action and action structure, and reasoning and decision (Part I). The notion of first-personal practical irrationality, as I take it, is an 'umbrella notion' covering two separate notions: conceiving of one's own actions as irrational and conceiving of oneself as irrational. My primary concern is with the notion of conceiving of one's own actions as irrational, because, firstly, that notion is presupposed by the notion of conceiving of oneself as irrational, and

secondly, all persons know the phenomenon of conceiving of an action of oneself as irrational, whereas some may not know the phenomenon of conceiving of oneself as irrational.

It will emerge in due course that, in my view, the phenomenon of conceiving of one's actions as irrational has to be distinguished from the phenomena of conceiving of one's actions as mistaken, non-rational, or wrong.[1] Although those other phenomena stand in need of thorough analysis too, they will only be discussed as far as clarification is needed for the benefit of the notion of first-personal practical irrationality as I see it. Especially the 'mattering-question' (Why does it matter to us whether we act …?) will not be asked with regard to actions that we conceive of as mistaken, non-rational or wrong.

Part I – Conceiving of one's own actions (or of oneself) as irrational

There are situations where we know that we deliberated about an action and decided to do what we do (or did), and at the same time conceive of that action as irrational. Thus, conceiving of our own action as irrational does not (necessarily) mean that we conceive of the action as not resulting from our evaluative reasoning. When a person conceives of an action as irrational, she knows her motive for her action and she knows that she decided (whether consciously or not) to perform that action, yet she in a certain way disapproves of her motive for her action. This preliminary claim (directly or indirectly) involves various concepts, like the concept of person, action, and reasoning. And we will have to clarify them in order to shed light on the concept of first-personal practical irrationality.

1 Agents and persons

A being that does things is an agent. Agents do things[2] in the sense that certain things would not have happened if the being in question had not had certain desires, beliefs, intentions, or in other words, if its actions cannot be explained by assigning desires, beliefs, and intentions to the being in question. Agents experience the world from their own point of view. They do things from their own perspective and for their own purposes; and to achieve those purposes, agents must be capable of more or less complex *instrumental reasoning*. Acting based on one's own purposes using instrumental reasoning

---

[1] These classifications do not necessarily exclude each other; for instance, a person may conceive of an action as non-rational *and* wrong. See also Chapter 4 (Wrongness from a self-related perspective), Section 6 (The overriding nature of the self-related conception of wrongness of action).

[2] Regarding the notion of 'doing things', see note 8 in Chapter 2 (The relation of oneself to one's self).

does not imply that the being in question knows that it acts, what action it performs at a certain moment, or the purpose of that action; neither does it imply that the being thinks of itself as a subject that is capable of instrumental reasoning. In short, acting based on one's own purposes using instrumental reasoning does not require that the being is conscious of its agency or of its reasoning.

To be conscious of one's agency requires a *first-person perspective*, that is to say, it requires that the being in question is aware of the fact that it has its own perspective and that this perspective differs from the perspective of others. A being that fulfils this characterization is a self-conscious being. But to be called a person requires even more. A person is able to conceptualize her own perspective and, because of this, to think of herself as a self. A person thinks of herself as the subject of her behaviour, thoughts, feelings and experiences. She thinks of herself as an agent, that is to say, as a being who wants to achieve her own purposes (by acting the way she does), has her own motives, and decides on her own actions by reasoning. These ways of thinking of oneself require a highly developed form of self-consciousness, or in other words, of the property of having a first-person perspective.[3] From now on, when I talk about a first-person perspective, I mean this highly developed form of a first-person perspective, which is fundamental for being a person.

Persons have cognitive and practical abilities that are based on a first-person perspective, for example[4]: (a) persons can think of themselves in the future and in other circumstances, and this means that they can imagine alternatives for their own futures; (b) persons are able to take a position regarding their motives for their actions and to reflect on these motives, and this means that they are capable of rational agency; and (c) persons can think of longer-term purposes and have the capability to think of (more or less complex) behavioural rules or guidelines that they may set for themselves, while pursuing a certain purpose, and this means that they are capable of a special form of reasoning which I label *structural reasoning*[5].

Structural reasoning is a highly developed form of instrumental reasoning, which only persons are capable of. Why is that? Firstly, it is a form of

[3] Self-consciousness, that is, the property of having a first-person perspective, may come in degrees, that is to say, non-human beings, like chimpanzees, may presumably have a certain form of a first-person perspective, but only beings that are able to conceptualize their own perspective and through this exist for themselves as selves are persons.

[4] It will emerge that the three mentioned abilities are crucial for a first-personal conception of practical irrationality. Regarding the first two examples, see also Baker 2000, p. 159. By the way, ability (c) is not independent of abilities (a) and (b); but I will not elaborate on that dependency here.

[5] Because the labels might look similar at first sight, I should mention that my notion of structural reasoning is not identical with Nida-Rümelin's notion of structural rationality as expounded in Nida-Rümelin 2001. Whereas structural reasoning, as I take it, is a highly developed form of instrumental reasoning, Nida-Rümelin's 'structural rationality' is a specific way of evaluating one's motivations and alternative actions, which he distinguishes from punctual optimalization.

instrumental reasoning, because the behavioural rules or guidelines, which the person sets for herself, are meant to achieve a certain purpose and are meant for the duration of the pursuit of that purpose;[6] and secondly, only beings that are persons are capable of structural reasoning, because setting behavioural rules or guidelines for oneself requires that the being in question (a) thinks of herself as a being who wants to achieve her own purposes and who decides on her own actions by reasoning; and (b) thinks of herself in the future.

On top of the property of having a first-person perspective, which is fundamental for being a person, most persons, or rather most human persons, have two other properties which are characteristic of (human) persons, namely the property of having a narrative self-conception and of making qualitative distinctions. A person with *a narrative self-conception* thinks of her life as an autobiographical narrative, that is to say, she interprets her experiences, traits, actions, practical beliefs and desires, connects them and positions them in relation to each other, and as a result of that, 'produces' the autobiographical narrative that she conceives of as her life. In short, a person with a narrative self-conception conceives of herself as the (developing) protagonist of her (developing) self-narrative. Such a person is able to think of long-term purposes for herself, or, in other words, of life projects or plans for herself, since thinking of life projects or plans for oneself requires that a person conceives of her life as a unity and as a developing entity. Furthermore, adopting long-term purposes requires complex structural reasoning, because the person has to think of behavioural rules which are meant to guide herself during a relatively long period of her life. Depending on the circumstances, she may have to adjust the behavioural guidelines she set for herself, but, as long as she wants to actualize those long-term purposes, they are meant to be taken into account in her reasoning.

The second property that I mentioned as being characteristic of (human) persons is the property of *making qualitative distinctions*. A person who makes qualitative distinctions uses a language of qualitative contrast, thinking of certain actions, purposes, kinds of persons, or kinds of life in terms of, for example, honourable and dishonourable, noble and despicable, brave and cowardly, self-disciplined and impulsive, caring and selfish, or merciful and merciless. Such a person conceives of herself as a person who has certain

---

[6] The capability for structural reasoning is not only required when initially thinking of a certain longer-term purpose the required actions for which can, at that moment, not be thought through in detail (for instance, due to the complexity of the required action structure, the expectation of various possible influencing conditions, or the need to acquire more information or knowledge), but also during the various stages of the pursuit of that purpose. After all, when a person initially has set behavioural rules for herself to support the achievement of a certain longer-term purpose, she has to take those behavioural rules into account when she, at various points in time, during her instrumental reasoning about appropriate actions for the benefit of that (longer-term) purpose, thinks and decides about concrete actions; it might also become clear then that certain behavioural rules are not appropriate (anymore) and therefore have to be adjusted or abandoned, or that new behavioural rules or guidelines are required (at that stage of the pursuit of the purpose in question). I refer also to Section 4 (Our evaluative agenda).

values, and, in general, she will be sensitive to situations where those values are at stake. To protect or realize her values, she may also set certain behavioural guidelines for herself. She knows that the values in question will only be values of herself as long as she lives up to these values. In other words, she will have to continuously re-affirm her values in actual situations of action, by the qualitative distinctions which she then makes and which she applies as qualitative standards to herself in her reasoning about the action in question.[7]

In a certain way values seem to be like long-term purposes, because, from both of them, we may derive behavioural rules or guidelines, which we set for ourselves and which are meant to be taken into account in our reasoning; but in another way they are different: purposes can be achieved whereas values cannot. For example, if a person has the ambition to become a surgeon, she will conceive of herself as having achieved this purpose as soon as she has received the required medical degrees, obtained the required practice as assistant physician, and, finally, found a job as a surgeon. In contrast, if the same person wants to be the kind of person who is honest to others, she will never conceive of her value of honesty as being finally realized. In actual situations, she may conceive of herself as acting or having acted in accordance with her value, but she will have to take her value into account throughout her life, at least as long as she still conceives of herself as the kind of person who cherishes this value. In short, purposes can be achieved, whereas values have to be re-affirmed, by acting the way we do.

Although I used the notion of reasoning in various contexts, namely evaluative reasoning, instrumental reasoning, structural reasoning, and taking something into account in one's reasoning, I did not thoroughly explain my view of reasoning. The next three sections will be dedicated to the subject of reasoning broadly construed.

2 Two sorts of evaluative reasoning

In my view, there are two basic sorts of reasoning: *non-evaluative reasoning* and *evaluative reasoning*. Non-evaluative reasoning is (more or less complex) instrumental reasoning taking a certain motivating reason, that is, a certain purpose[8], as given; in other words, it is a process of deciding how to achieve a given purpose. Evaluative reasoning, even though it involves instrumental reasoning too, is concerned with the question which purpose to pursue and, in case of alternatives to actualize that purpose, which method of realization to choose. This process of deciding what to do and how to do it is (more or less) iterative, because we do not only evaluate the motivating reason (to act)

---

[7] See also my discussion of the notion of strong evaluation in Section 2 (Two sorts of evaluative reasoning).

[8] I owe the idea to take the motivating reason for an action as its purpose, or – in the terminology of Michael Smith – as its goal, to Michael Smith, see Smith 1994, pp. 116, 125.

but also the results of our instrumental or structural reasoning (that is to say, the corresponding action structures and, if appropriate, the behavioural rules we are considering), which in turn may lead to a re-evaluation of the motivating reason, further instrumental or structural reasoning, or both. While all agents, whether they are conscious of their reasoning or not, are capable of (more or less complex) non-evaluative reasoning, only persons are capable of evaluative reasoning and of applying structural reasoning during such processes. After all, all agents are capable of (more or less complex) instrumental reasoning, but only persons have the ability (a) to think of themselves as beings which decide on their actions by their own reasoning, (b) to take a position regarding their purposes and alternative actions, and to reflect on them, and (c) to think of themselves in the future; and these three abilities are required to evaluate one's own purposes and alternative actions, and to set behavioural rules or guidelines for oneself (the latter being characteristic of structural reasoning).

Persons may apply either weak or strong evaluation when evaluating their purposes or the results of their instrumental or structural reasoning. When applying *weak evaluation* "motivations or desires count in virtue of the attraction of the consummations"[9], that is to say, when confronted with conflicting motivations, a person's decision to reject or postpone the other motivations for action in favour of one of them is based on her preference for the achievement of the corresponding purpose. When applying *strong evaluation*, however, "[m]otivations or desires do not only count in virtue of the attraction of the consummations, but also in virtue of the kind of life and kind of subject that these desires properly belong to"[10]; that is to say, a person's decision to reject a certain desire may be based on the fact that she does not want to be the kind of person who is moved by that desire; she might, for example, think of such a person as being selfish, whereas she wants to be the kind of person who cares for others.

Are all persons capable of both weak and strong evaluation? No. All persons are capable of weak evaluation because, owing to a first-person perspective, they are capable of reflecting on their motivating reasons for action (that is to say, their probable action purposes) and of rejecting certain purposes in favour of others, which they, weighing the alternatives, prefer to actualize. And because all persons can think of themselves in the future and in other circumstances, a person may even give more weight to certain longer-term purposes at the cost of the probable consummations of a number of short-term desires. In short, all persons are capable of weak evaluation, because it belongs to their capability of (weak) rational agency. But only persons that have the property of making qualitative distinctions are capable

---

[9] This clause belongs to the by now well-known citation: "Motivations or desires do not only count in virtue of the attraction of the consummations, but also in virtue of the kind of life and kind of subject that these desires properly belong to." (Taylor 1985, p. 25)

[10] Taylor 1985, p. 25.

of strong evaluation, because to ask oneself the question 'What kind of person would I be if I were moved by this motivation to act in a certain way?' requires that the person in question assesses motivating reasons and kinds of life in terms of 'higher' and 'lower'. Such a person may reject a certain desire or possible action, not only because it is contingently incompatible with another one which this person gives more weight to – as is the case with weak evaluation – but because she judges that being moved by that desire or performing that action would mean sacrificing (some of) her values.

In general, the decision to pursue a certain purpose may involve all mentioned sorts of reasoning: instrumental and structural reasoning, weak and strong evaluation. This does not mean that persons always apply the sort of reasoning which would be most applicable for them in their actual situation. For example, depending on the internal (like mood or physical conditions) and external circumstances (like interfering actions of other persons), a person might not apply strong evaluation even if her values are at stake. And a person may not apply structural reasoning even if the long(er)-term purpose she aims to achieve required her to do so.

To understand evaluation processes in their full complexity, we have to focus on the concept of action in more detail.

3 Actions and action structures

To achieve a certain purpose, we have to think of and perform appropriate actions. Normally, when we decide what to do, that is, to pursue a certain purpose, we also decide (more or less precisely) how to do it, that is, which action structure we will perform to actualize our purpose. The term *action structure* is more appropriate here than the term action, because what we generally call actions are in fact (more or less complex) action structures. When a person is engaged in the performance of a certain action structure (for the achievement of a certain purpose), she applies (whether consciously or not) instrumental reasoning to 'construct' that action structure[11]. If the person *non-consciously* decided on how she is doing what she is doing, then, for her, that action structure must be a well-known *action building block*. By contrast, action structures that are complex or new for the person in question are derived from the purpose in question by *conscious* instrumental reasoning. The person will stop that process of instrumental reasoning as soon as she, when breaking down the purpose into its constituent purposes, arrives at well-known action building blocks or thinks that she can fill in the

[11] Normally, action structures are not strictly sequential, but also consist of iterative and case structures. An action structure (or a certain part of it) has an iterative structure when a certain action (structure) has to be performed repeatedly, until a certain condition is fulfilled; it has a case structure when either action (structure) '1', '2', …, or 'n' has to be performed, depending on a certain condition.

details of the action structure at a later stage. Furthermore, when a person decides to pursue a certain purpose by performing a certain action structure, this does not mean that she starts executing the action structure immediately. Action structures may be scheduled to be executed at a certain moment, or the execution may be made dependent on certain internal (for example, the existence of other purposes that would also be supported by that action structure) or external conditions (for example, the availability of required resources, or supporting actions of other persons). In short, the result of a process of instrumental reasoning is a more or less detailed action structure, which the person, after or without further evaluation of the action building blocks involved, plans to perform at a certain moment or under certain conditions.

The performance of a certain action structure has a certain purpose, which is the motivating reason of the person in question to perform that action structure. The action building blocks that are part of that high-level action structure also have their own purposes, but these are *derived purposes*, that is to say, they result from breaking down a certain 'final purpose' into its constituent purposes, and only become purposes of the person because their actualization is required to achieve the purpose of the high-level action structure.[12] In other words, a derived purpose functions as an 'auxiliary purpose' with respect to a certain 'final purpose'. And a person who uses her capabilities for evaluative reasoning to the full will not only evaluate her non-derived purposes but also her derived purposes, which in turn may lead to a re-evaluation of the original motivating reason.

Sometimes persons think of purposes the realization of which cannot be thought of and planned in detail by the construction of appropriate action structures. Those purposes are long(er)-term purposes and require that the person in question sets certain behavioural rules or guidelines for herself for the duration of the pursuit of those purposes.

As I have argued, persons can think of complex action structures, which they plan to perform to achieve a certain purpose. They may even schedule the execution of those action structures (or parts of it) at a certain moment or make the execution dependent on certain conditions. Furthermore, they may think of and set behavioural rules or guidelines for themselves to support the achievement of their long(er)-term purposes. The question arises: How do persons take all their planned action structures and behavioural rules or guidelines into account when deciding on alternative actions? The next section will be concerned with that question.

---

[12] A certain action building block may even be part of more than one high-level action structure; furthermore, its purpose may be derived as well as non-derived.

4 Our evaluative agenda

When a person, as a result of her (conscious) evaluative reasoning, decides to pursue a certain purpose, she puts that purpose on her *evaluative agenda*[13], so to speak, because she may have to take this purpose into account in other situations where she again applies evaluative reasoning. The purpose in question will stay on her evaluative agenda until it is achieved or until, during a further process of evaluative reasoning, she decides to abandon or adjust that purpose, or to remove it from her evaluative agenda. Besides her purposes, a person's evaluative agenda also consists of the action structures the person has thought of to achieve her purposes, the behavioural rules or guidelines she may have set for herself to actualize her long(er)-term purposes, and the behavioural guidelines she may have adopted to protect her values. The action structures and the purpose-related behavioural rules or guidelines will stay there until the corresponding purposes are achieved, abandoned, or adjusted, or until the person adjusts the action structures or behavioural guidelines. The value-related behavioural guidelines will stay there until those values are not her values anymore.

Why do persons sometimes abandon or adjust the purposes or action structures, or modify the behavioural guidelines (or rules), which they decided on during earlier processes of evaluative reasoning? Here is a series of examples. (1) A person may abandon or adjust a certain purpose because, in her current process of weak evaluative reasoning, she gives more weight to another purpose, the achievement of which would be made impossible or hindered by pursuing the first one. (2) A person may abandon or adjust a particular purpose because she thinks that the pursuit of that purpose would prevent her from actualizing another purpose, which she, in her current process of strong evaluative reasoning, regards as fitting the kind of person that she judges to be qualitatively 'higher' than the one that might continue to pursue the first purpose. (3) A person may abandon a certain purpose because she recognizes that she is not able to execute the corresponding action structure or to act according to the corresponding behavioural rules (anymore). (4) A person may abandon or adjust a certain purpose because, during performance of its action structure, she recognizes that she 'sacrifices' her values by acting as she does. (5) A person may adjust the action structure or behavioural guidelines which belong to a certain purpose because she recognizes that the action structure or behavioural guidelines are not appropriate (anymore) to support the actualization of the purpose they were initially meant for. (6) A person may adjust the schedule for the execution of a certain action structure because the execution of that action structure would conflict with that of another, the corresponding purpose of which she has decided to achieve first, in her current process of evaluative reasoning.

---

[13] I owe the idea for the label 'evaluative agenda' to T.M. Scanlon, who briefly introduces the term 'deliberative agenda', see Scanlon 1998, pp. 46-47.

(7) A person may adjust the external conditions for the execution of a certain action structure because she recognizes, for example, that additional resources are needed for the execution of that action structure. (8) A person may adjust the internal conditions for the execution of a certain action structure because she recognizes, for example, that a certain bodily condition is required for the execution of that action structure, and this in turn may lead to new or adjusted behavioural rules or guidelines which the person sets for herself. (9) A person may remove certain value-related behavioural guidelines from her evaluative agenda because she recognizes that the value which these guidelines stand for does no longer count in her actual processes of strong evaluative reasoning.

Analysing the examples given, we discover that, in some of the mentioned examples, the update of the person's evaluative agenda takes place when the person decides to pursue a new purpose the pursuit of which would, in one way or other, be hindered by the pursuit of purposes that are on her current evaluative agenda, whereas, in other examples, the update is not triggered by a new purpose, but takes place, for example, because the person becomes more knowledgeable about a certain subject or her own capabilities, or because she recognizes that the action structure she chose conflicts with her values. This observation, together with the already mentioned facts that persons put new purposes they have decided to achieve on their evaluative agenda and remove certain purposes (together with the corresponding action structures, and behavioural rules or guidelines) from their evaluative agenda when the purposes in question have been achieved, shows that a person's evaluative agenda is a *planning and monitoring medium* for the purposes we decide or have decided to actualize. Therefore our evaluative agenda is subject to our continuous more or less radical adjustment (whether consciously or not).

But our evaluative agenda is not only the planning and monitoring medium for the purposes we have decided to pursue, it also functions as a *reference base* when we reflect upon our motivations and deliberate about our options for action, in a concrete situation. Why is that? When we have to decide how to act and, by conceiving of the situation in this way, are engaged in a process of evaluative reasoning, we assess our current desires or motivations, which could move us to act in a certain way, by applying weak or strong evaluation. In the first case (that is, when applying weak evaluation), our current desires or motivations (and the required action structures to achieve the respective purposes), besides being weighed against each other, are also weighed against our earlier adopted purposes (and the corresponding action structures, or behavioural rules or guidelines), for putting a certain purpose on one's evaluative agenda is a commitment to oneself[14] to pursue that purpose. We

[14] Committing oneself, first-personally, to pursue a certain purpose means to pursue that purpose as long as one does not decide to adjust or abandon the pursuit of the purpose in question; and this implies that one has to evaluate in concrete situations whether the action one is considering might hinder or obstruct one or more of one's adopted purposes. – I was inspired by Christine Korsgaard's article 'The normativity of instrumental reason' to see one's adopted purposes as commitments to oneself; see Korsgaard 1997, pp. 215-254, especially note 69 on pp. 248-249.

may conclude that the new purpose under consideration (1) does not conflict with purposes that are on our evaluative agenda already, and therefore could be added to it; or (2) conflicts with one or more purposes that are on our evaluative agenda already, and therefore could only be added to it if we removed, abandoned or adjusted one or more of the already adopted ones; or (3) conflicts with action structures or behavioural rules or guidelines that belong to certain purposes which are on our evaluative agenda already, and therefore could only be added to it if we were to remove, abandon or adjust one or more of those purposes, or if we, for one or more of those purposes, could determine alternative action structures that do not conflict with our new purpose.

In the second case, that is, when applying strong evaluation, we may equally use our evaluative agenda as a reference base, but in another and more limited way. We assess our current motivations and alternative actions either (a) by asking whether they conflict with our values, more specifically, by asking ourselves whether we want to be the kind of person that is moved by these motivations or that acts in the ways we are considering; or (b) by asking whether they conflict with the value-related behavioural guidelines which we adopted and which are therefore on our evaluative agenda, more specifically, by asking ourselves whether the motivations and alternative actions under assessment would fit the kind of person that lives in accordance with the behavioural guidelines in question; or (c) by a combination of both methods. We may conclude that the motivations and/or actions which are under assessment: (1) do not conflict with the value-related behavioural guidelines that are on our evaluative agenda, and that therefore the purpose in question could be added to our evaluative agenda; or (2) conflict with the value-related behavioural rules that are on our evaluative agenda, and therefore have to be rejected; or (3) conflict with our values, because we do not want to be the kind of person that is moved by this motivation, or that acts in the way we are considering, and that therefore the motivation and/or action under assessment has to be rejected, even though it does not conflict with the value-related behavioural guidelines that are on our evaluative agenda; this judgement may mean that our values have changed and might therefore lead to a reflection on and adjustment of certain behavioural guidelines which are meant to stand for our values; or (4) do not conflict with our values, because being moved by the motivation in question and acting in the way we are considering accord with the kind of person we (now) want to be, and that therefore the purpose in question could be added to our evaluative agenda, although the motivation under assessment conflicts with certain value-related behavioural guidelines that are on our evaluative agenda; this judgement may mean that our values have changed and therefore might lead to a reflection on and adjustment of certain value-related behavioural guidelines on our evaluative agenda.

In conclusion, when we are engaged in processes of evaluative reasoning, whether we apply weak or strong evaluation, our evaluative agenda functions

as a reference base. In the case of weak evaluation, we take our adopted purposes, their action structures, and the behavioural rules or guidelines which are meant to support the achievement of our long(er)-term purposes into account by weighing the motivations and alternative actions which are under consideration against them. In the case of strong evaluation, we take our values into account in one (or both) of the following ways. We may take them into account *indirectly* (that is to say, via behavioural guidelines that we have derived from our values and which are on our evaluative agenda) by asking ourselves whether the motivations and alternative actions under assessment fit the kind of person that lives in accordance with certain value-related behavioural guidelines that we have set for ourselves, or we may take them into account *directly* (that is to say, without referring to value-related behavioural guidelines that may be on our evaluative agenda) by asking ourselves whether we want to be the kind of person that is moved by the motivations under consideration or that acts in the ways we are considering.

Processes of evaluative reasoning may be simple or complex. The process is simple, for example, if a certain possible motivation for action (and the required action structure) is 'put to the test' once and, as a result, is adopted or rejected. Such a process of evaluative reasoning may take place non-consciously. A process of evaluative reasoning is extremely complex, and will take place consciously, if a number of iterations is needed before a person can decide what to do and how to do it; the person may, for example, have to adjust action structures for the achievement of some adopted longer-term purposes, to abandon an already adopted long-term purpose, and to set behavioural rules for herself to support the achievement of the new purpose she wants to actualize.

After having elaborated extensively on evaluative reasoning and, as part of that, having argued my view on the involved notions of agent and person, instrumental and structural reasoning, weak and strong evaluation, and action and action structure, and having introduced the concept of one's evaluative agenda, I will now deal with the notion of 'practical irrationality from a self-related perspective'.

5 Different ways of conceiving of one's own actions, regarding aspects of rational agency

The phenomenon of conceiving of one's actions as irrational has to be distinguished from the phenomena of conceiving of one's actions as mistaken, non-rational, or wrong, and, obviously, from the phenomenon of conceiving of them as rational. In the introduction of this chapter, I sketched a person who has put herself on a diet, and who, at a certain moment, eats chocolate. Let me use this example as a framework which, without claiming to be complete, can serve to illustrate and distinguish different ways in which persons may conceive of their actions (regarding aspects of rational agency).

For each case, I will first present the 'example-person' (EP) conceiving of her action of eating chocolate in a specific way, and then, abstracting from the example, describe that way of conceiving of one's action in general terms. By the way, I do not claim that that (general) description covers all variants of conceiving of one's action in the way under discussion.

Case 1: Persons may conceive of their actions as *rational*.
Example: EP may conceive of her action as rational, because she knows that she deliberated whether she should let herself be moved by her desire to eat chocolate and that she finally decided to eat chocolate.
Generally: A person may conceive of her action as rational if she was engaged in a process of evaluative reasoning about that action and acts (or acted) in accordance with her decision.

Case 2: Persons may conceive of their actions as *irrational*.
Example: EP may conceive of her action as irrational, because she knows that she decided (and still wants) to lose weight and therefore has put herself on a diet, and she is aware of the fact that she violates her diet by eating chocolate.
Generally: A person may conceive of her action as irrational if the action hinders or prevents the achievement of an adopted purpose that she, during her evaluative reasoning about the action in question, would rather have given more weight to (for example, a longer-term purpose, which she had adopted and had defined behavioural rules for).

Case 3: Persons may conceive of their actions as *rational as well as irrational*.
Example: EP may conceive of her action as rational as well as irrational, because she knows that she deliberated whether to eat chocolate or not and that she decided to do so, and, at the same time, she knows that she violates the objective she has set for herself to lose weight, which is a goal that she really wants to actualize.
Generally: A person may conceive of her action as rational as well as irrational if she, on the one hand, was engaged in a process of evaluative reasoning about that action and acts (or acted) in accordance with her decision, and, on the other hand, knows that her action hinders or prevents the achievement of an adopted purpose that she, during her evaluative reasoning about the action in question, would rather have given more weight to (for example, a longer-term purpose which she had adopted and had defined behavioural rules for).

Case 4: Persons may conceive of their actions as *non-rational*.
Example: EP may conceive of her action as non-rational, because she, having woken up at night and having got out of bed half-asleep, 'finds herself' eating chocolate.
Generally: A person may conceive of her action as non-rational if she (thinks

that she) was not engaged in a process of evaluative reasoning about that action, in other words, if (she thinks that) the action happens (or happened) to her, so to speak.

Case 5: Persons may conceive of their actions as *mistaken*.
Example: EP may conceive of her action as mistaken, because, when deliberating about eating chocolate, although she had decided to follow a special fat-free diet, she reasoned that she could compensate for eating the piece of chocolate by taking the calories of the piece of chocolate into account when having her next meal. After having eaten the chocolate, however, she recognizes that she made a mistake in her reasoning, which would have been appropriate for a low-calorie diet, but not for a fat-free one.
Generally: A person may conceive of her action as mistaken if she recognizes that she made mistakes in her instrumental reasoning (and more or less knows the mistakes), in other words, if she recognizes that the action structure she decided on in order to achieve her purpose or to adhere to the behavioural rules she had defined for herself to achieve a certain long(er)-term purpose is not appropriate (and if she more or less knows the mistakes in her reasoning).

Case 6: Persons may conceive of their actions as *wrong*.
Example: EP may conceive of her action as wrong, because she has eaten all the chocolate, whereas it was meant for this afternoon's birthday party of her little daughter. The fact that, by eating the chocolate, she hindered the achievement of her goal to lose weight is incomparably less important to her than the fact that, by acting the way she did, she cannot conceive of herself as a loving mother.
Generally: A person may conceive of her action as wrong if she does not want to be the kind of person who acts as she does (or did), in other words, if she recognizes that she 'sacrifices' (or 'sacrificed') her values; and this in turn means that she recognizes that she should have engaged in strong evaluation, because (one or some of) her values were at stake. Conceiving of an action as wrong seems to 'override' other possible ways of conceiving of that action, because, when conceiving of an action as wrong, it seems that it does not matter to us anymore whether this action, for example, hinders one of our longer-term purposes.

Summarizing the presented ways of conceiving of one's own actions, we may say: (1) To conceive of a certain action of oneself as rational or irrational means, firstly, that we conceive of the action as resulting from our evaluative reasoning, and, secondly, that we regard that action as respectively supporting or hindering the achievement of a certain purpose. (2) Because we can have decided on an action the purpose of which hinders or prevents the achievement of an adopted purpose that we (now) would rather have given more weight to, we may conceive of a certain action as rational as well

as irrational. (3) To conceive of a certain action of oneself as non-rational means that we (think that we) did not engage in (conscious) evaluative reasoning about that action. (4) To conceive of a certain action of oneself as mistaken means that we know that we made mistakes in our instrumental reasoning about that action. (5) To conceive of a certain action of oneself as wrong means that we do not want to be the kind of person who acts as we did (or do). Conceiving of one's action as wrong seems to 'override' the other possible ways of conceiving of the action in question.[15]

Having classified the various ways in which persons may conceive of their own actions (regarding aspects of rational agency) and, by way of that analysis, having narrowed down the possible interpretations of conceiving of one's own action as irrational, I will, from now on, only be concerned with the notions of acting irrationally and (later on) of being irrational.

6 Acting irrationally from a self-related perspective

At this stage of the analysis, our thesis about the notion of 'acting irrationally from a self-related perspective' may be as follows:

(T1) Conceiving of one's own action as irrational means (1) that we conceive of the action as resulting from our (conscious) evaluative reasoning, and (2) that we regard that action as hindering or preventing the achievement of a certain adopted purpose which we would rather have given more weight to (than the one we decided on) during our evaluative reasoning about the action in question.

The first condition implies that actions which are not conceived of as resulting from one's evaluative reasoning will, first-personally, not be conceived of as irrational. The second condition makes clear that conceiving of one's actions as irrational is an assessment in which we refer to specific adopted purposes, which function as standards, so to speak. Let me elucidate the (scope of the) thesis further by discussing a number of possible questions. I will first be concerned with the thesis's first condition, and then, in questions 4 to 8, with its second one.

Question 1: Why exactly is the first condition necessary? If we conceive of certain of our actions as not resulting from our evaluative reasoning, then we conceive of those actions neither as rational nor as irrational. To avoid

---

[15] Because this chapter is not concerned with the notion of conceiving of one's action as wrong as such, but explains that notion for the benefit of clarification of the notion of conceiving of one's own action as irrational, I do here not further discuss the phenomenon that conceiving of one's action as wrong seems to override other possible ways of conceiving of that action. Regarding the overriding nature of the self-related conception of wrongness of action, see Chapter 4 (Wrongness from a self-related perspective).

misinterpretation: conceiving of an action as resulting from one's evaluative reasoning does not imply that we were conscious of our evaluative reasoning before or during the performance of the action, but it means that we conceive of the action as one that we do or did deliberately, as opposed to an action that we conceive of as just happening to us. As examples of the latter, think of 'playing' with your ballpoint or your glasses while listening to a lecture.

Question 2: How can we know that we were non-consciously engaged in a process of evaluative reasoning about a certain action if we did not consciously deliberate about that action and its possible alternatives? The point is not that we, in such a situation, have to know for sure that we were non-consciously engaged in a process of evaluative reasoning about a certain action in order to assess that action in terms of rational or irrational, but that we conceive of the action as one that we did deliberately, as opposed to an action that we conceive of as just happening to us. In other words, if we conceived of that action as just happening to us, without our capability of evaluative reasoning being involved, then we would neither conceive of that action as irrational nor as rational.

Question 3: Do we assess an action of ourselves in terms of irrational or rational if someone put pressure on us to perform that action? Probably yes. If we conceived of that situation as still offering certain alternatives, then we in fact were engaged in evaluative reasoning and might assess our behaviour in terms of rational or irrational. Otherwise, that is, if we conceived of the situation as not offering any alternatives, for example, because we were violently forced to do what we did, then we would not assess our action in terms of rational or irrational.[16]

Question 4: At what moment, compared with the moment or period of performing an action, do we conceive of that action as irrational? A person may conceive of her action as irrational at three occasions: in retrospect, during the performance of her action, and right from the beginning of her action. She may conceive of her action as irrational *in retrospect* if she, after having achieved the purpose of that action, realizes that she would rather have decided otherwise. She may conceive of her action as irrational *during the performance* of that action if she, while busy executing that action structure, realizes that she would rather have decided otherwise. Depending on the situation, she may decide to abandon the performance of her current action structure for the benefit of the adopted purpose, the achievement of which would otherwise be hindered or prevented. She may conceive of her action as irrational *right from the beginning* of the performance of that action structure if she, during her process of evaluative reasoning, realizes that she in fact gives most weight to a purpose the decision for which (and thereby its corresponding action structure) she, at the same time, judges to be irrational. I believe that most of our actions which we conceive of as irrational belong to this category. In such cases we, on the one hand, deliberately decide to pursue

[16] We may, nevertheless, conceive of our action as violating our values.

a certain purpose and act accordingly, and, on the other hand, know that deciding and acting as we do is irrational.

Question 5: Is the purpose we have given more weight to necessarily a new purpose? Although that purpose may often be based on a strong spontaneous desire and so not be on our evaluative agenda already, this is not necessarily the case. The purpose we have decided to pursue may have been (together with the required action structure) on our evaluative agenda already, but we, for example, originally planned to pursue that purpose only in situations where the achievement of a certain adopted long-term purpose would not be hindered.

Question 6: Can the purpose that we, when we judge our action to be irrational, are referring to as the one that we would rather have given more weight to (than to the one we decided on) be a new purpose, that is, a purpose that was not on our evaluative agenda already? No, this is impossible. The purpose which we are referring to when we assess our action as irrational has to be an adopted purpose, and all our adopted purposes are part of our evaluative agenda. This does not exclude the possibility that a person in retrospect would rather have given more weight to another of the (contingently) conflicting new possible purposes (and alternative actions) that she took into account during her evaluative reasoning. But because that purpose was not already adopted as a result of a former process of evaluative reasoning and therefore not part of the person's evaluative agenda, it cannot function as a 'standard' during, what we might call, an irrationality assessment.

Question 7: Do we conceive of all actions which hinder or obstruct one of our adopted purposes and which we conceive of as resulting from our evaluative reasoning as irrational? No. To conceive of an action as irrational not only implies that we view an action as hindering or preventing the achievement of a certain adopted purpose, and as resulting from our evaluative reasoning, but it also implies that we would rather have given more weight to that adopted purpose during our evaluative reasoning about the action in question. When we, for example, decided to perform an action that hinders one of our adopted purposes, because we did (and do) not want to be the kind of person that, in the situation in question, pursues that earlier adopted purpose, then we do *not* conceive of our action as irrational, since we do not want to have decided otherwise.

Question 8: What does our action exactly mean for the adopted purpose the achievement of which we hinder(ed) or prevent(ed) with that action? It means that we have to adjust the adopted purpose in question, adjust the action structure that we planned to perform to achieve that purpose, adjust its schedule conditions, adjust or remove other adopted purposes the action structures of which could interfere with the adjusted purpose in question, or even remove that purpose from our evaluative agenda. But why should we? Because otherwise we will end up with an incoherent evaluative agenda, and this, in turn, might mean that we not only conceive of our action(s) as irrational, but that we, in the end, conceive of ourselves as irrational. This

statement obviously calls for further explanation; it will be discussed in the next section and in Section 9.

In conclusion, we can fine-tune our thesis about the notion of 'acting irrationally from a self-related perspective', and we may state:

(T2) Conceiving of one's own action as irrational means (1) that we conceive of the action as resulting from our (conscious) evaluative reasoning, and (2) that we (retrospectively, during the performance of the action, or right from the beginning) view that action as hindering or preventing the achievement of a certain adopted purpose which we would rather have given more weight to than to the (new) one we decided on (or are going to decide on) during our evaluative reasoning about the action in question.

7 Being irrational from a self-related perspective

Conceiving of one's own action as irrational involves conceiving of the corresponding decision to perform the action as irrational. After all, the decision to pursue the purpose in question (and to perform the corresponding action structure) conflicted with a purpose (its corresponding action structure and/or behavioural rules) which we had adopted in a former process of evaluative reasoning and which therefore was on our evaluative agenda when we made the decision in question, and which we would rather have given more weight to than to the (new) one we decided on during our evaluative reasoning about the action in question. As for situations where a person conceives of her action as irrational, a person may conceive of a decision (to act in a certain way) as irrational after performing the action she has decided on, during the performance of that action, or right from the beginning (that is, she already views the decision as irrational during her process of evaluative reasoning). In conclusion, we can legitimately say:

(T3) Conceiving of one's own action as irrational involves conceiving of the corresponding decision to perform that action as irrational.

If a person conceives of one of her decisions as irrational, she might also conceive of herself as irrational. Why is that? Because we think of ourselves as the subject of our decisions, and of our decisions as being part of our own evaluative reasoning, and because, in our self-conception, we cannot separate our own evaluative reasoning from ourselves, nor can we, in our self-conception, separate our decisions from ourselves. Therefore, we can state:

(T4) If a person conceives of one of her decisions as irrational, she might also conceive of herself as irrational, because, in our self-conception, we cannot separate our decisions from ourselves.

The course of my argument for this thesis may require an explanation. As I argued in Section 1 (Agents and persons), persons are beings that have a first-person perspective and that thereby conceive of themselves as themselves and exist for themselves as selves. Conversely, all beings that exist for themselves as selves and hence conceive of themselves as themselves, have a first-person perspective and are, therefore, persons. The capacity of persons to take a position towards their motives for action and to reflect on them, in other words, the capacity for evaluative reasoning, is based on the property of having a first-person perspective. Thus, only persons have the property of evaluative reasoning. And because the capacity for evaluative reasoning is based on that property which is fundamental for conceiving of oneself as a self, we can, in our self-conception, not separate our evaluative reasoning from ourselves. Does this imply that, in our self-conception, we can separate certain other capacities from ourselves? Yes, indeed. In our self-conception, we can separate all those capacities from ourselves which are not based on our property of having a first-person perspective, because we will not cease to exist (as a self) if we lost (some or all of) them. This is not to say that we do not conceive of those capabilities as our own capabilities, but they do not directly belong to our personhood. For example, our capacities to walk, make noise, hear, or taste are not based on our property of having a first-person perspective; therefore, if we lost (some or all of) these capacities, we would still exist as selves. But if the organism that now constitutes a person lost the capacity for evaluative reasoning, because it lost the property of a first-person perspective, then that organism would not constitute a person anymore.[17] Therefore, in our self-conception, we cannot separate our capacity for evaluative reasoning from ourselves, and because we conceive of our decisions as being part of our own evaluative reasoning, we cannot separate our decisions from ourselves.

Let us return to the thesis T4: If a person conceives of one of her decisions as irrational, she might also conceive of herself as irrational. This statement does obviously not imply that a person *will* (or even should) conceive of herself as irrational if she conceives of certain of her actions as irrational. She probably will not conceive of herself as irrational if she occasionally conceives of certain of her actions as irrational and can deal with the effects of those actions on her adopted (long(er)-term) purposes that she still wants to achieve. However, she might conceive of herself as irrational if she, for example: (1) regularly conceives of her own actions as irrational; (2) frequently has to think of and perform 'repair-actions' to undo the hindering or damaging effects of her irrational actions, because she still wants to achieve the purpose the actualization of which is hindered by her irrational action(s); (3) regularly has to (that is to say, feels the need to) rework her evaluative agenda to resolve any incoherence between her adopted purposes (and their corresponding action structures); or (4) has to (that is to say, feels the need to) remove long(er)-term purposes from her evaluative agenda, because she

[17] I use Baker's 'constitution view' of persons; see Baker 2000, for example, p. 91 or 116.

adopted purposes (and performed the corresponding action structures) which obstructed those earlier adopted long(er)-term purposes.

The aforementioned examples have in common that the person in question does not conceive of herself as irrational along with conceiving of one of her actions as irrational, but she, for example, conceives of herself as irrational after having conceived of her actions as irrational a number of times, or after having been confronted with the need to (more or less radically) rework her evaluative agenda; thus she conceives of herself as irrational as a result of reflection on herself qua self. We may say: Conceiving of oneself as irrational is a possible result of reflecting on oneself qua self. I believe that, even though a person may conceive of herself as irrational based on *one* action of herself which she conceives of as irrational, her conception of being irrational is not a direct result of her conceiving of that action as irrational, but an indirect one, that is to say, her conceiving of a certain action of herself as irrational has led her to reflect on herself[18], which in turn resulted in her conception of being irrational.

Taking these considerations into account, and combining the two theses T3 and T4, we arrive at the final thesis of this section:

(T5) Conceiving of one's own action as irrational involves conceiving of the corresponding decision to perform that action as irrational; and if a person conceives of one of her decisions as irrational, she might – via reflection on herself qua self – conceive of herself as irrational, because, in our self-conception, we cannot separate our decisions from ourselves.

But why do we bother at all whether our actions are irrational?

Part II – Why acting irrationally or being irrational matters to us

After having analysed the notion of conceiving of one's own action as irrational and having shed light on the notion of conceiving of oneself as being irrational, the question remains: Why does it matter to us whether we conceive of certain of our actions as irrational or of ourselves as being irrational? I claim: The brute fact that we act(ed) irrationally or *are* irrational bothers us, even apart from the fact that conceiving of one's own actions or of oneself as irrational means that we occasionally or regularly do not achieve the purposes we really want(ed) to achieve.

[18] As I will argue in Section 9 (Being irrational matters to us - the concern with coherence of our evaluative agenda), it requires a (certain degree of) narrative self-conception to reflect on oneself in the relevant way.

8 Acting irrationally matters to us – the concern with the soundness of our will

A person who conceives of her action as irrational did (or does) not do what she *really* wants to do, since according to our thesis T2, she (retrospectively, during the performance of her action, or right from the beginning) would rather have given (or rather give) more weight to another (new) adopted purpose than to the one she decided on (or is going to decide on), during her evaluative reasoning about the action in question. On the other hand, she herself decided (or is going to decide) to perform that action, because, and again according to our thesis T2, she conceives of her action as resulting from her (conscious) evaluative reasoning. And if a person decides to pursue a certain purpose (and to perform the corresponding action structure), then this person is moved by a desire or motivation (to act in certain way) which she, at that moment, actually wants to be moved by. Therefore we might say: A person who conceives of a certain action of herself as irrational did what she actually wanted to do at that moment, and yet, at the same time, she did not do what she really (would have) wanted to do. This is an unsurprising conclusion, because an action that we conceive of as irrational is irrational with respect to a certain earlier adopted purpose which functions as a standard in our assessment of irrationality, but our action is rational with respect to the purpose we actually have decided to pursue. Although the conclusion may well be unsurprising, it helps to illuminate our question 'Why does it matter to us whether we conceive of certain of our actions as irrational?'.

A person who, according to her self-conception, did what she actually wanted to do at that moment, and who, at the same time, did not do what she really (would have) wanted to do, cannot conceive of her will as a unity. Why are we concerned with that unity? Persons, because of their capability of rational agency, conceive of themselves as rational agents. Rational agents, when they act rationally, are motivated by reasons (to act in a certain way) which they want to be motivated by.[19] The problem with a person who

---

[19] This definition is qua formula inspired by Lynne Rudder Baker's notion of 'acting based on reasons of one's own' (see Baker 2004; Lynne Rudder Baker presented a preliminary version of that paper at the conference *Reasons of One's Own*, held at the University of Utrecht, The Netherlands, 25-28 April 2001), but the definitions are not identical qua meaning. Whereas Baker distinguishes between a person's acceptance and a person's acceptance cum approval of a certain reason for her to perform a certain action, and based on this distinction, distinguishes between actions that are based on reasons and actions that are based on reasons of one's own, I do not make this distinction. As I take it, a person who conceives of a certain action as resulting from her own evaluative reasoning wanted – at least at the moment when she actually decided to perform that action – to be moved by the motivating reason which she conceives of as the reason for her action. Yet motivating reasons for actions, even if the action is conceived of as resulting from one's own evaluative reasoning, can have different sources; they may result from processes of instrumental or structural reasoning, and/or weak or strong evaluation. To avoid misinterpretation, I should also mention that Baker's distinction between acceptance and acceptance cum approval of a certain reason is not identical with my distinction between weak and strong evaluation.

conceives of her action as irrational is that she in fact wanted to be moved by two conflicting motivations for action. On the one hand, she wanted to do what she actually did, while at the same time she did not do what she really (would have) wanted to do. Hence, this person is confronted with a substantial *divergence in her will*. Such being the case, she in fact cannot justifiably conceive of herself as a rational agent. But persons, because of their self-conception of being rational agents, strongly desire to justifiably conceive of themselves as rational agents, and are therefore concerned with the *soundness of their will*. This soundness of the will is what I also label the unity of the will. We can summarize the result of this argumentation in the following thesis:

(T6) A person who conceives of certain of her actions as irrational is confronted with a substantial divergence in her will, and can therefore, in the light of those actions, not conceive of her will as sound. But persons are concerned with the soundness of their will, because they desire to justifiably conceive of themselves as rational agents.

This thesis entails the claim that I made in the introduction: Conceiving of our actions as irrational matters to us, even apart from the specific purpose the achievement of which is threatened by acting the way we do (or did).

I should address a possible objection which expresses scepticism regarding the concern with the soundness of our will. It seems to be attractive to (certain) persons to deliberately act irrationally (with respect to a certain adopted long(er)-term purpose); hence, these persons do not bother whether they act irrationally; on the contrary, they are attracted by (occasionally) acting irrationally – or so the objection goes. This objection is based on a misinterpretation of the first-personal conception of acting irrationally. The objection (a) confuses a self-related conception of acting irrationally with a third-personal one; (b) overlooks the possibility of 'camouflage behaviour'; or (c) labels certain kinds of motivations for action as irrational, and transfers this labelling to the action in question. In what follows, I discuss the three misconceptions, which might be the basis for the sceptical objection, one by one.

Regarding the first misconception: If a person's action conflicts with a certain long(er)-term purpose which she announced (and probably still announces) as being a purpose of herself, an observer of that person's action may think that the person acts irrationally. But the observer cannot legitimately conclude that the person in question is not concerned with the soundness of her will. Why is that? The fact that an observer assesses the person's action as irrational does not mean that the person herself assesses her action as irrational. She might be content with her decision to act in conflict with her long(er)-term purpose and (a) therefore adjusted or will adjust that purpose (the corresponding action structure, or its internal or external schedule conditions); (b) may have removed (or will remove) that purpose from her evaluative agenda; or (c) does not need to rework her

evaluative agenda, because the action structure, or the internal or external schedule conditions of her long(er)-term purpose left room for the action that she actually performed.

Regarding the second misconception: There may be persons who 'flirt' with acting irrationally. That is why our objector might think that these persons are not concerned with the soundness of their will. I believe this conclusion is mistaken. A person who 'flirts' with acting irrationally (a) does not conceive of her own actions as irrational (she may, for example, have learned to build some slack into the action structures for certain long(er)-term purposes in order to allow for actions which she otherwise may have conceived of as irrational), but, for whatever reason, wants to be seen as a person who occasionally acts irrationally, or (b) conceives of certain of her actions as irrational and wants to avoid a negative judgement about herself from her interlocutors because of her irrational actions. I admit that this is a psychological interpretation of the person's behaviour, but I think that 'flirting' with acting irrationally is 'camouflage-behaviour' and can therefore only be explained psychologically.

Regarding the third misconception: Generally, we think of persons who are, in certain situations, moved by 'emotional motivating reasons', in a positive way. For example, we assign to them attributes like sensitive or gentle. Even persons who sometimes act unwisely are generally seen as more sympathetic than persons who seem to act wisely all the time; this is because certain weaknesses or feelings are typical of humans; after all, we are not machines. Such being the case, it is natural that we like to be seen as persons who are sometimes moved by 'emotional motivating reasons'. Our objector might have had this in mind when stating that persons are attracted by (occasionally) acting irrationally and are, therefore, not concerned with the soundness of their will. But he made a mistake, which has its source in folk-psychology. What exactly is his mistake? In folk-psychology, certain kinds of motivations for action, which people conceive of as emotional reasons, are often labelled as 'irrational', and people may transfer this label to those actions which are, according to their classification, based on emotional motivating reasons. But conceiving of the motivating reasons for one's action as emotional neither excludes that the person in question conceives of her action as resulting from her own (conscious) evaluative reasoning, nor does it mean that she (retrospectively, during the performance of the action, or right from the beginning) views that action as hindering or preventing the achievement of a certain adopted purpose which she would rather have given more weight to than to the (new) one she decided on (or is going to decide on), during her evaluative reasoning about the action in question. Furthermore, it should not pass unnoticed that people, knowing their own weaknesses and thus situations in which they probably might act unwisely, are able to anticipate situations where they otherwise might conceive of their actions as irrational by taking their own weaknesses into account when planning the action structures, or internal or external schedule conditions

for certain long(er)-term purposes which they decide to pursue. In other words, persons are able to anticipate (in their evaluative agenda) that they are sometimes moved by certain ‘emotional motivating reasons’, and through this try to keep themselves from situations where they otherwise might be confronted with a substantial divergence in their will.

In conclusion, we can say: The objection to thesis T6 that it seems to be attractive to (certain) persons to deliberately act irrationally (with respect to a certain adopted long(er)-term purpose), and that these persons thus do not bother whether they act irrationally, is not tenable, but rests on a misinterpretation of the first-personal conception of acting irrationally, as it is defined in thesis T2.

9 Being irrational matters to us - the concern with coherence of our evaluative agenda

To approach the question ‘Why does it matter to us whether we are irrational?’, I recapitulate thesis T5 regarding conceiving of oneself as irrational: Conceiving of one’s own action as irrational involves conceiving of the corresponding decision to perform that action as irrational; and if a person conceives of one of her decisions as irrational, she might – via reflection on herself qua self – conceive of herself as irrational, because, in our self-conception, we cannot separate our decisions from ourselves. In short, persons who conceive of certain of their actions as irrational and who reflect on themselves qua self in the relevant way might conceive of themselves as irrational. Two questions arise: Firstly, what is the relevant way of self-reflection which, presupposing the person conceives of certain of her actions as irrational, may lead to conceiving of herself as irrational? Secondly, do all persons reflect on themselves qua self in the relevant way, and if not, which persons engage in the relevant way of self-reflection, and why? When we know the relevant way of self-reflection, which persons engage in such self-reflection, and what exactly they are concerned with in that self-reflection, we presumably know why conceiving of ourselves as irrational bothers us.

Regarding the first question: The kind of self-reflection that may lead to conceiving of oneself as irrational is a reflection on one’s own decisions, processes of evaluative reasoning, and evaluative agenda, across a period of time. This statement may be unsurprising, but it does not follow that it is not useful, nor does it follow that nothing illuminating can be said about the kind of self-reflection under consideration. Let me describe our evaluative reasoning in the light of the involved *maintenance of our evaluative agenda* in order to illustrate how one’s evaluative agenda may become incoherent. Our decisions (for action) are part of our evaluative reasoning. Confronted with various motivations for action and hence with various alternatives, we decide to pursue certain short-term, medium-term or even long-term purposes.

Together with the decision to pursue a certain purpose, we also decide which action structures we plan to perform to achieve that purpose, the internal or external schedule conditions, and, probably, a set of behavioural rules to support the actualization of the purpose in question. Deciding to pursue a certain purpose means adopting that purpose, and, therefore, putting the purpose on one's evaluative agenda in order to take that purpose into account when being engaged in further processes of evaluative reasoning. Sometimes the decision to pursue a certain purpose will not fit in with our evaluative agenda. In such situations, the decision to pursue that purpose may involve adjusting, abandoning or removing certain earlier adopted purposes (together with their corresponding action structures, or internal or external schedule conditions), or we may decide to refrain from pursuing that purpose because of the conflict with earlier adopted purposes. If we used our capability of evaluative reasoning in the way described, we would never 'find ourselves' in situations where we conceive of our own actions as irrational. But we are not perfect.

We sometimes make decisions to act (in a certain way) which we (retrospectively, during the performance of the action, or right from the beginning) conceive of as irrational. When we think that one of our actions hindered (or hinders) of even prevented (or may prevent) the achievement of an adopted purpose, then we normally try to repair the situation, or rework our evaluative agenda, or both. In the first case (that is, if a person thinks that her action hinders the achievement of an adopted purpose), the person may still be in a position to abandon her irrational action without hindering effects on the adopted purpose that she wants secured, or it may be possible for her to undo the hindering effects of her irrational action by performing 'repair-actions', so to speak. If she cannot abandon the irrational action, because, for example, it cannot be stopped half-way or is already completed, or if she cannot undo the hindering effects of her irrational action on the adopted purpose that she still wants to achieve and which she has not given enough weight to during her process of evaluative reasoning, then, normally, the person will adjust the adopted purpose in question, adjust the action structure that she planned to perform to achieve that purpose, adjust its schedule conditions, or adjust or remove other adopted purposes the action structures of which could interfere with the purpose the achievement of which she wants to secure. In the second case (that is, if a person thinks that her action makes the actualization of a certain adopted purpose impossible), the situation is similar. The person may still be in a position to abandon her irrational action without damaging effects on the adopted purpose in question, or it may be possible for her to undo the damaging effects of her irrational action by performing 'repair-actions'. If the obstruction of her adopted purpose is a fact, then, normally, the person will remove that purpose from her evaluative agenda.

In conclusion, if a person adopts new purposes which do not fit in with her evaluative agenda without reworking that agenda, or, when conceiving of

a certain action of herself as irrational, does not abandon that action, perform 'repair-actions', or rework her evaluative agenda according to the new situation, then this person's evaluative agenda has become incoherent. And a person with an incoherent evaluative agenda will, quite likely, regularly 'find herself' in situations where she may conceive of her actions as irrational. Each of these situations may trigger her to rework her evaluative agenda to achieve a situation in which all the parts of that agenda (that is to say, the purposes together with the corresponding action structures, behavioural rules, or internal or external schedule conditions) fit together well, in other words, to achieve coherence of her evaluative agenda. And each of these situations may also lead to a reflection about the fact that she regularly 'finds herself' in situations where she conceives of her own actions as irrational and feels the need to (more or less radically) rework her evaluative agenda. By asking herself 'why do I regularly end up in situations where my goals don't fit together?', she will reflect on her decisions, processes of evaluative reasoning, and the state of her evaluative agenda, across a period of time. This is the relevant way of self-reflection that those persons engage in who (as a result of that reflection) might conceive of themselves as irrational.

We can state so far:

(T7) A person who – as a result of reflecting on herself qua self in the relevant way, that is, by reflection on her decisions, processes of evaluative reasoning, and evaluative agenda, across a period of time – conceives of herself as irrational is confronted with an incoherent evaluative agenda.

Let us now turn to the second question: Do all persons reflect on themselves qua self in the relevant way, and if not, which persons engage in the relevant way of self-reflection, and why? Only persons with a narrative self-conception (being a property which is characteristic of adult human persons) will engage in self-reflection regarding the state of their evaluative agenda across a period of time. Why is that? Persons with a narrative self-conception conceive of their life as a unity. They connect their actions, experiences, traits, desires, and goals (and position them in relation) with each other, and, as a result of that, they 'produce' the autobiographical narrative that they conceive of as their life. Such persons conceive of themselves qua self as the protagonist of their continuously developing self-narrative. This self-conception requires a certain degree of coherence in the person's self-narrative. And the coherence of a person's autobiographical narrative is threatened, disturbed or even disrupted by actions which the person herself conceives of as irrational, because these actions do not accord with certain goals which are still part of her self-narrative. The person will either have to 'rewrite' her self-narrative, which in this situation means that she will have to adjust or dismiss certain of her goals, or she will, due to a *fragmentation of her self-narrative*, not be able to conceive of herself as the *well-defined*

protagonist of her autobiographical narrative anymore. The latter case, however, would frustrate her, because, owing to her narrative self-conception, she desires to be able to conceive of herself as the well-defined protagonist of her self-narrative.[20]

To sum up, a person's concern with the coherence of her self-narrative involves a concern with the coherence of her evaluative agenda. After all, an incoherent evaluative agenda will quite likely lead to situations where a person will conceive of her own actions as irrational, which in turn – via reflection on herself in the relevant way – might lead to conceiving of herself as irrational, which is a threat to her self-conception as being the well-defined protagonist of her developing self-narrative, because, strictly speaking, a person with an incoherent evaluative agenda has a fragmented self. Even if a person reworked her evaluative agenda in each situation where she conceived of actions of herself as irrational or in each situation which otherwise would have led to incoherence of her evaluative agenda, she will, at least if this happens frequently, at a certain moment no longer be able to conceive of herself as the well-defined protagonist of her self-narrative, because her evaluative agenda is seriously incoherent across a period of time. We can summarize the answer to our second question in the following thesis:

(T8) A person who conceives of herself as irrational is confronted with an incoherent evaluative agenda (across a period of time). But persons (with a narrative self-conception) are concerned with the coherence of their evaluative agenda, because incoherence of a person's evaluative agenda implies incoherence of her self-narrative; persons, however, desire to be able to conceive of themselves qua self as the well-defined protagonist of their autobiographical narrative and thus are concerned with the coherence of their self-narrative.

And combining theses T7 and T8, we can answer this section's question 'Why does it matter to us whether we are irrational?':

(T9) A person who – as a result of reflecting on herself qua self in the relevant way, that is, by reflection on her decisions, processes of evaluative reasoning, and evaluative agenda, across a period of time – conceives of herself as irrational is confronted with an incoherent evaluative agenda (across a period of time). But persons (with a narrative self-conception) are concerned with the coherence of their

[20] I should mention that, in my view, the property of having a narrative self-conception can be developed to a higher or lower degree. Whether and to what degree one suffers from a fragmented self depends on the strength of one's desire for coherence of one's self-narrative. And this in turn depends on the degree to which the corresponding property is developed. In short, different persons may be satisfied with different degrees of coherence of their self-narrative. Therefore, different persons may be bothered by different degrees of incoherence of their respective evaluative agendas. See also Chapter 2 (The relation of oneself to one's self).

evaluative agenda, because incoherence of a person's evaluative agenda implies incoherence of her self-narrative; persons, however, desire to be able to conceive of themselves qua self as the well-defined protagonist of their self-narrative and thus are concerned with the coherence of that narrative.

10 Concluding summary

I analysed the notion of first-personal practical irrationality, and argued the following main theses, each of them related to one of the major subtopics:

(T2) *Conceiving of one's own action as irrational:* Conceiving of one's own action as irrational means (1) that we conceive of the action as resulting from our (conscious) evaluative reasoning, and (2) that we (retrospectively, during the performance of the action, or right from the beginning) view that action as hindering or preventing the achievement of a certain adopted purpose which we would rather have given more weight to than to the (new) one we decided on (or are going to decide on) during our evaluative reasoning about the action in question.

(T5) *Conceiving of oneself as irrational:* Conceiving of one's own action as irrational involves conceiving of the corresponding decision to perform that action as irrational; and if a person conceives of one of her decisions as irrational, she might – via reflection on herself qua self – conceive of herself as irrational, because, in our self-conception, we cannot separate our decisions from ourselves.

(T6) *Conceiving of one's own action as irrational matters to us:* A person who conceives of certain of her actions as irrational is confronted with a substantial divergence in her will, and can therefore, in the light of those actions, not conceive of her will as sound. But persons are concerned with the soundness of their will, because they desire to justifiably conceive of themselves as rational agents.

(T9) *Conceiving of oneself as irrational matters to us:* A person who – as a result of reflecting on herself qua self in the relevant way, that is, by reflection on her decisions, processes of evaluative reasoning, and evaluative agenda, across a period of time – conceives of herself as irrational is confronted with an incoherent evaluative agenda (across a period of time). But persons (with a narrative self-conception) are concerned with the coherence of their evaluative agenda, because incoherence of a person's evaluative agenda implies incoherence of her self-narrative; persons, however, desire to be able to conceive of themselves qua self as the well-defined protagonist of their self-narrative and thus are concerned with the coherence of that narrative.

To avoid misinterpretation, I should re-emphasize that this chapter did not analyse the notions of conceiving of one's own actions as non-rational, mistaken or wrong, although we saw the shape of the theory that would emerge concerning these notions.

Chapter 4
Wrongness from a self-related perspective

In certain situations, persons may conceive of their actions as wrong. This phenomenon calls for further explanation, because questions like the following arise: When do we conceive of our actions as wrong? Do all persons sometimes have self-related conceptions of acting wrongly? What standards do we apply (whether consciously or not) when conceiving of our actions as wrong? Does it matter to us whether we act wrongly, and if so, why? Let me illustrate the phenomenon of conceiving of one's action as wrong with an example, which, although it does not cover all aspects of the notion it is meant to illustrate, may clarify my view of a self-related conception of wrongness of action, and specifically the distinction I draw between that conception and a self-related conception of irrationality of action.

A couple has a savings account which they dedicated to periodical or unexpected household expenses (for example, the yearly premium for the fire and theft insurance, or the cost of a new refrigerator); every month, they pay a fixed amount of money into that account. One day, when doing some window-shopping, the woman sees a beautiful dress; and, while she knows that she cannot effort it, she tries on the dress. It suits her very well, and she really wants to have that dress. She thinks about the possibility of drawing some money out of the special savings account; she first hesitates, but eventually she draws money out of that account and buys the dress. Having bought the dress and come back home, she feels regret. On the one hand, she conceives of her action as irrational, because she knows that it hinders the actualization of her (and her husband's) goal to be prepared for special household expenses; on the other hand, and incomparably more important to her, she conceives of her action as wrong, because she broke the agreement that she has made with her partner (namely to use the special savings account only for the expenses that it is meant for, and by this to secure their standard of living). She blames herself, because her action does not fit the kind of person she wants to be, namely a person who is reliable.

I claim: Whenever a person conceives of one of her actions as wrong, this way of conceiving of her action overrides all other possible ways of conceiving of that action (regarding aspects of rational agency). And further: Conceiving of our actions as wrong bothers us, because it threatens our desire for loyalty to ourselves.

Before arguing for these claims, I will elucidate the notion of conceiving of one's action as wrong, which in turn will lead to some reflection about the notions of person, action, purpose and value, and to an elaboration of the different functions of values and purposes in our evaluative reasoning. I will conclude the first part of this chapter by extending the self-related conception of wrongness of action to a conception of wrongness of oneself. In the second part, I will be concerned with the 'mattering-questions'. I shall begin with the overriding nature of the self-related conception of wrongness of action with respect to other ways of conceiving of one's action, and then turn to the question 'Why does acting wrongly bother us?'. Finally, I will try to answer the question 'What is it about us that makes wrongness of ourselves matter to us?'. I should emphasize that, although the notion of conceiving of oneself as wrong is the subject of the final sections of both Part I and Part II and might therefore be seen to be the culmination of this chapter, my primary concern is with the first-personal conception of wrongness of *action* (and with the corresponding 'mattering-question'). Once that notion has become clear, the notion of conceiving of *oneself* as wrong develops naturally, so to speak.

Let me first dispose of a possible question concerning the subject of this chapter: Why am I concerned with the first-personal notion of *wrongness* of action instead of focussing on the first-personal notion of *rightness* of action? Our moral emotions relate predominantly to actions that we conceive of as wrong. We blame ourselves for having acted as we did, we feel regret or remorse, or we may be full of contrition when we conceive of certain of our actions as wrong. Those feelings may lead to reflections on our motivations for action and on the 'qualitative standards' we apply to ourselves. Admittedly, we may also have feelings when we conceive of certain of our actions as right. A person may, for example, feel relieved that she managed to uphold her values although she risked her career by acting as she did. But explicitly conceiving of our actions as right happens less frequently than conceiving of our actions as wrong, furthermore the corresponding feelings are less vivid and they do not typically lead to further reflections on our motivations for action and on the 'qualitative standards' we apply to ourselves. It is the deviation from our own self-regarding qualitative standards that raises moral emotions and can threaten our self-conception.

Part I – Conceiving of one's own actions (or of oneself) as wrong

In my view, conceiving of one's action as wrong differs essentially from conceiving of one's action as irrational. Although both ways of conceiving of one's action require reflection on (one's motivations for) one's action, they differ essentially in the reference criterion that is applied during that reflection: when a person conceives of her action as wrong, that conception results from her concern with her values, whereas when a person conceives

of her action as irrational, that conception results form her concern with the achievement of her adopted purposes. Therefore, to make a start with my account of a first-personal conception of wrongness, I will first distinguish the two kinds of reflection on (one's motivations for) one's actions, and elucidate the notion of a person's values by drawing the distinction between a person's values and a person's purposes.

1 Prerequisites of conceiving of one's own action as wrong

A person may conceive of (her motivation for) her action as wrong when she thinks and feels that she does not want to be the kind of person who acts in the way she does (did or considers doing). This preliminary thesis implies that a being that may conceive of its own actions as wrong necessarily fulfils three requirements: (1) it is conscious of its agency; (2) it has the ability to reflect on (its motivations for) its actions; and (3) it has the ability to apply self-regarding qualitative standards when assessing its actions. Let me explain these three requirements.

To be conscious of one's agency requires a *first-person perspective*, that is to say, it requires that the being in question is aware of the fact that it has its own perspective and that this perspective differs from the perspective of others. A being that fulfils this characterization is a self-conscious being. But conceiving of one's actions as actions of oneself requires more than a rudimentary form of self-consciousness; it requires that the being is able to conceptualize its own perspective and, because of this, to think of itself as itself. Such a being is called a person. A person thinks of herself as the subject of her behaviour, thoughts, feelings and experiences. She thinks of herself as an agent, that is to say, as a being that wants to achieve her own purposes (by acting the way she does) and has her own motives, and, more specifically, as an agent that decides on her own actions by reasoning. These ways of thinking of oneself require a highly developed form of self-consciousness, in other words, of the property of having a first-person perspective. From now on, when I talk about a first-person perspective, I mean this highly developed form of a first-person perspective, which is fundamental for being a person. Owing to a first-person perspective, persons (among other abilities which are based on a first-person perspective) (a) can think of themselves in the future and in other circumstances, and this means that they can imagine alternatives for their own futures, and (b) are capable of taking a position regarding their motives for their actions and of reflecting on these motives, and this means that they are capable of rational agency.

In conclusion, because of their property of having a first-person perspective, persons fulfil the first two of the aforementioned prerequisites for a self-related conception of wrongness of action: they are conscious of their agency and have the ability to reflect on (their motivations for) their actions. What about the third prerequisite, that is, the ability to apply self-regarding

qualitative standards when assessing one's actions? In my view the ability to reflect on (one's motivations for) one's actions does *not* entail in itself the ability to apply *self-regarding qualitative* standards.

Based on their property of having a first-person perspective, persons are able to reflect on (their motivations for) their actions (before, during, or after the performance of actions) in a number of different ways. When reflecting on their motivations (to act in a certain way) and their alternative actions *before* performing a certain action[1], persons typically evaluate whether (the purpose of) the action they are considering fits in with other (longer-term) purposes that they want to achieve, whether a certain alternative is more effective or efficient than another to fulfil a current spontaneous desire, or, in case of conflicting motivations for action, which action purpose they want to give more weight to. When reflecting on their actions *during* or *after* their performance, persons evaluate, for example, whether their action is (or was) an effective or efficient means to achieve the purpose they want(ed) to actualize, or whether they, for the benefit of a certain longer-term purpose that they had adopted earlier, should rather have rejected the spontaneous desire which they had decided to fulfil.

An essentially different way of reflecting on one's motivations and alternative actions is to ask oneself questions like 'What kind of person do those motivations characteristically belong to; and do I want to be that kind of person?' or 'What kind of person would I be if I performed that action?'. Reflections of this kind require that the person in question makes *qualitative distinctions* between certain kinds of persons (or kinds of life), assessing them in terms of 'higher' and 'lower' (for example, honourable versus dishonourable, or caring versus selfish), and that she applies these qualitative standards to herself qua person (or to her life). When reflecting on her motivations (to act in a certain way) *before* performing a certain action, such a person may reject a certain motivation for action, because she does not want to be the kind of person that is moved by that (kind of) motivation. She may judge that she would act wrongly if she were moved to act (in a certain way) by the motivation in question. When reflecting on her action *during* its performance, such a person may judge that, by acting as she does, she is the kind of person that she does not want to be. She may judge that she acts wrongly, may feel regret, and might abandon her action or try to undo its consequences by performing 'repair-actions'. And when a person who makes qualitative distinctions between kinds of persons or kinds of life reflects on her actions *retrospectively*, she may judge that she frequently performed actions which belong to a kind of life which she thinks of in negative terms (or in other words, which she regards as qualitatively 'low').

---

[1] The term *action structure* would be more appropriate here than the term action, because what we generally call actions are in fact (more or less complex) action structures. In the context of this chapter, however, I will nevertheless use the term action, as long as it does not make any difference to the meaning.

She may judge that she acted wrongly and might be disappointed with herself.

To sum up: The ability to apply *self-regarding qualitative standards* when assessing one's actions, which is the third prerequisite for a self-related conception of wrongness of action, is not implied by the property of having a first-person perspective; it requires that the person has the property of making qualitative distinctions, which is the property of assessing (one's motivations for) one's actions in terms of 'higher' and 'lower'; after all, she must be able to reflect on (her motivations for) her actions in terms of the kind of person or kind of life that these (motivations for) actions characteristically belong to. [2, 3, 4]

Let me – before unfolding my account of regarding one's actions as wrong any further – consider a possible objection to the (preliminary) thesis, which is included in my above explanations concerning a person's ability to make qualitative distinctions and to apply them as qualitative standards to herself, namely: a person may conceive of her action as wrong when she thinks and feels that she does not want to be the kind of person who acts in the way she does, did or considers doing. This thesis may at first sight seem to require one thought too many. The woman in the introductory example, for instance, does not ask herself – or rather, does not have to ask herself – what kind of person she wants to be in order to regard her action as wrong. She blames herself, because she did not stick to her agreement with her partner. Period. Or so the objection goes. Admittedly, when we conceive of one of our actions as wrong, because we think of that action in a negative way and experience reflexive emotions of regret or remorse, we do not

---

[2] The property of making qualitative distinctions, although not being implied by the property of a first-person perspective, obviously requires that property; after all, all kinds of self-reflection require a first-person perspective.

[3] I should repeat that I owe the inspiration for this aspect of my view on a self-related conception of wrongness of action to Charles Taylor's explications of what 'strong evaluation' means. "Motivations or desires do not only count in virtue of the attraction of the consummations, but also in virtue of the kind of life and kind of subject that these desires properly belong to." (Taylor 1985, p. 25).

[4] By the way, my view that only persons – or rather, only persons with the property of making qualitative distinctions in the sense described above – have the ability to apply self-regarding qualitative standards when assessing their actions does obviously not entail that only beings that are persons may behave in ways that we would describe, for example, by attributing empathetic emotions or feelings to those beings. There are research reports (I refer to a special lecture (NWO/Huygenslecture 2002) given by the psychologist and 'apes expert' Frans B.M. de Waal, organized and published by NWO (the Dutch organization for scientific research), original title (freely translated): 'Man as a social ape, and the reduced popularity of greed') about the behaviour of certain species of anthropoid apes that can give rise to believe that we are born with the building blocks of morality, such as consolation, peacemaking, and mutual help. However that may be, only beings that are able to reflect on themselves (which requires the property of a first-person perspective) by applying self-regarding qualitative standards to themselves (which requires the property of making qualitative distinctions) may conceive of certain of their actions as wrong.

always explicitly think – as a kind of after-thought – "and I do not want to be such kind of person". However, that thought is entailed if I disqualify one of my actions. If I conceive of my action as wrong, I regard my action – or rather, my motivation for that action – as belonging to a kind that I myself condemn for myself, and I also have corresponding negative reflexive emotions. Thus, conceiving of my action as wrong and having bad feelings about it definitely involves the thought, even if not engaged in explicitly, "I do not want to perform actions of that kind" or "I do not want to be moved by the kind of motives that moved me to perform that action". But these thoughts in turn entail that I do not want to see myself as the kind of person that performs such actions or has such motivations. It may be true that if we blame ourselves for having been insensitive to someone's feelings, we do not automatically conceive of ourselves as a person who is insensitive to the feelings of others. Still, the fact that we blame ourselves is only intelligible if we want to be a person who is sensitive to the feelings of others. Given the seriousness of our reflexive emotions, our acts of insensitiveness will soon threaten our self-conception of actually being sensitive or really striving to be sensitive. Potentially, every action that we conceive of as wrong is a threat to our self-conception of actually being or really striving to be a particular kind of person. That is why I think that the woman in the introductory example, who regards her action as wrong because she did not stick to her agreement, does not have to engage in the after-thought "I do not want to be a person who does not stick to her agreements" in order to implicitly think of herself in that way. I am therefore drawn to the view that a qualitative assessment of one's action, when it is part of a self-related conception of wrongness of action, entails not wanting to be the kind of person that acts as one does, did or considers doing.[5]

2 A person's purposes and a person's values

Persons generally have the ability to reflect (whether consciously or not) on their motivations and alternative actions in two ways: (1) weighing them, by assessing them in the light of conflicting spontaneous desires or adopted long(er)-term purposes, and (2) judging what kind of person those (motivations for) action characteristically belong to, by applying self-regarding qualitative standards which represent the person's values. What exactly are a person's values? Let me clarify the notion of a person's values by discussing essential differences between values and purposes.

Persons, like all agents, are capable of purposeful action and of more or

[5] I was encouraged to make this point clear when I presented a paper, which was based on an earlier version of this chapter, at the symposium 'The Constitution View of Persons and Artefacts', held at the Technical University of Delft, The Netherlands, 20 October 2003.

less complex instrumental reasoning. Although beings without a first-person perspective do not conceive of themselves as subjects of their actions, their actions can be explained, as for persons, by assigning desires, beliefs, and intentions to the beings in question. But only persons, because they can think themselves in the future and in other circumstances and are able to reflect on (their motives for) their actions, can think of longer-term purposes, adopt them, and take them into account when reflecting on (their motivations for) their actions. Furthermore, only persons can think of and adopt (long-term) purposes the achievement of which requires that the person thinks of and sets behavioural rules or guidelines for herself that she plans to take into account (when evaluating her motivations for action) until those purposes are achieved. All adopted purposes and the purpose-related behavioural rules stay on a person's evaluative agenda[6] until the purpose in question is achieved, modified or abandoned. To sum up: A person's adopted purposes have 'natural life-cycles', that is to say, they appear on a person's evaluative agenda when they are adopted, they are taken into account when the person evaluates her motivations for action while pursuing those purposes, and they disappear from a person's evaluative agenda when they are achieved. Furthermore, a person can modify or abandon adopted purposes by her own decision, that is to say, as a result of her evaluative reasoning.

A person's values, by contrast, do not have 'life-cycles' and cannot be modified or removed by her decision. Let me explain. A person cannot *decide* to 'have' a certain value and by this add it to her 'set of values', so to speak. Furthermore, a value is never realized and will therefore not disappear from her 'set of values' because of its realization. And a person cannot *decide* to modify certain values or remove them from her 'set of values'; thus she cannot decide *not to apply* certain self-regarding qualitative standards anymore. This is why: A person applies certain self-regarding qualitative standards because she *cares* about being a person of a certain kind or about living a certain kind of life, and she cannot decide *not* to care.[7] Persons who care about the kind of person that they are (by acting as they do and by having the motivations for action that they have), and thus ask themselves 'kind-of-person questions', think of their motivations and actions in terms of, for example, honourable and dishonourable, noble and despicable, brave and cowardly, self-disciplined and impulsive, caring and selfish, merciful and merciless, or loving and loveless. And making these qualitative distinctions in actual situations of

---

[6] As already mentioned in note 13 of the preceding chapter, I owe the idea for the label 'evaluative agenda' to Scanlon, who briefly introduces the term 'deliberative agenda'; see Scanlon 1998, pp. 46-47.

[7] I was inspired by Harry Frankfurt's essay *The importance of what we care abou*t (Frankfurt 1982) to use the notion of care in this way. Frankfurt characterizes 'caring', for example, in the following way: "The fact that someone cares about a certain thing is constituted by a complex set of cognitive, affective, and volitional dispositions and states. [...] It certainly cannot be assumed that what a person cares about is generally under his immediate voluntary control." (p. 85)

action is constitutive of 'having' certain values, of being sensitive to other situations where those values are at stake, and of being inclined to apply the relevant self-regarding qualitative standards in those situations. In other words, making qualitative distinctions makes the values in question normative for oneself, 'it is equivalent to committing oneself, first-personally'[8], to accord with these values.

How do we know our values if we cannot decide to have them, as we (can) do with purposes? We do not know that we have a certain value as long as we do not apply that value as a self-regarding qualitative standard when reflecting on (our motivations for) our actions. Only by assessing, in a concrete situation, (our motivations for) our actions in terms of 'higher' and 'lower' (for example, as honourable versus dishonourable, merciful versus merciless), and thus thinking of the kind of person that those actions or motivations would belong to in those terms, we may become conscious of caring about being the kind of person that we ourselves can conceive of as a person whom a certain value can be justifiably attributed to. Having made a certain qualitative distinction in a concrete situation of reflection on (one's motivation for) one's action, a person may conceive of herself as *having* a certain value, in the sense that she is conscious of caring about being the kind of person whom that value can be justifiably attributed to. We might say: Whereas persons can *decide* to pursue a certain purpose, they *get to know* which values they care about. In other words, persons get to know that they committed themselves, first-personally, to accord with certain values.

When do we experience evidence of having a certain value? We experience evidence of having a certain value (at a certain moment) when we *effectively apply* that value as a self-regarding qualitative standard or when we feel and think that we *should have applied* that value effectively. What does it mean to *effectively apply* a certain value as a self-regarding qualitative standard? It means that a person (a) reflects on a certain action *before* its performance, (b) makes a certain qualitative distinction when assessing her motivations and alternative actions, *and* (c) actually acts in accordance with her self-regarding qualitative judgement. Consequently, a person may feel and think that she *should have applied* a certain value effectively if she (a) failed to reflect on a certain action before its performance, (b) failed to judge her motivations and alternative actions by asking herself 'kind-of-person questions' at all, or failed to judge them in accordance with the value in question, *or* (c) failed to act in accordance with her self-regarding qualitative judgement. In both cases – when a person actually applies a certain value effectively *or* feels and thinks that she should have done so – the evidence of having a certain value is experienced *retrospectively*, that is to say, when she judges that she has (or should have)

---

[8] I borrowed this formulation from a nearly parallel formulation regarding the normativity of instrumental reason by Christine Korsgaard; see Korsgaard 1997, especially p. 245.

effectively applied the value in question. By contrast, the evidence of having a certain *purpose* is experienced when one adopts a purpose, plans actions to achieve an adopted purpose, or thinks of and sets behavioural rules for oneself (for the duration of the pursuit of a certain adopted purpose). That is why failing to effectively take a certain adopted purpose into account (when evaluating one's motivations for action in a concrete situation) in most cases[9] perfectly accords with conceiving of oneself as still having that purpose. After all, a person can adjust her planned actions to achieve the purpose in question, she can plan 'repair-actions' in case her original action hinders the achievement of her adopted purpose, or she can adjust the behavioural rules that she has set for herself for the duration of the pursuit of that purpose. In short, as long as a person plans actions to achieve a certain adopted purpose, she conceives of herself as having that purpose. We might say: Whereas the first-personal evidence of having a certain value is experienced when one retrospectively judges that one has (or should have) effectively applied that value during one's evaluative reasoning concerning a certain action, the first-personal evidence of having a certain purpose is experienced when we plan actions (including probable 'repair-actions' in case we hindered the achievement of that purpose by failing to take that purpose effectively into account when evaluating our motivations for action in concrete situations) in support of the achievement of that purpose.

How long will a person have a certain value? As long as she cares about being the kind of person whom that value can be justifiably attributed to. And that means that a person *has* a certain value as long as she reaffirms that value in concrete situations of reflection on (her motivations for) her action. She will never conceive of a value as definitively realized as she can conceive of purposes as definitively achieved. Although she may, in a concrete situation, conceive of herself as having realized a certain value in that situation, because she effectively applied a certain value as a self-regarding qualitative standard when reflecting on (her motivations for) her action, this will not stop her caring about being the kind of person whom that value can be justifiably attributed to. This is *not* to say that a person *cannot* stop caring about herself in that way, but she will not do so because of the realization of the value in question or because she decides not to care anymore about being the kind of person whom that value can be justifiably attributed to. How then can she stop caring? A person may stop caring (in some cases by first caring less) about being the kind of person whom a certain value can be justifiably attributed to because of certain changes in her personal

[9] Only if the achievement of the purpose in question is definitively obstructed (by failing to effectively take that purpose into account in a certain situation) *or* if the person decides to abandon the pursuit of that purpose, for example, because the required 'repair-actions' would be too costly, in other words, only if the person stops planning actions to achieve a certain purpose, will she no longer experience evidence of having that purpose.

circumstances (in the broadest sense). A person might (whether she is aware of it or not) not assess (her motivations for) her action (anymore) in terms of, for example: (a) brave and cowardly when becoming older; (b) honourable and dishonourable after having become a victim of racial prejudice; (c) merciful and merciless when making a successful career as a 'captain of industry'; or (d) self-disciplined and impulsive after having had therapy to conquer her obsessive-compulsive behaviour. Such a person does not apply (whether consciously or not) a certain value as a self-regarding qualitative standard anymore when reflecting on (her motivations for) her action. A person who stopped 'having' a certain value may *not* even be aware of not having that value anymore. In other words, a person's first-personal commitment to accord with a certain value may (just) disappear.

What are the sources of a person's values? That question obviously may arise, but I will *not* deal with it. Why not? I am concerned with the *function* of values in our evaluative reasoning, not with their sources; and I can ignore the question, because my account of a first-personal conception of wrongness is neutral with respect to the sources of a person's values.

Do all persons have values? No. As argued in Section 1 (Prerequisites of conceiving of one's own action as wrong), only persons who have the property of making qualitative distinctions – which is the property to assess (her motivations for) her actions in terms of 'higher' and 'lower' – have values. Although persons characteristically have that property[10], having the property does not imply that those persons always take their values into account when reflecting on (their motivations for) their action. Furthermore, even if a person, in a concrete situation, originally asks herself 'kind-of-person questions' before performing an action, she may then (consciously) deviate from certain self-regarding qualitative standards (that she originally applied) in order not to hinder or obstruct certain long(er)-term purposes of herself, with the result that her actual action may violate the value(s) she originally took into account.

Can one's values conflict with each other? Yes! In a concrete situation, a person's values may conflict with each other. A person may, for example, care about being the kind of person that keeps her promises and that helps people who need her help. In a concrete situation, she may only be able to accord with one of those values. She may solve the conflict by 'feeling' and thus knowing that she cares more about one value than about the other, in that situation. Or she may be in a dilemma, because she either does not know (beforehand) which of the values she cares about more, or she cares equally about both values.[11] Does this mean that a person should not have those two values (at the same time)? No. In contrast with a person's adopted purposes

---

[10] Persons characteristically have the property of making qualitative distinctions, but the property may be not (fully) developed (yet) or (temporarily) lost; furthermore the property may be developed to different degrees.

[11] See also Frankfurt 1982, especially p. 85.

that should be coherent with each other on pain of conceiving of oneself as irrational[12], a person's values may conflict in concrete situations.[13] In other words, while it is up to the person in question *not* to have conflicting adopted purposes, a person cannot avoid to be first-personally committed to values that may conflict in concrete situations.

Can one's adopted purposes conflict with one's values? Yes, in at least four circumstances. Firstly, a person may adopt a certain purpose without reflecting on that purpose by applying self-regarding qualitative standards, in other words, without 'testing' that purpose against her values. She may later on, when performing actions in support of the achievement of that purpose, recognize that her adopted purpose conflicts with her values and probably abandon that purpose. Secondly, a person – after having reflected on that purpose by applying self-regarding qualitative standards – may adopt a certain long-term purpose. Later on, however, she may judge that certain actions, which she performs to support the actualization of her long-term purpose, do not accord with her values. She might then recognize that her values have changed during the pursuit of that purpose, and that her purpose is in fact in conflict with her (current) values. In that case, she may decide to abandon the long-term purpose in question. Thirdly, a person may adopt a certain long-term purpose which 'passes the test' (that is to say, does not conflict with her values) without knowing in detail which actions will be required to actualize that purpose. Later on, when thinking of certain concrete actions to achieve her purpose, she may judge that those actions do not accord with her values. She may then decide to abandon her purpose. Fourthly, a person may adopt a purpose despite her judgement that the purpose conflicts with her values. Later on, when performing actions to achieve that purpose, she may again reflect on that purpose, and may abandon it because it conflicts with her values. In the aforementioned four circumstances of conflicts between one's purposes and one's values, the person can decide to abandon the purpose in question or not. If a person, when confronted with a conflict between her purposes and her values, does *not* decide to abandon (or adjust) the purpose in question but to further pursue that purpose, she can either not conceive of herself anymore as having the value in question or (as I will explain in the next section) she pursues a purpose on pain of conceiving of (her motivations for) her actions as wrong.

Let me sum up. I have described six essential differences (Dn) between one's purposes and one's values:

---

[12] In the context of this chapter, I will not argue why a person's adopted purposes should be coherent with each other on pain of conceiving of oneself as irrational. See, for that argumentation, Chapter 3 (Practical irrationality from a self-regarding perspective).

[13] This is not to preclude that certain values might conflict per se and not only *per accidens*, but I cannot think of such a pair of values; that is why I am tempted to think that, for a certain person, values can only conflict in concrete situations.

(D1) *Role*: Taking one's purposes into account when reflecting on (one's motivations for) one's action means weighing them, by assessing them in the light of conflicting spontaneous desires or adopted long(er)-term purposes. Taking one's values into account, by contrast, means judging what kind of person those (motivations for) actions characteristically belong to, by applying self-regarding qualitative standards which represent the person's values.

(D2) *'Life-cycle'*: Contrary to a person's adopted purposes, a person's values do not have 'life-cycles'. A person's purposes appear on a person's evaluative agenda when they are adopted, and (as long as the person does not remove or modify them before their achievement) they disappear 'naturally' from a person's evaluative agenda when they are achieved. By contrast, values do not disappear from a person's 'set of values' because of their realization, since values are never definitively realized.

(D3) *Control*: Whereas a person can decide to pursue a certain purpose (and by this put that purpose on her evaluative agenda), she cannot decide to care about being the kind of person whom that value can be justifiably attributed to, in other words, a person cannot add a value to her 'set of values', so to speak. Furthermore, whereas a person can, abandon adopted purposes and so remove them from her evaluative agenda, she cannot remove a value from her 'set of values' by her own decision. A person cannot, by her decision, stop caring about being the kind of person whom that value can be justifiably attributed to.

(D4) *Knowledge*: Whereas a person can decide to pursue a certain purpose, she gets to know which values she cares about. Only by actually applying a certain self-regarding qualitative standard in concrete situations of reflection on her actions does a person become conscious of caring about being the kind of person whom the value in question can be justifiably attributed to.

(D5) *Evidence*: Whereas the first-personal evidence of having a certain purpose is experienced when a person plans actions in support of the achievement of that purpose, the first-personal evidence of having a certain value is experienced when one judges that one has (or should have) effectively applied that value during one's evaluative reasoning concerning a certain action.

(D6) *Conflicts*: While it is up to the person in question *not* to have conflicting adopted purposes, a person cannot avoid to be first-personally committed to values that may conflict in concrete situations.

3 A first-personal conception of acting wrongly

Persons, normally, have adopted (longer-term) purposes *and* they have values. Both are 'meant' to be taken into account when one acts. A person's adopted purposes as well as her values are based on first-personal normative acts, namely the acts of adopting a purpose and of effectively making a certain self-regarding qualitative distinction (during a process of evaluative reasoning)

respectively. The former – adopting a purpose – is equivalent to committing oneself, first-personally, to pursue that purpose; the latter – effectively making a certain self-regarding qualitative distinction – is equivalent to committing oneself, first-personally, to accord with the value in question.[14] But making a commitment entails that one can deviate from that commitment; thus persons can deviate from their first-personal commitments regarding both their adopted purposes and their values. When a person deviates from her first-personal commitment to pursue a certain purpose, she may (under certain further conditions) conceive of her action as irrational. When a person deviates from her first-personal commitment to accord with a certain value, she may (under certain further conditions) conceive of her action as wrong. Because my concern here is with the notion of a first-personal conception of wrongness, let me focus on deviation from one's first-personal commitment to accord with certain values, and on the further conditions for conceiving of one's own actions as wrong.

From a first-personal perspective, we can distinguish different cases of deviation from a person's commitment to accord with her values: a person may be conscious of her deviation or not, and the reasons for her deviation may be different. For each case, I will first characterize the reason for the deviation, I will then describe the circumstances of that situation, and finally, I will discuss whether a person may conceive of her action as wrong under the described conditions. (By the way, I do not claim to cover all possible cases.)

Case 1: A person may *consciously* deviate from her former first-personal commitment to accord with a certain value because her values have changed. The person recognizes that she, although she performed, performs, or is going to perform an action which is not in accordance with a value she thought she had, does not view (her motivation for) her current action as not belonging to the kind of person that she wants to be. She becomes aware that she does not care anymore about being the kind of person whom that value can be justifiably attributed to. In other words, she recognizes that her values have changed, and this means that she is no longer committed, first-personally, to the value in question.
Conclusion: A person who deviates from her former commitment to accord with a certain value, because she recognizes that she actually does not care anymore about being the kind of person that has the value in question, in other words, because she recognizes that her values have changed, does *not* conceive of the action in question as wrong.

---

[14] I was inspired by Korsgaard's view on the normativity of instrumental reason (Korsgaard 1997) to see 'adopting a purpose' and 'effectively making a certain self-regarding qualitative distinction' as first-personal commitments. Korsgaard states: "[W]illing an end just is committing yourself to realizing the end. Willing an end, in other words, is an essentially first-personal and normative act. [...] So willing an end is equivalent to committing yourself, first-personally, to taking the means to that end." (Korsgaard 1997, p. 245)

Case 2: A person may *consciously* deviate from her first-personal commitment to accord with a certain value, because, in the actual situation, that value conflicts with another value that she cares about more.
The person is sensitive to the fact that two of her values are at stake in a certain situation, and she thinks that all available alternatives would violate one of the two values. Reflecting on herself, she finds out which value she cares about more, by recognizing that she 'cannot'[15] act otherwise than in a way that accords with that value. In other words, the person becomes aware which of her values she cares about more, and she consciously deviates from her first-personal commitment regarding the other value.
Conclusion: A person who consciously violates one of her values because, when confronted with a conflict between two of her values (in a concrete situation), she recognizes that she cares more about another value, does *not* conceive of the action in question as wrong.

Case 3: A person may *consciously* deviate from her former commitment to accord with a certain value, because she is confronted with a conflict of two values which she cares about evenly.
The person is sensitive to the fact that two of her values are at stake in a certain situation, and she thinks that all available alternatives would violate one of the two values. Reflecting on herself, she cannot find that she cares more about either of the two values; she seems to care evenly about both of them. Because she (thinks that she) cannot avoid actualizing one of the alternatives, she acts while being conscious of violating a value that she definitely wants to accord with.
Conclusion: A person who, in a situation of two conflicting values that she cares about evenly, violates one of them, may conceive of her action as wrong, although she thinks that she has no other alternative than violating one of the values which she cares about evenly.

Case 4: A person may *consciously* deviate from her former commitment to accord with a certain value, because she wants at all costs to pursue a certain purpose the achievement of which would be obstructed if she were to accord with the value in question.
The person is sensitive to the fact that a certain value is at stake in the concrete situation, but she strongly desires to achieve a certain purpose (whether it is an adopted long(er)-term purpose or based on a spontaneous desire); and, while being aware that she cannot conceive of herself anymore

[15] This is not to say that the person does not have the capacities to act otherwise. Frankfurt illustrates the impossibility in question with Luther's famous declaration "Here I stand; *I can do no other.*" Frankfurt explains: "It is clear, of course, that the impossibility to which Luther referred was a matter neither of logical nor of causal necessity. After all, he knew well enough that he was in one sense quite able to do the very thing he said he could not do; that is, he had the capacity to do it. What he was unable to muster was not the *power* to forbear, but the *will*." (Frankfurt 1982, p. 86; italics in original).

as the kind of person whom that value can be justifiably attributed to if she acts as she considers acting, she ignores her judgement that she will violate one of her values, and decides to pursue that purpose, despite of her judgement that she sacrifices her value.
Conclusion: A person who, for the benefit of a purpose that she wants to achieve at all costs, consciously ignores her judgement that she will violate (one of) her values if she acts as she considers acting, may conceive of her action as wrong.

Case 5: A person may deviate from her (former) first-personal commitment to accord with a certain value *without being conscious* of that deviation.
This could be the case in different situations. (a) The person does not ask herself 'kind-of-person questions' when reflecting on (her motivations for) her action, because (for her) none of her values seems to be at stake in the concrete situation; she weighs the alternatives in the light of her adopted long(er)-term purposes, decides on what to do and acts accordingly. (b) The person does not reflect on the alternatives at all (whether consciously or not), and thus performs a 'wanton action'[16]. (c) The person, although she asks herself 'kind-of-person questions' when reflecting on (her motivations for) her action, does not take the value in question into account, because – without being aware of that change – she does not care anymore about the value in question. (d) The person, although she asks herself 'kind-of-person questions' when reflecting on (her motivations for) her action, does not take the value in question into account, because she is not sensitive to the fact that the value in question is at stake in that situation.
Conclusion: A person who is not sensitive to the fact that, when being moved by a certain motivation (to act in a certain way) or acting as she does, a (former) value of herself is at stake, does *not* conceive of her action as wrong.

Having analysed possible reasons for deviation from one's (former) commitment to accord with a certain value, we have individuated two circumstances – the circumstances of case 3 and case 4 – when a person may conceive of certain of her actions as wrong. In both situations, the person conceives of her action as wrong *right from the beginning*, that is to say, she is conscious of going to act wrongly, during her deliberation concerning that action. We might phrase the following thesis:

---

[16] As already mentioned in note 27 of Chapter 2 (The relation of oneself to one's self), I borrow the term 'wanton' from Harry Frankfurt; see Frankfurt 1971, especially p. 16.

(T1) A person may conceive of her action as wrong *right from the beginning* of that action
- (a) if she, for the benefit of a purpose that she wants to achieve at all costs, consciously ignores her judgement that she violates her value if she acts as she considers acting;
- (b) *or* if she, in a situation of two conflicting values that she cares about evenly (and thinking that she has no other alternative than violating one of them), violates one of those values.

But persons not only reflect on (their motivations for) their actions *before* they act, they also (can) reflect on their actions *during* or *after* their performance, and there are circumstances when persons conceive of their actions as wrong *then*. The situations of case 5 could be possible candidates, because, in those situations, a person deviates from her (former) first-personal commitment to accord with a certain value *without* being conscious of that deviation, *at that moment*.

Case 5a: Reflecting on her action during (or after) its performance, a person may recognize that she was not sensitive to the fact that a certain value is (or was) at stake, and that she violates (or violated) that value by acting as she does (or did). She further recognizes that she evaluated (her motivations for) her action without applying self-regarding qualitative standards at all. This person may judge that she should have been sensitive to the fact that the value in question was at stake and that she should have effectively applied the value in question as a self-regarding qualitative standard during her reflection, because she cares about being a person whom the value in question can be justifiably attributed to. In that case, she conceives of her action as wrong.

Case 5b: Reflecting on her action during (or after) its performance, a person may recognize that she started that action as a wanton (or that she performed a 'wanton action'), and that she violates one of her values by acting as she does (or that she violated one of her values by acting as she did). This person may judge that she should have reflected on (her motivations for) her action before performing the action, *and* that she should have effectively applied the value in question as a self-regarding qualitative standard during that reflection, because she cares about being a person whom the value in question can be justifiably attributed to. In that case, she conceives of her action as wrong.

Case 5c: Reflecting on her action during (or after) its performance, a person may recognize that, although she reflected on her action – by applying self-regarding qualitative standards – before starting to act, her action does not accord with a value which she thought she had. She may now become aware of the fact that she does not care anymore about being the kind

of person whom that value can be justifiably attributed to; thus she may now become aware of the fact that her values have changed. In that case, she will *not* conceive of her action as wrong. (By the way, this seems to me a highly hypothetical situation, but I did not want to omit the possibility of such a case.)

Case 5d: Reflecting on her action during (or after) its performance, a person may recognize that, although she reflected on her action – by applying self-regarding qualitative standards – before starting to act, she was not sensitive to the fact that a certain value is (or was) at stake, and that she violates (or violated) that value by acting as she does (or did). This person may judge that she should have been sensitive to the fact that the value in question was at stake and that she should have effectively applied the value in question as a self-regarding qualitative standard during her reflection, because she cares about being a person whom the value in question can be justifiably attributed to. In that case, she conceives of her action as wrong.

Based on this analysis, we might phrase a second thesis:

(T2) A person may conceive of her action as wrong *during* or *after its performance* if she, reflecting on her action, recognizes that she violates (or violated) a certain value by acting as she does (or did), while she cares about being the kind of person whom the value in question can be justifiably attributed to, *and* if she judges that she should have effectively applied the value in question as a self-regarding qualitative standard.

To summarize: I constructed a number of possible cases of deviation from one's (former) commitment to accord with a certain value, and, taking into account the possibility of reflection on one's actions before, during and after their performance, I arrived at two theses. Although we lose certain nuances and information about the conditions, we might, on the basis of T1 and T2, formulate the following more general thesis:

(T3) Conceiving of one's own action as wrong (before, during, or after the performance of that action) means: a person judges that her action does not accord with one (or more) of her self-related qualitative standards which she thinks and feels she should not violate or have violated (in the situation in question), because she cares about being the kind of person whom the value(s) in question can be justifiably attributed to.

4 Critical analysis of my view on conceiving of one's own action as wrong

Let me submit, with the knowledge of T1 and T2, thesis T3 to a critical analysis by asking some questions.

Question 1: A person is forced (under threat of violence) to perform a certain action, which she does not want to perform because performing that action means violating her values. The person succumbs to the threat. Might this person conceive of her action as wrong? The answer depends on the specific circumstances. (a) The person may conceive of her action as wrong if she, although she cannot (or could not) resist succumbing to the threat of violence, nevertheless thinks and feels, before or during that action (or thinks, feels, and judges after that action), that she should not succumb (or not have succumbed) to the threat, and should accept (or have accepted) the consequences, however bad they may be. (b) The person does *not* conceive of her action as wrong if she decides to succumb to the threat, because she judges, for example, that nothing is worth sacrificing one's life for. This does not preclude that the person, retrospectively, may judge that she should have decided otherwise and may conceive of her action as wrong then. (c) The person does *not* conceive of her action as wrong if she, although she cannot resist succumbing to the threat of violence, approves the fact that she succumbs to the threat, because she, for example, thinks that nothing is worth sacrificing one's life for. This does not preclude that the person, retrospectively, may judge that she should not have succumbed to the threat and may conceive of her action as wrong then.

Question 2: A person feels the need to perform a certain action under certain circumstances, no matter what. For example, a person feels the need to have her house cleaned up whenever she leaves it. This leads – at least, according to the judgement of others – to absurd behaviour. She, for example, does not respond adequately when a neighbour asks for help, after all, she first has to clean up her house before helping her neighbour. Will this person conceive of her action as wrong? It depends. If the person cannot resist the need to clean up her house, although she feels and judges that she should resist that need and that another action would be required to accord with her values, then she will conceive of her action as wrong. In that case, she does not conceive of 'having the house cleaned up' as a value but as a compulsion. By contrast, if she really cares *most* about having her house cleaned up, that is to say, if she applies that self-related qualitative standard when evaluating her alternatives for action and, in case of conflicting values[17], really thinks that she should not violate it, no matter

[17] See also case 2 in Section 3 (A first-personal conception of acting wrongly) and the discussion of conflicts between one's values in Section 2 (A person's purposes and a person's values).

what, then she will not conceive of her action as wrong. In that case, her caring about being a certain kind of person is grounds for her need to have her house cleaned up. (Obviously, others might judge that having cleaned up one's house is not worth caring about most.)

Question 3: A person is physically forced by external influences to do what she does. For example, she is given a certain medication as a result of which she exactly follows the instructions of another person (performing the bodily movements that person tells her to perform). After that event she finds what she did detestable, in other words, she does not want to be the kind of person who does such things. Might this person conceive of her action as wrong? No.[18] If a person was forced to perform a certain bodily movement by external influences, that is to say, if the person simply could not avoid performing that movement because of those external influences, she will not conceive of what she did as wrong, because she does not conceive of what she did as her *action*.

Question 4: At a certain moment, a person consciously deviated from her former commitment to accord with a certain value. She did not conceive of that action as wrong, neither before, during, or right after the action. Is it possible that the person, later on, conceives of that action as wrong, nonetheless? Yes. A person can retrospectively reflect on her actions more than once and during different periods of her life, and with different results.[19] Depending on her current set of values, a person's judgement concerning the wrongness of certain of her past actions may differ at different points in time.

Question 5: By performing a certain action, a person consciously sacrifices one of her values for the benefit of another. Might this person conceive of her action as wrong? Yes or no. She might conceive of her action as wrong retrospectively if she thinks and feels that she in fact cares more deeply about the sacrificed value than about the other one, thus if she judges that she should have decided otherwise. But she will *not* conceive of her action as wrong if she still thinks and feels that, in a situation where she could not accord with both values, she accorded with the value she cares about most.

Question 6: A person, having reflected on (her motivations for) her alternative actions (with or without applying self-regarding qualitative standards), performs a certain action. Her action has an unforeseen detestable effect by which one of her values is violated. She thinks that she would not have performed that action if she had foreseen these consequences, in other words, she thinks that she would have taken the sacrificed value effectively into account. Might this person conceive of her

---

[18] The answer is based on my view on the notion of action: A being that *acts* does something that can be explained by assigning desires, beliefs, and intentions to the being in question.

[19] See also my discussion of conflicts between one's purposes and one's values in Section 2 (A person's purposes and a person's values).

action as wrong? Yes or no. If she now judges that she could have foreseen those consequences by reflecting on that action more carefully, she may conceive of her action as wrong. By contrast, if she now judges that she by no means could foresee those consequences, she will *not* conceive of her action as wrong.

Question 7: A person performs a certain action (with or without reflection on that action beforehand), which, according to her own retrospective judgement, would have had detestable consequences if the process of that action had not been interrupted by another person or event. Might this person conceive of her action as wrong? Yes or no. She might *not* conceive of her action as wrong if she did reflect on the action beforehand by taking her values into account and if she retrospectively judges that she could by no means have foreseen the consequences in question. She may, however, conceive of her action as wrong if she judges that she did not reflect at all on her action, or that she, although reflecting on (her motivations for) her action, was not sensitive to the fact that one of her values was at stake. It is important to note that her conception that she acted wrongly is independent of the fact that the consequences of her action did not occur.

In conclusion, the answers to the six questions, all of them being allowed by thesis T3, explain that:

(1) being forced (under threat of violence) to perform a certain action does not imply that the person necessarily will not conceive of that action as wrong; she may (retrospectively) judge that she should not succumb (or have succumbed) to the threat.

(2) feeling internally forced to perform a certain action under certain circumstances, no matter what, and fulfilling that need does not imply that the person will not conceive of that action as wrong; she may judge that she should be able to resist that internal force (presupposing that she does not experience that force as an expression of her caring about being the kind of person whom a certain value can be justifiably attributed to).

(3) conceiving of an action as wrong implies that the person conceives of what she does (or did) as her *action*.

(4) a person's judgement concerning the wrongness of her past actions may differ at different points in time, depending on her current set of values.

(5) being conscious of having sacrificed (or sacrificing) one of two conflicting values does not imply that a person conceives of her action as wrong, as long as she judges that she, in that situation, could not accord with both values and that her action accorded (or accords) with the value which she cares about more.

(6) the fact that a person's action has consequences by which, according to her own retrospective judgement, her values are violated does not imply that she conceives of that action as wrong, as long as she judges that she by no means could have foreseen those consequences.

(7) the fact that a person's action did not have the (according to the person's own judgement) detestable consequences that it would have had if the process of that action had not been interrupted by another person or event does not imply that the person may not conceive of that action as wrong.

5 A first-personal conception of one's own wrongness

In my view, persons, under certain conditions, may conceive of *themselves* as wrong. Let me explain those conditions and argue my view. Persons, generally, not only have the ability to reflect on their actions but to reflect on themselves qua self. A person may ask herself, for example, whether she has planned appropriate actions to achieve her adopted purposes, or why she frequently acts in ways which hinder or obstruct her long(er)-term purposes. Besides reflecting on the way she pursues her adopted purposes, a person who has the property of applying self-regarding qualitative standards may also ask herself whether she can justifiably conceive of herself as a person who has the values which she conceives of as being her values. Both kinds of reflections are *meta-reflections*, because the person does not focus on one concrete situation of action, but she focuses on herself qua self (taking a number of concrete actions into account). I claim: conceiving of oneself as wrong can only result from a meta-reflection on oneself – or rather on oneself qua being a certain kind of person.

In my view, we can distinguish two situations in which a person might conceive of herself as wrong. Firstly, a person might conceive of herself as wrong if she, reflecting on herself, recognizes that she frequently violated (one or more) values which she conceives of as being her values. The person in question may be full of contrition, judging she *should* effectively apply her values during her evaluative reasoning, and wanting that judgement to affect her reflections in future concrete situations of action. She wants to change herself, so to speak. If this is the case, then the person (presumably) conceives of herself as wrong. But it is also possible that the person, reflecting on herself, becomes aware that she was misconceived about still having certain values, that is to say, she recognizes[20] that it does not bother her anymore to have violated those values. If that is the case, the person does *not* conceive of herself as wrong.

Secondly, a person may conceive of herself as wrong if she, reflecting on herself, thinks and feels that the kind of person which she (until now or during a certain period of her life) thought of in positive terms (or in other words, which she regarded as qualitatively 'high') is not the kind of person she cares about anymore; on the contrary, she now thinks of such a kind of

[20] See also Section 2 (A person's purposes and a person's values), D5 (Evidence) and the corresponding explanation.

person in negative terms. She becomes aware that her 'set of values' has changed (completely). Such a person may conceive of herself as wrong retrospectively, especially (but not exclusively) if she cannot imagine anymore that she could care about being the kind of person she cared about being.

There seems to be a certain tension between, on the one hand, the fact that it is *not* a matter of a person's decision to care about being the kind of person which she cares about being (and by this to care about the specific values that she cares about)[21] and, on the other hand, the fact that a person may conceive of herself as wrong retrospectively, because she, during a certain former period of her life, cared about being the kind of person (or living a kind of life) which she now assesses in negative terms (or in other words, which she now regards as qualitatively 'low'). Without denying that tension, we can understand why persons may conceive of themselves as wrong retrospectively, although it is not a matter of their decision to care about what they care about. Let me explain. Caring about certain values, that is to say, effectively applying certain self-regarding qualitative standards is a first-personal normative act; it is equivalent to committing oneself, first-personally, to accord with the values in question.[22] Therefore a person may now think and feel that she *should have always* applied the self-regarding qualitative standards in question; after all, she is now committed to accord with these values. Consequently she might now judge that she had the wrong values, in other words, that she assessed the wrong 'things' as qualitatively 'high'. That is why she may conceive of herself as wrong retrospectively.

We can summarize by stating:

A person, by way of a meta-reflection on herself,

(a) may conceive of *herself* as wrong *now* if she recognizes that she frequently violated (one or more) values which she conceives of as being her values, is full of contrition, judges that she *should* effectively apply those values during her evaluative reasoning, and wants that judgement to affect her reflections in future situations of action – in short, if she suffers from having (frequently) violated her values and therefore wants to change herself;

(b) *or* may conceive of *herself* as wrong *retrospectively* if she thinks and feels that the kind of person which she (until now or during a certain former period of her life) thought of in positive terms (or in other words, which she regarded as qualitatively 'high') is not the kind of person she cares about anymore, on the contrary, she now thinks of such a kind of

---

[21] For example, a person's psychological and physical condition, and the content of certain experiences may affect the self-regarding qualitative standards which she applies to herself.

[22] See Section 2 (A person's purposes and a person's values), D2 (Control) and D4 (Knowledge) and the corresponding explanations as well as the first paragraph of Section 3 (A first-personal conception of acting wrongly).

person in negative terms – in short, if the person conceives of herself as having had 'wrong values'.

By shortening this description, we arrive at the following thesis:

(T4) A person, by way of a meta-reflection on herself, may conceive of *herself* as wrong *now* if she suffers from having (frequently) violated her values and therefore wants to change herself, *or* she may conceive of *herself* as wrong *retrospectively* if the person conceives of herself as having had 'wrong values'.

Let me conclude this part of the chapter with two remarks concerning the possible effect of meta-reflection, or rather the kind of meta-reflection under discussion. Firstly, I believe that (regular) reflection on oneself, by asking oneself whether one justifiably can conceive of oneself as a person who has the values which one conceives of as being one's values, will normally have effect on one's reflections in future concrete situations of action, because it will sharpen one's sensitivity to situations where one's values are at stake, and strengthen one's inclination to ask oneself 'kind-of-person questions' when reflecting on one's actions in concrete situations.

Secondly, a radical reflection about oneself qua kind of person might affect the values one cares about, even though it is not a matter of decision. For example: A man, who plays a major role in public life and who, in his self-conception, derives his dignity from being famous, is confronted with a serious disease of one of his friends. Owing to the pressure of work, he only visits his ill friend once before the friend dies. The death of his friend acts as a trigger for him to reflect on his own life and on the things that are important to him. Doubts arise as to whether being famous is really worth caring about; and he starts thinking that being the kind of person who justifiably can conceive of himself as loving and caring is in fact much more important. His thoughts and judgement about what is worth caring about most may affect the things he really cares about. I believe that doubting whether the values one cares about are worth caring about and thinking that being the kind of person whom certain other values can be justifiably attributed to would be much more important may affect the values one cares about, that is to say, the person may – in future concrete situations of evaluative reasoning – find herself applying other self-related qualitative standards than before. The values she thought of as more important may have become the values she really cares about more.

Part II – Why wrongness of (actions of) ourselves matters to us

As stated in Part I, effectively making certain self-regarding qualitative distinctions, during processes of evaluative reasoning, is equivalent to committing oneself, first-personally, to accord with the values in question.

But why exactly do persons find it necessary to act in accordance with their values? Why, for example, do we feel committed to certain values in the face of adopted purposes that we (might) fail to achieve because of according with our values? Why does it bother us whether we act wrongly? And what is it about us that we might even conceive of ourselves (retrospectively) as *being* wrong? Before trying to answer these 'mattering-questions', I will argue my 'overridingness claim' – that is, the claim that conceiving of one's action as wrong overrides all other possible ways of conceiving of that action (regarding aspects of rational agency) – since the argumentation for this claim will give us insights that will be helpful in dealing with the 'mattering-questions'.

6 The overriding nature of the self-related conception of wrongness of action

When we conceive of one of our actions as wrong, we assign the negative attribute 'wrong' to that action based on our judgement concerning a certain aspect of rational agency, namely whether we effectively reflected on that action by applying certain self-regarding qualitative standards. But we may also evaluate (whether consciously or not) our actions concerning other aspects of rational agency, and conceive of them (depending on our judgement) as rational, irrational, non-rational, or mistaken. My claim is that conceiving of one's action as wrong overrides all other possible ways of conceiving of that action. I will, in the context of this chapter, not elaborate on the different ways of conceiving of one's own actions (concerning aspects of rational agency)[23] but limit myself to the relevant definitions.

Non-rational: When we conceive of our action as non-rational, we do *not* conceive of that action as resulting from our evaluative reasoning.

(Ir)rational: When we conceive of our action as either rational or irrational, we, in both cases, conceive of that action as resulting from our evaluative reasoning (whether it happened consciously or not), that is to say, we reflected on that action (whether consciously or not) and acted according to our decision.

Irrational: In the case of conceiving of our action as irrational, we regard that action as hindering or preventing the achievement of a certain adopted (longer-term) purpose which we, during our evaluative reasoning concerning the action in question, would rather have given more weight to than the one we decided on (or are going to decide on).

[23] See Chapter 3 (Practical irrationality from a self-related perspective), Section 5 (Different ways of conceiving of one's own actions, regarding aspects of rational agency).

Mistaken: When we conceive of our action as mistaken, we recognize that we made mistakes in our instrumental reasoning (and more or less know the mistakes), in other words, we recognize that the action structure or behavioural rules we decided on to achieve a certain (longer-term) purpose are not appropriate (and we more or less know the mistakes in our reasoning).

Let me complete this list by adding thesis T3 (its formulation is slightly modified for the sake of consistency of this list).

Wrong: When we conceive of our action as wrong, we judge that our action does not accord with one (or more) of our self-related qualitative standards which we think and feel we should not violate or have violated (in the situation in question), because we care about being the kind of person whom the value(s) in question can be justifiably attributed to.

When I state that a first-personal conception of wrongness of action overrides all other possible ways of conceiving of one's action (concerning aspects of rational agency), I do *not* claim that the conception of wrongness *excludes* the others, but that the others are pushed into the background, so to speak, as soon as we conceive of a certain action as wrong.[24] Let me illustrate my point by presenting situations where a person conceives of her action as both wrong *and* either non-rational, rational, irrational, or mistaken.

Case 1 (Non-rational and wrong): A woman has put herself on a diet because she wants to lose weight. Having woken up at night and having got out of bed half-asleep, she 'finds herself' eating chocolate cake, which was meant for next day's birthday party of her little daughter. On the one hand, she conceives of her action as non-rational, because she (thinks that she) was not engaged in a process of evaluative reasoning concerning that action; on the other hand, and much more important for her, she conceives of her action as wrong, because what she did will sadden her little daughter, and she does not want to be the kind of person who saddens someone she cares so much about. She blames herself for acting as she did.

Case 2 (Rational and wrong): A couple has a savings account which they dedicated to periodical or unexpected household expenses (this is the example from the introduction of this chapter). One day, although not

[24] By the way, certain other ways of conceiving of one's action (concerning aspects of rational agency) do not exclude each other either; we may, for example, conceive of one of our actions as mistaken *and* irrational. But the possible combinations of conceiving of one's action – if they do not include wrongness of action – are not relevant to the claim under discussion.

without hesitating, the woman takes money from that savings account to buy a dress. Having bought the dress, she feels regret that she decided to act as she did. She conceives of her action as wrong, because she broke the agreement that she made with her partner (namely to use the special savings account only for the expenses that it is meant for, and by this securing their standard of living). She blames herself, because her action does not fit the kind of person she wants to be, namely a person who is reliable.

Case 3 (Irrational and wrong): A woman, who has put herself on a diet because she wants to lose weight, feels like eating chocolate. She tries to take her mind off it, but she cannot think of anything else. Eventually, conscious of violating her diet, she eats chocolate and other sweet things, including the ones that were meant for this afternoon's birthday party of her little daughter. At other times, when she deliberately ate sweet things while being on a diet, she conceived of those actions as irrational, because they hindered the achievement of her goal to lose weight. In the current situation, however, the fact that her action will sadden her little daughter (because there are no sweets left for the birthday party) is incomparably more important to her than the fact that, by eating the chocolate, she hinders the achievement of her goal to lose weight. After all, she does not want to be the kind of person who saddens someone she cares so much about.

Case 4 (Mistaken and wrong): A couple has a savings account which they dedicated to certain household expenses. One day, the woman wants to buy a dress that she actually cannot afford. She thinks about taking money from the savings account but hesitates. Suddenly, there seems to be a solution; after all, she will receive her holiday pay at the end of the month. The woman decides to take money from the savings account to buy the dress, and to pay the same amount back into the account when she receives her holiday pay. Having bought the dress, she recognizes that she made a mistake in her reasoning: her holiday pay is not an unexpected additional income, but it is part of the couple's yearly budget which their financial plans and agreements are based on. Nevertheless, what really bothers her, when reflecting on her action, is not that her action was mistaken, but that her action was wrong anyway. After all, she broke an agreement that she made with her partner (namely to use the special savings account only for the expenses that it is meant for). She feels regret and blames herself, because her action does not fit the kind of person she wants to be, namely a person who is reliable.

Why does the first-personal conception of wrongness of action push other possible ways of conceiving of one's action (concerning aspects of rational agency) into the background? I can think of the following arguments (An):

(A1) *Threat*: Acting wrongly is a threat to one's self-conception of being a certain kind of person (who has certain values). By contrast, occasionally acting non-rationally, irrationally, or mistakenly is *not* a threat to our self-

conception of being a rational agent (that is, an agent that has the ability of evaluative reasoning and who acts in accordance with the results of that reasoning). After all, we generally allow ourselves to occasionally act non-rationally, mistakenly or irrationally, because we know that we are not machines.

(A2) *Control 1*: When we recognize that one of our actions irrevocably hinders (or hindered) the achievement of one of our adopted purposes, we can adjust or abandon that purpose. By contrast, when we judge and feel that we irrevocably sacrificed one of our values by acting the way we did, we cannot deal with our wrong action by deciding to adjust our 'set of values'. After all, if we were able to remove the value in question from our 'set of values' by our own decision, then we would not really care about being the kind of person whom that value can be justifiably attributed to, and thus would not conceive of our action as wrong in the first place.

(A3) *Control 2*: We may set the pursuit of a certain adopted purpose aside for a certain period of time or until certain conditions are in place without ceasing to have that purpose, whereas, when we care about being a certain kind of person, we cannot postpone that caring.

(A4) *Hierarchy*: Our values are more important to ourselves than our purposes, for we reflect on our purposes by applying self-regarding qualitative standards, and reject, adjust, or abandon (adopted) purposes if they do not 'pass the test', and not the other way round, that is to say, we do not 'test' our values with respect to our purposes.

(A5) *Having versus being*: Our purposes are linked to what we achieve, whereas our values are linked to the kind of person we are. And for persons who 'deeply' care about being a certain kind of person, being (that person) is more important than having (achieved purposes), or so it seems to me.

Obviously, one might raise objections against each of the five arguments. Let me therefore present and reject some possible objections (O-An).

(O-A1) *Threat*: There may be persons for whom every irrational action means a threat to their self-conception. Think of a person who cares about being the kind of person who always (a) takes her adopted (longer-term) purposes into account when being engaged in a process of evaluative reasoning; (b) decides to pursue an action that accords with her adopted purposes; and (c) acts accordingly. For this person, conceiving of an action as irrational will push all other possible ways of conceiving of her action into the background, or so the objection might go. Our objector makes a mistake, however. The person in question, besides conceiving of her action as irrational, conceives of her action as wrong; after all, her action does not accord with her value of having control of herself. The fact that she – according to her own judgement concerning the (ir)rationality of her action – acted irrationally, means – according to her own judgement concerning the rightness or wrongness of her action – a violation of her value of having control of herself. That is why her action is a threat to her self-conception of being the kind of person who has control of herself. And

this threat to her self-conception of being a certain kind of person whom a certain value can be justifiably attributed to is what really bothers her.

(O-A2/A3) *Control*: A person may conclude from one of her actions that she is prepared to give up a certain value for the benefit of the achievement of certain (longer-term) purposes, and she may think that she will decide accordingly in comparable situations. Couldn't we say that this person deals with her wrong action by adjusting her values? No. This person does not really conceive of her action as wrong; after all, she does not think, feel and judge that she should have accorded with the value in question. And a person who does not think, feel and judge that she should accord with a certain value does not really care (anymore) about being the kind of person whom the value in question can be justifiably attributed to. Hence, the person in question, firstly, does not deal with a *wrong* action of herself, since she does not conceive of her action as wrong, and, secondly, does not adjust her values, because she does not have these values (anymore).One may wonder how a person can think that she has a certain value (that is what the objection suggests) whereas she actually does not have that value – as I argue. Here is why: As long as a person has not actually been in situations where a certain value is at stake, or if a person has not been in such situations for a period of time, she might think that she has that value without experiencing evidence of having it. Let me give two examples to illustrate my point. (a) A person may think that she has a certain value, because, during her upbringing, that value was presented to her as being important to accord with. However, when confronted with a situation of that value being at stake, she consciously violates the values and recognizes that she does not really care about being a person whom the value in question can be justifiably attributed to. (b) A person may think that she has a certain value, because, in an earlier period of her life, she abandoned the pursuit of a certain purpose to accord with that value. However, confronted with a new situation of that value being at stake, she is not willing to abandon the pursuit of a certain purpose, which she would have to give up to accord with the value in question. She recognizes that she does not care anymore about being the kind of person whom that value can be justifiably attributed to, in other words, she recognizes that her values have changed. In conclusion: when not being in the concrete situation of a certain value being at stake, one might *think that one has* that value, but one will only experience *evidence of having* that value when one effectively applies that value during one's evaluative reasoning concerning a concrete action (or judges and feels that one should have done so).[25]

(O-A4) *Hierarchy*: Persons can derive purposes from their values. For example, a person who cares about being the kind of person that helps

[25] See also Section 2 (A person's purposes and a person's values), D5 (Evidence) and the corresponding explanation.

others might derive her long-term purpose to become a physician from that value. Such purposes are as important to us as our values and have the same function as values, or so our objector might argue. I agree that a long-term purpose which a person once derived (or thinks that she has derived) from one of her values is important to that person and that she normally will give that purpose a lot of weight when she evaluates alternative actions in concrete situations, but such a purpose does *not* take over the function of a value. After all, if the person who wants to become a physician, instead of (also) asking herself the relevant 'kind-of-person questions' (during concrete processes of evaluative reasoning) *only* asks herself which alternative best agrees with that long-term purpose, then this person is not sensitive anymore to situations where her value of being helpful is at stake. Indeed, by pursuing a 'value-derived' purpose at all costs, a person might not even be sensitive anymore to situations where her original value, which she derived that purpose from, is at stake.

(O-A5) *Having versus being*: Achieving purposes is of great importance to (most) people; it happens, more often than not, that people violate their values for the benefit of their purposes, and they do not seem to suffer from violating their values. This objection is not only an objection against my 'overridingness claim', but also against the claim that acting wrongly bothers us. I will therefore discuss this objection in the next section, which is devoted to that topic.

7 Acting wrongly matters to us – the desire to be loyal to oneself

We – presupposing that we are persons who have the property of making qualitative distinctions, in other words, persons who apply qualititative standards to themselves[26] – conceive of ourselves as persons who have values. But this way of conceiving of ourselves would be threatened if we did not actually act in accordance with our values, that is to say, if we, in concrete situations, when certain of our values are at stake, did not evaluate our (motivations and) alternative actions by applying self-regarding qualitative standards; we then could not justifiably think of ourselves as persons who *have* values. This fact can explain why the property of making qualitative distinctions not only is the basis for conceiving of ourselves as having values but *also* for our concern with our values. In short, conceiving of oneself as having values involves concern with one's values.

Being concerned with one's values means having a strong desire to justifiably conceive of oneself as a person who has certain values, in other words, to conceive of oneself as the kind of person whom these values can be justifiably attributed to. And this in turn means, for every concrete action

---

[26] See also Section 1 (Prerequisites of conceiving of one's own action as wrong).

which we perform, that conceiving of the action as wrong (whether right from the beginning, during the action, or retrospectively) is a threat to our conception of being the kind of person who 'really' cares about the value in question. The fact that we experience a first-personal conception of wrongness of action as a threat to our self-conception of being a certain kind of person confirms my earlier statement that effectively[27] making a certain self-regarding qualitative distinction is equivalent to committing oneself, first-personally, to accord with the value in question.[28]

We might say: Effectively making certain qualitative distinctions between kinds of persons (or kinds of life) is equivalent to committing oneself, first-personally, to accord with the value in question; it is the basis for both one's self-conception of having a certain value *and* for one's concern with that value. This concern is a strong desire to justifiably conceive of oneself as the kind of person whom a certain value can be justifiably attributed to; it is the phenomenological evidence of our first-personal commitment to accord with the value in question. The frustration of that desire – and a first-personal conception of acting wrongly definitely *is* such a frustration – is therefore a threat to one's self-conception of being the kind of person who 'really' cares about the value in question.

I call our desire to justifiably conceive of ourselves as the kind of person whom certain values can be justifiably attributed to *our desire to be loyal to ourselves*. Why do I use this label? A person who is loyal to someone remains firm in her support for that person; in other words, by being loyal to someone a person exhibits a certain constancy. In case of our desire to justifiably conceive of ourselves as the kind of person whom certain values can be justifiably attributed to, we desire to be constant in our support of ourselves to be the kind of being that we care about being. And this desire also entails that we conceive of ourselves as constant in our caring about being a certain kind of person, since in case of supporting oneself, subject and object are one. Thus, it is the constancy of oneself that is entailed by the term 'loyalty to oneself' which makes that term appropriate. Such constancy of oneself, if I am right, is determined by one's values. Let me explain.

If we effectively apply certain self-regarding qualitative standards during processes of evaluative reasoning, and, by doing so, become conscious of caring about being the kind of person whom the corresponding values can be justifiably attributed to, then we cannot, at the same time, think that we will only care about being that kind of person for a certain period of time, or as long as certain conditions are in place. If we 'really' care about something, we cannot plan to stop caring at a certain moment or under certain conditions.

---

[27] Effectively making qualitative distinctions means that a person (a) reflects on a certain action b*efore* its performance, (b) makes a certain qualitative distinction when assessing her motivations and alternative actions, and (c) actually acts in accordance with her self-regarding qualitative judgement.

[28] See also Section 3 (A first-personal conception of acting wrongly).

This is not to claim that we *may* not stop caring at a certain moment or under certain conditions (we presumably will), but the notion of caring precludes planning (or deciding) (not) to care. The notion of caring about something entails the continuity of that caring. That is why conceiving of oneself as having a certain value entails conceiving of oneself as a constant 'thing'. We might say: The constancy of oneself, which one's desire for loyalty to oneself entails, is determined by one's values.

Let me summarize my argumentation by stating the following thesis:

(T5) A person who conceives of certain of her actions as wrong is confronted with disloyalty of herself to herself. But persons have the desire to be loyal to themselves, because, by virtue of caring about being the kind of person who has certain values, they are concerned with their values and desire to justifiably conceive of themselves as persons who have these values.

One might object that the constancy of ourselves, as it may appear to us at the moment of caring about being a certain kind of person, is an illusion, and that we know that it is an illusion, and hence that the loyalty to ourselves cannot really matter to us. This objector confuses our ability to conclude from our experiences with others and with ourselves (because our values actually may have already changed during our life, and we may be aware of that change) that our values can change with the first-personal phenomenology of caring. First-personally, caring entails, as I argued, the continuation of that caring. Therefore, even if we know that our values might change, we cannot help experiencing a first-personal conception of violating our values as not being loyal to ourselves.

Another objector might argue that, to most people, according with their values is of less importance than achieving certain purposes. After all, it happens, more often than not, that people violate their values for the benefit of their purposes. Hence, most people do not bother about acting wrongly. This is objection O-A5 (*Having versus being*) from the last section, and my response will be threefold.

Firstly, our objector may confuse a first-personal perspective with a third-personal perspective. The fact that we (from a third-personal perspective) think that a person does not bother about violating one of her values, in a concrete situation, does not imply that we know that the person herself conceives of her action as wrong: (a) the person might (still) not be aware that one of her values was at stake in the situation in question, and hence not conceive of her action as wrong; (b) she might have been confronted with conflicting values, may have acted in accordance with the value she cares about more, and therefore does not conceive of her action as wrong; (c) her values may have changed, and that is why she did not take her former value(s) into account when reflecting on her alternative actions; she consequently does not conceive of her action as wrong; or (d) we might be

mistaken in believing that the person had the value(s) we thought she had.

Secondly, our objector might confuse effectively applying certain self-regarding qualitative standards with making people (and probably even oneself) believe that one's actions are guided by certain values. A person who, more often than not, violates what *we* thought are her values, and who generally does not conceive of those actions as wrong, does not have those values. After all, having certain values means effectively applying certain self-regarding qualitative standards, or thinking and feeling that one should have effectively applied those standards[29], and thus conceiving of one's own action as wrong, in the latter case.

Thirdly, our objector may be right in her assessment that for (most) people it is of the utmost importance to achieve certain purposes, and that people, in most situations, aim for the achievement of their purposes at the expense of their values. But then, (a) she is either mistaken in her assessment that those people violate *their* values – in that case we are back at my second response; or (b) she is mistaken in her assessment that these people do not conceive of certain of their actions as wrong; after all, it might be the case that (most) people *often* think and feel that they act wrongly.

In short, if a person 'really' has certain values, in other words, if she really cares about being the kind of person whom the values in question can be justifiably attributed to, then it bothers her when she – according to her own self-conception – acts wrongly.

Let me present a final objection which says that it can be doubted whether one's desire for loyalty to oneself leads to (socially) desirable behaviour. One's desire for loyalty to oneself and thus one's acting in accordance with one's values – no matter what other values are at stake – can lead to absurd behaviour, so the objection goes. The objector's example is a woman to whom nothing is more important than having her house cleaned up before leaving it. As a result, the woman exhibits socially unacceptable behaviour: for example, she does not take her children to school in time, and she does not respond adequately when her mother, who lives next door, calls for help. The objector claims (1) that this woman deeply cares about her house being cleaned up whenever she leaves it, since no matter what reason she has to leave her house for, she first has to clean it up, and (2) that the woman's socially unacceptable behaviour is a result of her desire to accord with her deepest value, in other words, of her desire to be loyal to herself. I will reject both parts of the claim. Firstly, I doubt that the woman cares – in the right sense of caring – about having her house cleaned up. Caring in the right sense implies *endorsement* of that caring; and I doubt that the woman endorses her caring about having her house cleaned up no matter what. In other words, caring about a certain

---

[29] See Section 2 (A person's purposes and a person's values), D5 (Evidence) and the corresponding explanation.

value implies caring about being the kind of person whom that value can be justifiably attributed to; and I doubt that the woman really wants to be the kind of person who cares about having her house cleaned up no matter what. I rather think that she feels an internal force to clean up her house, and that she cannot resist succumbing to that internal force. In that case she does not conceive of 'having her house cleaned up no matter what' as a value but as a compulsion. Secondly, if I am right that the woman does not endorse her caring about having her house cleaned up no matter what, but conceives of that 'caring' as an internal force that she cannot resist succumbing to, then the woman's socially unacceptable behaviour is not a result of her desire to accord with her deepest value, but a result of an obsessive-compulsive disorder. By the way, it may even be the case that the woman conceives of her action of first cleaning up the house before responding to her mother's call for help as wrong; after all, she might think that she should be able to resist the internal force (to first clean up the house).[30] I must admit: my argumentation is primarily a rejection of the presented example and the related claim, but may not rule out the objector's initial doubt whether one's desire for loyalty to oneself leads to (socially) desirable behaviour.

The point is that I will not argue against the objector's doubt whether one's desire for loyalty to oneself leads to (socially) desirable behaviour, because I did and do not claim that one's desire for loyalty to oneself necessarily leads to socially desirable behaviour. A person may, for example, care more about being famous than about being a loving husband, or more about being self-disciplined than about being merciful, while our objector might think that caring about being a loving husband or being merciful will lead to socially more desirable behaviour. And taken into account that persons by virtue of caring about being the kind of person who has certain values are concerned with their values and desire to justifiably conceive of themselves as persons who have these values, and thus will exhibit a certain constancy in their caring about these values (as I argued when I introduced thesis T5), a person's 'value-based' behaviour (whether it is socially desirable or not, according to our objector's judgement) will be considerably resistant to changes, because – by virtue of the nature of caring – a person's 'set of values' is relatively resistant to changes.

In conclusion, one's desire to be loyal to oneself entails that one cares about being the kind of person whom certain values can be justifiably attributed to, which in turn entails constancy of that caring and resistance to changes in one's 'set of values'.

---

[30] See also Section 4 (Critical analysis of my view on conceiving of one's own action as wrong), Question 2.

8 Wrongness of oneself matters to us - the desire to be worthy of self-esteem

Conceiving of one's own actions as wrong is not only a matter of one's thinking but also a matter of one's feeling. One's thought that one acted wrongly is accompanied by feelings of regret or remorse. Sometimes it may seem to us that the thought (that we acted wrongly) was primary, and that our feelings of regret result from our judgement that we acted wrongly, whereas in other situations, it may seem to us that the judgement that we acted wrongly is secondary, and that we first felt that we acted wrongly. However it may be, conceiving of one's own actions as wrong is accompanied by feelings of regret or remorse (that one acted as one did). Conceiving of *oneself* as wrong, however, is normally accompanied by even stronger feelings – such a person is full of contrition.

By saying that conceiving of oneself as wrong *normally* is accompanied by even stronger feelings than feelings of regret or remorse, it is apparent that there are exceptions. These exceptions belong to the second category of a first-personal conception of wrongness of oneself. Before characterizing the exceptions, let me recapitulate thesis T4 about conceiving of oneself as wrong:

> A person, by way of a meta-reflection on herself, may conceive of *herself* as wrong *now* if she suffers from having (frequently) violated her values and therefore wants to change herself, *or* she may conceive of *herself* as wrong *retrospectively* if the person conceives of herself as having had 'wrong values'.

A person who *retrospectively* conceives of herself as having had 'wrong values' judges that she (during a certain period of her life) cared about the wrong 'things'. She conceives of herself as having changed essentially. Such a person may still be full of remorse about the fact that she cared about the wrong 'things', but those feelings may not come to the surface frequently (anymore) or may have disappeared completely. The person may manage to avoid those feelings, for example, by attributing the fact that she cared about the wrong 'things' to certain circumstances of her childhood, to the requirements of her job, or to the influence of 'wrong friends'. But psychological explanations are not relevant in the context of this chapter. The point is that a person who retrospectively conceives of herself as having been wrong (during a certain period of her life) and whose judgement is *not* accompanied by corresponding feelings does not really bother (anymore) that she was wrong. These are the exceptions which I mentioned earlier and which I will leave aside for the rest of my analysis of the question 'Why does it bother us to conceive of ourselves (qua self) as wrong?'

Persons who have the property of making qualitative distinctions regarding kinds of persons (or kinds of life) and who thus have the ability to

apply self-regarding qualitative standards, when reflecting on (their motivations for) their actions, conceive of themselves as beings who have the ability to judge whether (their motivations for) their actions accord with their values, or in other words, as beings who are able to assess their purposes and actions as right or wrong. But this way of conceiving of ourselves would be threatened if we were to recognize (by judging and feeling) that we in fact (frequently) do not use that ability, or did not use that ability during a certain period of our life, or were mistaken in the qualitative distinctions that we made in that period – in short, if we conceived of ourselves as wrong. When a person, reflecting on herself qua self, conceives of herself as wrong, she cannot justifiably conceive of herself anymore as a person who is able to assess (her motivations for) her actions as right or wrong.

I believe that a being which in principle has the ability to apply self-regarding qualitative standards strongly desires to be able to justifiably conceive of herself as a being which has that ability, and that this desire is frustrated when a person conceives of herself as wrong. I call that desire our *desire to be worthy of self-esteem*. Let me explain why I use that label. A person who has the ability to assess her purposes and actions as right or wrong conceives of herself as the kind of person who has that ability, and she strongly desires to justifiably conceive of herself in that way, in other words, she cares about being a person who assesses her purposes and actions as right or wrong. Caring means behaving in a way that ensures the flourishing of the object of one's caring, and being sensitive to situations where the object of one's caring is in danger. And a person who in fact cares about herself as a person who assesses her purposes and actions as right or wrong – otherwise she would not be able to conceive of herself as wrong – and conceives of herself (qua self) as wrong, is being torn by remorse that she in fact did not care about herself as the kind of person who assesses her purposes and actions as right or wrong. After all, in a (substantial) number of concrete situations or during a certain period of her life, she did not use her ability to reflect on (her motivations for) her actions in the relevant way; that is why such a person feels unworthy of self-esteem. By the way, there may be situations where persons, by way of self-reflection, feel unworthy of self-esteem because of 'just' one action by which they sacrificed their 'deepest values'.

We might phrase the following thesis:

(T6) A person who conceives of herself (qua self) as wrong is confronted with her own judgement – accompanied by strong feelings of remorse – that she actually did not care about herself as a person who is able to assess her purposes and actions as right or wrong. But persons who care about themselves in that way strongly desire to justifiably conceive of themselves as persons who assess their purposes and actions as right or wrong. That is why a person who conceives of herself (qua self) as wrong feels unworthy of self-esteem.

I do not know whether it happens frequently that persons conceive of themselves as wrong, and what we might conclude if it happened seldom. I can think of four possibilities: (a) persons seldom reflect on themselves qua self in the relevant way, that is, qua being the kind of person who has the ability to assess her purposes and actions as (right or) wrong; (b) persons rarely perform actions which they conceive of as wrong; (c) most persons have an emotional barrier to conceiving of themselves as wrong, and thus as not worthy of self-esteem; or (d) a combination of the other three explanations.

Presuming that persons characteristically have the property of making qualitative distinctions between certain kinds of persons (or kinds of life) and are able to apply corresponding self-regarding qualitative standards when reflecting on (their motivations for) their actions, we can exclude possibility 'b'. Possibility 'a' has a higher probability, because there may be a considerable number of persons who are (mentally) unable to perform the required meta-reflection, have not developed that capability, or 'just' do not engage in the required kind of self-reflection. But, in my view, the main reason is given by possibility 'c': It might be a mechanism of self-preservation, that people have an emotional barrier to feeling unworthy of self-esteem; after all, people who judge themselves unworthy of self-esteem are in a deep mental crisis.

9 Concluding summary

I explained some essential differences between purposes and values and showed their respective roles in our evaluative reasoning. Based on my view on the notion of 'having values', I developed the notion of a first-personal conception of wrongness, and argued the following theses:

(T3) *Conceiving of one's own action as wrong:* Conceiving of one's own action as wrong (before, during, or after the performance of that action) means: a person judges that her action does not accord with one (or more) of her self-related qualitative standards which she thinks and feels she should not violate or have violated (in the situation in question), because she cares about being the kind of person whom the value(s) in question can be justifiably attributed to.

(T4) *Conceiving of oneself as wrong:* A person, by way of a meta-reflection on herself, may conceive of herself as wrong *now* if she suffers from having (frequently) violated her values and therefore wants to change herself, *or* she may conceive of herself as wrong *retrospectively* if the person conceives of herself as having had 'wrong values'.

I then discussed the overriding nature of a first-personal conception of wrongness of action with respect to all other possible ways of conceiving of one's action (regarding aspects of rational agency). Finally I tried to answer the 'mattering-questions' and presented the following theses:

(T5) *Conceiving of one's own action as wrong matters to us:* A person who conceives of certain of her actions as wrong is confronted with disloyalty of herself to herself. But persons have the desire to be loyal to themselves, because, by virtue of caring about being the kind of person who has certain values, they are concerned with their values and desire to justifiably conceive of themselves as persons who have these values.

(T6) *Conceiving of oneself as wrong matters to us:* A person who conceives of herself (qua self) as wrong is confronted with her own judgement – accompanied by strong feelings of remorse – that she actually did not care about herself as a person who is able to evaluate her purposes and actions as right or wrong. But persons who care about themselves in that way strongly desire to justifiably conceive of themselves as persons who assess their purposes and actions as right or wrong. That is why a person who conceives of herself (qua self) as wrong feels unworthy of self-esteem.

Chapter 5
Selfhood: Unity in changeability

Looking at a photograph labelled 'my school class, 14 years old', which I had not seen since then, I tried to find out which of the girls was me. Eventually, I found myself. I tried to dig out memories of that period of my life: I could think of my first dancing lesson, a bicycle accident that brought me into hospital, and the teacher who made me enthusiastic for mathematics. I didn't have any doubt that it was *me* who attended that dancing lesson, that it was *me* who had the accident and went to hospital, and that it was *me* who enjoyed the mathematics lessons of that teacher; I even seem to be able to recall how *I* felt in those situations. Thinking about those three events now, I connect them to certain character traits that I had when I was young, to certain experiences that followed those events, and to certain goals that I, later on, tried to achieve. But do those three events from my youth really stand in these relations to my self? Do I really know myself? After all, thinking about myself when I was fourteen years old, I only came up with three events; I couldn't even recognize myself on a photograph. I decide to talk to my parents and school friends, when I visit them next year, about that period of my life. By the way, I have no doubt that *I* will be the one who will talk to them. However, it is possible that by then I will not be interested anymore in analysing those events and their connection to myself as I am now; other things may then be much more important to me.

The example illustrates, on the one hand, that we are *sure* to be the ones who remember, feel, reflect upon, consider, and plan certain things at a certain moment, and, on the other hand, that we are *not sure* about our past experiences and actions, their interrelations, their relations with our character traits and with our current beliefs, goals, and plans, about the stability of those goals and plans in the future, and about the things that might become important to us. In short, the example presents two pre-philosophical intuitions: our certainty of our continued selfhood, and our uncertainty about who we are, that is to say, our uncertainty about our personality[1].

The tension is problematic to us, for we have a desire for unity of our self.

---

[1] I use the term 'personality' in a broad sense, covering, for example, one's character traits, dispositions, adopted long-term goals and plans, lasting practical beliefs, the objects (for example, traditions, ideals or persons) one cares about or which are important to oneself, or moods that are characteristic of oneself.

I will argue that one's desire for unity of one's self – a desire that expresses itself in certain ways of self-reflection – is based on one's sense of one's continued selfhood. The view I am drawn to will emerge as I successively address, in the seven sections of this chapter: (1) one's sense of selfhood; (2) one's sense of one's continued selfhood; (3) one's self-conception as a rational agent; (4) one's awareness of the changeability of one's self; (5) embedded self-reflection and the desire for unity of one's self; (6) explicit self-reflection and the desire for unity of one's self; and (7) meta-reflection about oneself and the desire for unity of one's self.

1 One's sense of selfhood

Persons are beings that conceive of themselves as themselves.[2] This means in particular (the following list is not meant to be complete but to cover those aspects that are relevant at this stage of the argument) that a person: (a) thinks of herself as the subject of her behaviour, thoughts, feelings and experiences; (b) is aware of the fact that she has her own perspective and that this perspective differs from the perspective of others; (c) thinks of herself as a subject that is aware of itself; (d) can think of herself qua self in the future and in other circumstances; and (e) can take a position regarding her behaviour and reflect on it. Such a being conceives of itself as a self.

Conceiving of oneself as a self means more than mere self-consciousness; it is an expression of reflective self-consciousness. Reflective self-consciousness means thinking of oneself as a subject that is conscious, whereas 'mere' self-consciousness means being pre-reflectively aware of oneself. The latter – pre-reflective self-consciousness – is a prerequisite for reflective self-consciousness; after all, one can only reflect on something that is already 'given' to oneself. How could one ever identify oneself as the 'thing' one wants to relate to in an act of self-reflective thinking or feeling if one did not already 'know' oneself?[3] We would be caught up in an infinite regress, wouldn't we?[4] But we are immune to the error of misidentification,

---

[2] As I argued in the foregoing chapters, especially in Chapter 2 (The relation of oneself to one's self), the property of conceiving of oneself as oneself, which is the property of having (a highly developed form of) a first-person perspective, is the fundamental property of persons. On top of that property, persons characteristically also have the non-fundamental properties of a narrative self-conception and of applying self-regarding qualitative standards; and those two properties – as I argued earlier – are both not entailed by the property of a first-person perspective, although they require it.

[3] See also Frank 2002, pp. 127-128, especially "I cannot identify any object *as myself* if I did not have *self*-knowledge before the identification" (translated by myself; italics in original, correspondingly).

[4] For explications of this infinite regress, see, for example, Frank 2002, pp. 127-128 and 240, Shoemaker 1986, p. 13, or Zahavi and Parnas 1999, pp. 259 and 261.

[5] For a denial of the possibility of misidentification, see, for example, Baker 2000, pp. 136–137, Shoemaker 1986, pp. 11-12, or Zahavi and Parnas 1999, p. 262.

because we do not identify ourselves at all.[5] We have a pre-reflective sense of identity. Let me explain.

One's subjectivity reveals itself to oneself[6], so to speak, in one's being conscious of experiencing a certain external object and in being conscious of one's feelings or thoughts. We may call this kind of awareness – which means to be conscious not only of an external object but of one's experience of the object as well[7] – self-awareness in a fundamental sense. To avoid misinterpretation, I should emphasize that the distinction between being conscious of an external object and being conscious of one's experience of that object is a conceptual distinction that does not imply an epistemological or a phenomenological distinction. More specifically, it does not imply that a creature's being conscious of an external object is presented to a certain feature of that creature, and that the creature then becomes conscious of experiencing that external object and of (to use Thomas Nagel's well-known expression) 'something it is like' to experience that object.[8] Being conscious of an external object is always a *subject's* consciousness of experiencing that object. That is why we might call that kind of consciousness self-awareness in a fundamental sense. It is fundamental for two reasons: it is the most basic kind of self-awareness and it is pre-reflective.

But self-awareness in this fundamental sense is not sufficient for being pre-reflectively self-conscious, at least if, as I take it, pre-reflective self-consciousness can provide us with 'something' we can relate to in an act of reflective self-awareness. Consciousness of one's experience of an external object or of one's feeling does not 'give' oneself to oneself as an object that one could relate to in an act of reflective self-awareness. For example, a subject's consciousness of feeling tired and its consciousness of watching television do not entail in itself that the subject experiences itself as one and the same tired and television watching subject. The pre-reflective sense of being one and the same subject in both acts of pre-reflective self-awareness is, however, a prerequisite for the subject to perform acts of reflective self-awareness in that concrete situation, for example thinking about whether to go to bed or keep watching television. What is lacking in self-awareness in its fundamental sense in order to count as pre-reflective self-consciousness is a pre-reflective *sense of selfhood*. A sense of selfhood

---

[6] The expression is derived from Zahavi and Parnas 1999, p. 255: "It is possible to speak of self-awareness the moment I am no longer simply conscious of a foreign object, but of my *experience* of the object as well, for in this case my subjectivity reveals itself to me." (Italics in original)

[7] Zahavi and Parnas 1999, p. 255; I refer to the same sentence as in the foregoing note.

[8] See for example: Manfred Frank's argumentation for the functional unity of being conscious of something ('Empfindungsbewusstsein') and pre-reflective self-consciousness, in Frank 2002, p. 133; Zahavi and Parnas' elaboration on the 'first-personal givenness of phenomenal consciousness', which they speak of in terms of self-awareness, in Zahavi and Parnas 1999, pp. 255-256; or Sydney Shoemaker's 'objection to the idea that there is introspective perception of anything whatever', in Shoemaker 1986, pp. 19-20.

entails more than consciousness of one's experience of an external object or of one's feeling, it entails synchronic unity; more specifically, it requires that the subject pre-reflectively experiences itself as one and the same perceiving, thinking, and feeling subject in various acts of pre-reflective self-awareness that may appear at the 'same' moment. Otherwise – that is, without a sense of selfhood – a being might experience itself as the subject that perceives a certain external object and as the subject that has a certain feeling without experiencing the two subjects as one and the same.[9] But only a creature that pre-reflectively experiences itself as one and the same perceiving, thinking, and feeling subject, in the present moment[10], exists as a self for itself, in that moment. Its self appears to that creature as the certainty of *being-self*, and by this as 'something' constant that it can relate to, in that moment, in reflective acts of, for example, evaluating conflicting desires or explaining why it is doing what it is doing. Such a being has a sense of identity[11], in the present moment.

The question might arise how we pre-reflectively come to know ourselves in the relevant way. Developmental psychology may be able to give an answer here.[12] George Butterworth, for example, describes observations in support of ecological and interpersonal aspects of self that are directly perceived and are important "in underpinning later, more cognitively defined aspects of self".[13] According to him "[t]he mental reflective self is just one relatively late-developing component of self. Although adult introspection may (unreliably) suggest differently, introspection cannot reveal the inter-related, mutually-embedded levels of self awareness which have been shown by empirical inquiry." [14]

---

[9] Manfred Frank raises a similar point when arguing that selfhood requires a certain intimacy with oneself ('Vertrautheit'), without which we could not talk about *the* phenomenon of the self. Without that intimacy with oneself one could not identify one's being conscious of experiencing a certain external object and one's being conscious of one's feelings or thoughts as two manifestations of one peculiar phenomenon. We, possibly, would be other subjects (or none at all) as perceiving beings than as thinking beings. (See Frank 2002, p. 144) This passage is not a translation of Frank's argumentation but my interpretation of his point and applied to my line of thought.

[10] Admittedly, a moment still is a certain period of time. But I will not enter into the discussion here about what period of time may count as a moment in the relevant sense. According to Galen Strawson's 'Pearl view', for example, that moment might be defined as an 'uninterrupted or hiatus-free period of consciousness'; see Strawson 1997, p. 21.

[11] I was inspired by Frank 2002 to use the term 'sense of identity'. He cites (p. 171) Christoph Meiners (1776), who uses the German term 'Gefühl der Identität', and who characterizes it (according to Frank, and free translated by myself) as the ability to keep various (internal and external) mental modifications 'consciously' present in the same moment and to relate them to a numerical unique entity, the person, that does not lose itself in the change of impressions.

[12] See, for example, Maria Legerstee's proposal about the mechanisms that lead to the development of self-awareness in human infants during the first months (Legerstee 1999), and George Butterworth's developmental-ecological approach (Butterworth 1999). Both authors react, in the referred articles, to Galen Strawson's view (see Strawson 1997) that our sense of self is the sense of a mental self.

[13] See Butterworth 1999, p. 206.

To sum up: Acts of reflective self-awareness – for example, deliberating about what to do in a situation of conflicting desires, which could move us to act in a certain way – require a pre-reflective sense of selfhood, which in turn entails a sense of identity. After all, when I, in a concrete situation of conflicting desires, take a position regarding those desires and reflect on them, I already know – without performing any reflective act of identification – that *I* am the one who has those different conflicting desires; and that means that I pre-reflectively exist for myself as a self – at least, in the present moment. But most, if not all, capabilities that are linked to personhood – that is to say, which are based on the essential property of persons to conceive of themselves as themselves – seem to require not only a sense of selfhood in the present moment but also a sense of continued selfhood; think, for example, of a person's ability to think of herself qua self in the future and in other circumstances.

2 One's sense of one's continued selfhood

Let us imagine a person who has a sense of selfhood, in the present moment. Such a person pre-reflectively experiences herself as one and the same perceiving, thinking, and feeling subject, in the present moment, and therefore has the certainty of *being-self* and a sense of identity, in the present moment. But the person might not have any sense of personal continuity from the past to the present and into the future; she may, for example, have lost that sense due to brain damage. This person would be very, very different from us.[15] She can, for example (to recall some of the capabilities that are relevant to

---

[14] See Butterworth 1999, p. 204. Butterworth refers to Neisser (Neisser, U. (1988), 'Five kinds of self knowledge', *Philosophical Psychology,* 1, pp. 35-39) and writes (p. 204): "Although Neisser's taxonomy [of self-knowledge] was not intended to be a developmental description, the aspects of self he describes make their appearance at different times in development and help in explaining the origins of a distinctively mental sense of self. The five aspects of self are: 1. The ecological self, which is directly perceived with respect to the physical environment. 2. The interpersonal self, also directly perceived, which depends on emotional and other species-typical forms of communication. 3. The extended self, which is based on memory and anticipation and implies a representation of self. 4. The private self, which reflects knowledge that our conscious experiences are exclusively our own, and this is also dependent on representation. 5. The self concept, defined as a theory of self based on socio-cultural experience."

[15] Andrew Brook gives the following illustration: "Oliver Sacks has explored cases in which virtually all sense of continuity over time has been lost (for example, cases in which the patient has to [be] introduced to the medical staff anew on every meeting). While these patients continue to be aware of themselves as they are in the present moment (ask if they are happy and they can answer yes or no), something is profoundly different. […] Now consider a hypothetical patient with no ability to anticipate a future as his or her own. […] A person who could not think of him- or herself as existing in the future, who could not, for example, plan what to have for dinner, wonder what tomorrow will hold, think about who might visit next week, would be even more uncanny than the patients with no ability to form memories." (Brook 1999, p. 46)

(fully developed) personhood) not take a position regarding her behaviour of a few moments ago and reflect on that behaviour, and she cannot think of herself in the future and in other circumstances. The question could be raised whether this human being still is a person, but I will not address this question here. Whether we call such a human being a person or not[16], it seems clear to me that a person without a sense of *continued* selfhood lacks capabilities that are essential for (fully developed) personhood.[17]

Unity and permanence of oneself qua self, that is diachronic unity, are implied in various acts of reflective self-awareness that are characteristic of persons. Unity and permanence of oneself are, for instance, implied in: (a) the willingness to accept, or the attempt to avoid the consequences of one's actions; (b) the feeling of responsibility for one's actions; (c) the adoption of long-term purposes; (d) the explanation of one's actions by referring to one's purposes or desires, to certain earlier experiences, to certain adopted goals, or to a combination of (some or all of) those in the form of a self-narrative; (e) the act of setting behavioural rules for oneself; and in (f) caring about being a certain kind of person.[18] Let me recall why all those acts are acts of reflective self-awareness: they are reflective, because they require that one thinks of oneself as oneself, and thinking of oneself as something is a reflective act of self-awareness. But the aforementioned acts do not only require that one thinks of oneself as oneself, they require that one thinks of oneself as a self *over time* – the self that is, has been and continues to be. The question is whether such permanence and unity that are implied in those acts are based on a pre-reflective sense of *continued* selfhood (just as a person's conception of herself as herself is based on a pre-reflective sense of selfhood) – or, rather, on a person's conception of herself as herself and on her belief in her long-term continuity.

When we envisage concrete situations of performing the aforementioned reflective acts, there seems to be a major difference between the way permanence and unity are implied in, for example, feeling in general responsible for one's actions, or being willing to accept or wanting to avoid the consequences of one's actions, on the one hand, and in setting behavioural rules for oneself, or adopting long-term purposes, on the other hand. When we feel responsible for our actions, we conceive of ourselves as

---

16 There may be various reasons to call a being that has lost the sense of continued selfhood a person. For example, we may: (a) not know whether the damage is irreparable; (b) not know whether *all* sense of continuity is lost; or (3) want to keep that human being within the community of persons for legal or ethical reasons.

17 Andrew Brook, for example, argues: "For one thing, beings with *no* sense of personal continuity would also a*nd for that very reason* be cognitively impaired in a certain way. The ability to retain the contents of earlier experiences and, via remembering (or q-remembering) them from the inside, synthesize them with current experiences is an unsubstitutable feature of virtually all forms of cognition of any complexity. But when I (q-)remember an earlier experience from the inside, I also remember that experience as though *I* had it." (Brook 1999, p. 46; italics in original)

18 Regarding examples (a) and (b), see also Gallagher and Marcel 1999, p. 296.

agents that are responsible for what they do. Our feeling responsible for our actions in this general sense is not linked to particular actions or to particular periods in our life, but to every event – without having any specific event in mind – which we, at a certain moment, might conceive of as our own action. Such a feeling of being in general responsible for one's actions requires a conception of oneself as an agent who acts based on her own motives, and that in turn means that it requires a conception of oneself as a continuous self rather than as a momentary one. And it seems to me, because I believe that experiencing ourselves as agents belongs to our pre-reflective self-awareness[19], that this conception is simple instead of twofold. One does not first conceive of oneself as a momentary self and then comes to realize that one has been and continues to be. We pre-reflectively experience ourselves as one and the same perceiving, thinking, feeling, and *acting* subject. And because actions can consist of more or less complex behavioural structures across a period of time and may need adjustment during their performance, a creature that pre-reflectively experiences itself as an acting subject is 'given' to itself as a continuing self. That is why our conceiving of ourselves as existing over time is not a twofold act of reflective self-awareness but a simple one. Consider the example of being willing to accept or attempting to avoid the consequences of one's actions. When I recognize that an action which I am thinking of performing has certain consequences for myself, I may deliberate whether I am willing to accept or would rather avoid those consequences and thus whether to perform, slightly modify, or completely refrain from that action; but whatever I decide, the very fact that I think of certain consequences of the action as consequences for myself means that I relate, in the act of reflecting on my action, to myself as a self that continues to be.

Our experience confirms that there is not first a momentary self that is then, by way of a reflective act of self-awareness, extended to the conception of a continuing self. If we recognize that an action, which we are thinking of performing, will or may – immediately or in the short term – have undesirable consequences for ourselves, we normally will evaluate whether we are prepared to accept these consequences or whether we prefer to refrain from that action. However, if the undesirable consequences for ourselves are to be expected in the long term, we sometimes ignore those consequences or take them less seriously into account. In other words, we seem to be less inclined to take the long-term consequences into account than the short-term ones. A possible explanation of this phenomenon is that we perceive the immediate or short-term consequences of our actions as directly connected

---

[19] Shaun Gallagher and Anthony J. Marcel write: "[E]cological self-awareness gives us more than just a snap-shot profile of our posture, location and action. Implicit in this kind of self-awareness is a sense of *what I have just been doing*, and, of equal importance, *what I can do*, and *what I am just prepared to do*, a sense of capability which goes beyond the momentary." (Gallagher and Marcel 1999, p. 292; italics in original)

to the action in question, whereas we have to think of ourselves qua self in the future and in other circumstances in order to conceive of the long-term consequences of our actions as consequences for ourselves. This explanation would support our line of thought concerning a pre-reflective sense of continued selfhood in the following way: because, firstly, the direct consequences of one's actions are perceived as connected to the action in question, secondly, experiencing ourselves as acting subjects belongs to our pre-reflective self-awareness, and, thirdly, actions and their direct consequences for oneself range over a period of time, a creature that pre-reflectively experiences itself as an acting subject is 'given' to itself as a continuing self.

Now think of the way we set behavioural rules for ourselves and adopt long-term purposes. When we think of long-term goals for ourselves, we, firstly, have to think of ourselves in the future and in other circumstances. Secondly, we will have to think about ways to achieve those goals, because the processes required for their realization will not be readily available to us. Thirdly, we sometimes – for example, because of the complexity or vagueness of the required processes – will define behavioural rules in support of the achievement of our long-term goals. All three aspects make intelligible that we will not conceive of the complex of processes that is required for the actualization of a long-term goal as an *action* – the required behavioural structures are far too complex and too dependent on incalculable circumstances. And if, as I argued so far, our pre-reflective sense of continued selfhood is predominantly linked to pre-reflectively experiencing ourselves as acting subjects, then the unity and permanence of oneself qua self that are implied in adopting long-term purposes and setting behavioural rules for oneself seem to require conceptualization of oneself as a self over time – more specifically, as existing in particular circumstances in the future (or in the past) – instead of being solely based on a pre-reflective sense of continued selfhood. This conclusion is supported by the fact that we often take our age and finitude into account when thinking of long-term goals for ourselves. Think of a person who decides to put some of her monthly income aside in order to save up enough to buy a house. Normally, a person will not make that plan if she cannot realistically expect to be still alive by the time she will have saved up enough money. And taking one's age or finitude into account requires a conceptualization of oneself as a self that exists through time.

To sum up: Various acts of reflective self-awareness that are characteristic of persons imply the unity and permanence of oneself qua self. I argued that we can distinguish between two kinds of those acts: firstly, acts in which we relate to ourselves as subjects of our actions, which in turn may consist of complex behavioural structures across a period of time and may need adjustment during their performance, but which we still perceive as actions; and, secondly, acts that require that we think of ourselves in the future or in other circumstances and in which we relate to

ourselves as initiators of processes or as subjects that are involved in processes. In the former acts of reflective self-awareness, one's self is pre-reflectively given to oneself as a self that has been and continues to be; the latter, by contrast, require conceptualization of oneself as a self over time – or even as existing at a particular point in time or in particular circumstances – instead of being solely based on a pre-reflective sense of continued selfhood.[20]

A further question might be what period of time is covered by our pre-reflective sense of continued selfhood. I must admit, that I am not sure about the answer. But if I am right that our pre-reflective sense of continued selfhood is tied to our perceiving ourselves as agents, then the period of time that is covered by our pre-reflective sense of selfhood seems to be connected to the degree of complexity of our actions. I am therefore drawn to the view that our pre-reflective sense of continued selfhood covers a certain period of time, the length of which may vary from person to person and during the life of a person. Anyway, our pre-reflective sense of continued selfhood does *not* cover the complete period from birth, or more specifically from becoming a person, to a distant future. In situations in which we are engaged in acts of reflective self-awareness that involve conceiving of ourselves as living a life – as is the case when we, for example, think of long-term goals for ourselves, explain our actions by presenting an autobiographical narrative, or ask ourselves what kind of person we want to be – the unity and permanence that are implied in those acts are not directly based on an immediate sense of such unity and permanence, but on a conceptualization of oneself qua self as existing through time. And conceptualization of oneself qua self as existing through time is definitely not a pre-reflective sense of continued selfhood but a detached, reflective sense of continued selfhood, whereby the latter presupposes the former; after all, without a pre-reflective sense of continued selfhood we would lack the 'givenness' of a self that has been and continues to be, and thus would have nothing to relate to and conceptualize.

To conclude: Firstly, a pre-reflective sense of continued selfhood provides us with the certainty of *being-self* over time. We pre-reflectively experience ourselves as a self that has been and continues to be, and by this as 'something' constant over time. A pre-reflective sense of continued selfhood thus entails *a sense of identity over time*. Secondly, the period of

---

[20] See also Gallagher and Marcel's view on the temporally extended self, as presented in Gallagher and Marcel 1999, Section 'Ecological self and temporal extension', pp. 292-294. According to their view, "proprioceptive and ecological awareness also must include a sense of self over time" (p. 292). They further argue that the self expressed in action is temporally extended, more specifically, that one's actions are nonreflectively informed "by one's past experience, by beliefs, lasting attitudes, moral positions, by one's personal knowledge, concerns, and practical interests" (p. 293), but they doubt that people have an immediate sense of their long-term continuity.

time which is covered by our pre-reflective sense of continued selfhood and identity is not a lifetime or, for example, a fixed period before and after the current moment, but, if I am right, depends on the time frames of events that we pre-reflectively experience as actions; after all, our pre-reflective sense of continued selfhood seems to be linked to pre-reflectively experiencing ourselves as acting subjects. Thirdly, given our pre-reflective sense of continued selfhood and identity, and because we, moreover (after all, we are persons), have the property to conceive of ourselves as ourselves, we conceive of ourselves as the self that has been and continues to be – as existing as one and the same 'thing' over time. But as far as our property to conceive of ourselves as ourselves is concerned, that reflective act of self-awareness means nothing more than conceiving of oneself as *being-self*. The self is a content-free self, so to speak. And identity in this sense is a 'neutral identity'[21], precisely because a notion of the self as *being-self* is the notion of a 'content-free' self. Finally, owing to our reflective capabilities, we are able to think of ourselves qua self in the future and in other circumstances, that is to say, beyond the future and the circumstances that are within the scope of a concrete action. And such a reflective act requires conceptualization of our existence as a self over time, a conceptualization which in turn depends on one's conceiving of oneself as a self.

3 One's self-conception as a rational agent

Conceiving of ourselves as ourselves is fundamental for personhood. Based on that property and on capabilities that are dependent on it – especially the capability to reflect on (one's motives for) one's actions, we have the general self-conception of being rational agents; that is to say, we generally think of ourselves as agents that decide what to do and how to do it.[22] This does, however, not imply that every individual person would assent to the sentence 'I think of myself as being a rational agent' or 'I think of myself as an agent that decides what to do and how to do it'. One might even lack the capability to understand these sentences and yet comprehend oneself as

---

[21] The term 'neutral identity' is derived from Charles Taylor (Taylor 1989). See note 7 in Chapter 2 (The relation of oneself to one's self).

[22] More specifically: (a) we conceive of our actions as our actions, in the sense that the actions would not have happened if *we* did not have certain beliefs, desires, and intentions; (b) we conceive of our actions as having motivating reasons and as, in general, being intelligible, presupposing one knows those motivating reasons; and (c) we conceive of our actions as being motivated by motives which we want to be motivated by to perform the action in question. This is not to deny that, in a concrete situation of doing something, one or more of the aforementioned characteristics may be lacking. We may, for example, find ourselves doing something without knowing why we are doing what we are doing, or we may know that a certain desire has moved us to act but we would rather have rejected that desire. Nevertheless, such situations do not normally influence a person's general conception of herself as being a rational agent.

being a rational agent. The self-conception of being a rational agent does not have to be highly reflective, much less articulated.[23] The fact that we conceive of ourselves as rational agents becomes evident when we are engaged in acts of reflective self-awareness that belong to the domain of rational agency, for example when we explain why we are doing what we are doing by providing the purpose or the motive for our action, or when we reject a certain motive for action because it would be selfish to perform that action. But do we, in those acts, conceptualize ourselves as being rational agents? And do we conceptualize ourselves as being a self over time if we, for example, decide to postpone a particular action until the weekend?

Rational agency, in a concrete situation, does not necessarily entail a conceptualization of one's continued selfhood or of oneself as being a rational agent, because certain acts of reflective self-awareness that are most basic for rational agency – for example, specifying the purpose of one's present action, or evaluating one's alternatives in a concrete situation – are not concerned with oneself qua self but with oneself as an agent in the context of a particular action.[24] Let me explain. A person who responds to "What are you doing?" by providing an explanation of her action in terms of the purpose of that action, performs an act of reflective self-awareness, since she takes a position regarding her action and reflects on her action. Hence, such a person is able to provide an account of herself as an agent in the context of action. And that ability obviously requires that she conceives of herself as a rational agent, which in turn requires – after all, (most) actions cover a period of time – that she conceives of herself as a self over time. But, if I am right, providing an account of herself as an agent in the context of a particular action does not entail conceptualization of herself as a self over time or of herself as being a rational agent, because – in the act of reflective self-awareness in question – the person's conception of herself is not detached from herself as the agent in the context of a particular action but embedded in the reflexive act of explaining her action. I label acts of reflective self-awareness that are embedded in the context of action as *embedded self-reflection*.

There are, however, other kinds of acts of reflective self-awareness, which also – like the aforementioned ones – belong to the domain of rational

---

[23] Owen Flanagan writes: "One might simply not have a very reflectively held self-conception. One might even lack the ability to understand and thus to assent to (or dissent from) the appropriately formulated sentences when they are expressed in abstract philosophical terms. Both possibilites are compatible with being a self-comprehending creature." He writes further (p. 52): "[T]here is no incoherence whatsoever in thinking that […] self-comprehension can accrue in environments that are relatively impoverished linguistically and by means of all manner of intrapersonal and extrapersonal feedback mechanisms: by way of feelings of coordination, integration, and integrity, of fit with the social world mediated by the body language of others, and so on." (Flanagan 1990, p. 44)

[24] I was inspired to develop this view by Shaun Gallagher and Anthony J. Marcel's essay 'The self in contextualized action', see Gallagher and Marcel 1999.

agency, but which indeed require a detached stance towards oneself qua self and therefore entail conceptualization and symbolization of one's self. Think (I deliberately use examples in which the person's conceptualization and symbolization of her self is very explicit) of a person who tries to justify her insensitive behaviour towards family and friends by telling an extensive autobiographical narrative, explaining how and why she became the person that she is now; or think of a person who consciously strives to be a certain kind of person and who regularly engages in radical self-reflection in order to change and improve herself qua self. In this kind of acts of reflective self-awareness, the person's attention is explicitly directed at herself qua self. She has torn herself away, so to speak, from the level and context of particular actions. I label this kind of acts of reflective self-awareness as *meta-reflection about oneself qua self*.

There is still another category of acts of reflective self-awareness, also belonging to the domain of rational agency. If we take the degree of detachment from engagement in concrete action as criterion – or rather, the degree of detachment of oneself as the agent in the context of concrete action, this kind of acts of reflective self-awareness has to be positioned between embedded reflection and meta-reflection. Imagine: A man is asked by his colleague about his plans for the weekend; he responds that he will do some work in the garden. His colleague says "I did not know that you liked gardening, I didn't even know that you had a garden." The first man explains that he recently got a girlfriend, who has a garden and really loves gardening, and that he wants to spend as much time as possible with her. The second man, thinking of his own experiences in relationships, remarks "But you shouldn't give up all your own hobbies. Didn't you train for the marathon?" The answer to the first question, presupposing that the man had already made his plans for the weekend before being asked, is an expression of an act of embedded self-reflection; the man states the purpose of a particular, planned action. The second answer is an expression of an act of reflective self-awareness that belongs to the new category. The man interprets his planned action, and explains the planned action by providing a self-narrative. He takes a more detached stance towards the action – as part of his reflection about the planned action – by conceptualizing himself qua self. Eventually, moved by the last remark of his colleague, the man might engage in meta-reflection about himself qua self later on. I label the second kind of acts of reflective self-awareness, which comprises reflective acts that, firstly, are not entailed by one's conceiving of oneself as a rational agent in the context of a concrete action, secondly, are nevertheless related to certain of one's actions, but thirdly, require conceptualization of oneself qua self, as *explicit self-reflection*.

To sum up: Persons conceive of themselves as rational agents, that is to say, as beings that act based on reasons. They engage in various acts of reflective self-awareness that belong to the domain of rational agency. Those acts can be classified into three kinds: embedded self-reflection, explicit self-

reflection, and meta-reflection. In the case of embedded self-reflection, the reflective act is embedded in the context of a particular action, for example when providing the motivating reason for our present or planned action, or when evaluating our motives for action in the current situation. In the case of explicit self-reflection, we take a detached stance towards a particular action and towards ourselves as agent in the context of that action, and conceptualize ourselves qua self. For example, we justify a past action by providing a self-narrative in which we interpret that action in a certain way. In the case of meta-reflection, we take an even more detached stance; our attention is not directed anymore towards a particular action and towards ourselves as subject of that action, but we are explicitly concerned with ourselves qua self. It should be mentioned that in concrete situation of self-reflection the three kinds of acts of reflective self-awareness may sometimes blend into one another. We then switch, so to speak, from one category to the other within one process of evaluative reasoning.

Whereas embedded self-reflection 'merely' requires conceiving of oneself as the self that has been and continues to be, explicit self-reflection and meta-reflection require, on top of it, conceptualization of oneself qua self, which in turn means that we symbolize our self – and symbolization requires content. And as most of us will confirm from their own experience with meta-reflection about themselves qua self, our self seems to have a certain vagueness and instability. We have a 'sense' of who we are in terms of character or personality; after all, we often refer to our character traits when justifying our actions. But that 'sense' does not provide us with a stable self-image. And if we engage in meta-reflection by asking ourselves who we really are, our self seems to slip away. A conception of our self as a vague and unstable 'thing' is, however, in conflict with the self-conception of being 'something' constant over time, which in turn inherently belongs to our conceiving of ourselves as ourselves and which is based on our pre-reflective sense of continued selfhood and identity over time. Let me therefore, in the next section, call attention to a sense of self which conflicts with our pre-reflective sense of continued selfhood, namely one's awareness of the changeability of one's self. It imposes tension on one's (overall) self-conception: *the tension between constancy and changeability*.

4 One's awareness of the changeability of one's self

Persons are able to ask themselves 'Who am I?'. In answering that question, a person may think of her character traits and dispositions, and of certain experiences, goals and values of herself. By answering the same question after a certain period of time, the person may recognize that her answer has changed, that she, for example, now assigns other goals to herself, that certain experiences have lost their meaning, or that certain values have faded away. And when asking the question periodically, at a certain moment, the

person will recognize that her personality has slightly or dramatically changed. But do we indeed periodically ask ourselves 'Who am I?' I am inclined to deny this. Nevertheless, we are aware of the changeability of our self.

The question is whether one's awareness of the changeability of one's self is a pre-reflective sense of self, or a kind of reflective self-awareness. In the first case, two opposite and incompatible pre-reflective senses of self would be 'given' to oneself: firstly, the sense of continued selfhood, which entails a sense of identity over time and which means that one's self appears to oneself as the certainty of *being-self* over time, as 'something' constant that has been and continues to be; and, secondly, the sense of changeability of one's self, which means – at least if we really had a pre-reflective sense of changeability of one's self – that one's self appears to oneself as 'something' unstable over time. The psychological condition of a creature with such a mental make-up is hard to imagine. But from the very moment that this creature conceived of itself as itself, it would be in utmost confusion about itself, to say the least. And if the sense of the changeability of the self were indeed a pre-reflective sense of self, that situation of utmost confusion would be the psychological point of departure of all persons, and thus of each of us. However, to my knowledge, there is no evidence of such being the case. Small children, who conceive of themselves as themselves, do not seem to experience confusion about themselves qua selves. This fact seems to refute the thesis about the pre-reflectiveness of one's sense of the changeability of one's self. But there is still the theoretical possibility that one of the two pre-reflective senses of self is dominant, presumably our sense of identity over time (which is entailed by our sense of continued selfhood). In that case our not being in utmost confusion about ourselves becomes intelligible if we assume that our conceiving of ourselves as ourselves is grounded in the dominant pre-reflective sense of self.[25] This proposal, however, opens the possibility that for

---

[25] Sydney Shoemaker discusses a similar proposal in a different but related context. When analysing a case presented by Parfit, he imagines the impossible case that two streams of consciousness were realized in the two hemispheres of a brain in such a way that there would be no possibility of integrating their contents. For the sake of argument he supposes that what we would have is *one* person with two streams of consciousness, and that a particular experience occurs in one stream but not in the other. It seems to be, so he writes, that such a person is conscious of that particular experience in, say, the right hemisphere of the brain, but not in the left. "This suggests that the property of *being conscious* is really a relational one, and that our example should be described as one in which the experience is conscious relative to one stream and not relative to the other." (Shoemaker 1996, p. 191, italics in original) He then speculates that "the relativization of the notion of consciousness" provides a strategy in dealing with certain issues (for example, issues of repressed wishes, desires, intentions and beliefs, or of psychological continuity): "Let us say that in the normal case, unlike in Parfit's case, there is a "dominant" stream of consciousness; a temporally extended set of states and events that is highly integrated, i.e., characterized by a high degree of unity of consciousness, and includes most of the subject's mental states, or at any rate significantly more of them than any other such set." (p. 192)

some persons the other pre-reflective sense might be the dominant one, which would imply that their conceiving of themselves as selves is grounded in a sense of an unstable self. To my knowledge, there is no empirical evidence for that theoretical option.

Anyway, according to my account of selfhood, there is a conceptual problem with the aforementioned speculative proposal. Awareness of the changeability of the self requires awareness of oneself as a 'thing' that has certain content, in other words, it requires *awareness of personality* – the awareness of, for example, having certain character traits, dispositions, or moods that are characteristic of oneself. But now we are confronted anew with the question whether that self-awareness is pre-reflective or reflective. Let us for a moment presuppose that awareness of personality is a pre-reflective kind of self-awareness. How could we – without any reflection – be aware of the changeability of that personality? Awareness of the changeability of one's self requires awareness of one's personality and is, for that reason, an act of reflective self-awareness. But awareness of one's personality is, contrary to the assumption just made, not conceivable as being a pre-reflective kind of self-awareness either. One's awareness of oneself as a self with a personality, as evidenced by assigning character traits or dispositions to oneself, is an act of reflective self-awareness; after all, assigning something to oneself requires a conception of oneself as oneself and is, therefore, an act of reflective self-awareness. Moreover, it is an act of reflective self-awareness that implies unity and permanence of oneself, for aspects of personality are by their very nature not momentary. After all, being self-consciously aware of a particular character trait of oneself is not like being self-consciously aware of a momentary feeling or thought.

Having clarity about the reflectiveness of one's awareness of one's personality, the question about the required *kind* of reflective act arises: embedded self-reflection, explicit self-reflection, or meta-reflection about oneself qua self. The answer is that all three kinds are possible. Firstly, as we know from our own experience, persons can engage in meta-reflection about themselves qua personality. A person may, for example, ask herself whether certain repetitions in her behaviour happen by accident, or whether they show that she has a disposition to behave that way. Secondly, acts of awareness of one's personality are often acts of explicit self-reflection. Think, for instance, of a person who responds to a request to justify some of her past actions and who, in this context, refers to certain of her character traits or dispositions. Thirdly, acts of awareness of one's personality can also be embedded in the context of a particular action. They are characteristically entailed by processes of evaluative reasoning concerning the action one is considering, the motive for that action, or one's alternatives. A person may, for example, reject an alternative because she assesses that action as risky and she knows that she does not like risky actions; she may decide to perform a particular action because that action will support her long-term career-goal; or she may reject a particular motive for action, because it would be selfish

to be moved by that motive. In these situations, the act of awareness of her personality is embedded in the person's evaluative reasoning concerning what to do and/or how to do it, in the context of a concrete action; in other words, the person takes her personality into account without leaving the context of the present action.

Awareness of one's personality is a prerequisite for awareness of the changeability of one's self. A person will, obviously, not be able to explicate the motives for her actions by referring to a change in her plans, goals, the things that are important to her, or by referring to a change in her character – all those explications being expressions of her awareness of the changeability of her self – without being aware of herself as a self with a personality. But, as for the sense of one's personality, there is the question about the required *kind* of reflective act. Awareness of the changeability of one's self can, firstly, be an act of meta-reflection about oneself qua self. Think of a person who repeatedly, having decided on a career in a particular profession or to practice a certain hobby, is sidetracked by other careers or hobbies. She may engage in meta-reflection, asking herself why she cannot stick to the plans that she makes for herself. That person is clearly aware of the changeability of her self. But a person's awareness of the changeability of her self may, secondly, also be an act of explicit self-reflection. When asked to justify her former actions, a person may, for example, refer to the fact that she, in those days, had other plans and goals, and that other things were important to her then. A person who provides this kind of account of herself is definitely aware of the changeability of her self. Awareness of the changeability of one's self can, thirdly, be an act of embedded reflective self-awareness, namely when it is part of one's evaluative reasoning concerning a concrete decision for action. Take the following example. A person is busy planning her holidays. She sees an advertisement for a summer course about African dance; she thinks of immediately booking that course; but a few moments later she rejects that alternative, because she thinks that she might be not in the mood for African dance when the time comes. This person reasons, in the context of action, as an agent who is aware of her changeable personality.

To sum up: One's awareness of the changeability of one's self presupposes awareness of one's personality. Both are forms of reflective self-awareness, and they imply unity and permanence over time. However, they do not necessarily imply a detached stance towards oneself. In other words, they do not necessarily imply that the person in question conceptualizes and symbolizes her self as a self with a certain (changeable) personality: awareness of one's personality and of the changeability of one's self may be acts of meta-reflection about oneself qua self or of explicit self-reflection, but they can also be embedded in the context of a concrete action.

Although one's awareness of the changeability of one's self is *not* a pre-reflective sense of self and therefore does not conflict on one and the same level with one's pre-reflective sense of identity over time, there is still a

tension – the tension between constancy and changeability – between one's conceiving of oneself as existing as one and the same 'thing' over time[26], which is based on one's pre-reflective sense of continued selfhood, and one's reflective awareness of the changeability of one's self. In the forthcoming sections, I will argue that we strive – whether consciously or not – to relieve that tension. We desire unity of our self, despite our awareness of the changeability of our self. This desire expresses itself in certain ways of conceiving of our actions and of ourselves, and in certain self-reflexive emotions. And in my view, our desire for unity of our self is based on our pre-reflective sense of identity over time.

5 Embedded self-reflection and the desire for unity of one's self

When we explain a present or planned action by stating the purpose or motive of that action, or when we, in a concrete situation of deciding what to do and/or how to do it, evaluate (our motives for) our alternative actions, we are engaged in acts of reflective self-awareness. Those acts are acts of *embedded* self-awareness, assuming that one reasons as an agent in the context of a concrete action, more specifically, assuming that a person does *not*, for example, provide an extensive autobiographical narrative to make her action intelligible to her interlocutors, or ask herself whether the action she is considering fits in with the picture that her friends have of herself. And frequently and regularly – when reasoning concerning a current or planned action – we do not take such a detached stance towards ourselves and conceptualize ourselves qua self. Therefore, providing the purpose or motive for a present or planned action, and evaluating (one's motives for) one's action when deciding what to do or how to do it, are, frequently and regularly, acts of *embedded* reflective self-awareness.[27]

In the numerous occurrences of acts of embedded reflective self-awareness, we sometimes conceive of our actions in particular ways or have certain reflexive emotions. For example, when providing the purpose or motive of our present action, we may, suddenly, regard our action as wrong and experience regret or remorse. Or, when considering performing a certain action, we may come to realize that the action would be irrational. Such ways of conceiving of our actions exhibit, as I will argue, a desire for unity of our self. This is not to say that they *only* exhibit one's desire for unity; a person's regarding her action as wrong because of acting impulsively, for example, also – and perhaps primarily – exhibits her caring about being self-disciplined. Before turning, in more detail, to the question how the desire for unity of one's self expresses itself in acts of embedded self-reflection, we have

---

[26] See Section 2 (One's sense of one's continued selfhood).
[27] See also Section 3 (One's self-conception as a rational agent).

to pay attention to the fact that there are two kinds of evaluative reasoning concerning a concrete action.

When engaged in processes of evaluative reasoning concerning a concrete action, persons may, for instance, give more weight to the fulfilment of certain momentary desires than to others; they may give more weight to the fulfilment of certain momentary desires than to the pursuit of their adopted longer-term purposes; or they may reject their momentary desires for the benefit of the pursuit of certain adopted purposes. These persons decide what to do and/or how to do it based on weighing alternatives. But persons also decide on other grounds whether to pursue a certain purpose or not.[28] They may reject certain desires or alternative actions, adjust purposes that they adopted earlier, or remove certain adopted purposes from their 'evaluative agenda', because performing the corresponding actions would not be in accordance with the self-regarding qualitative standards which they apply to themselves. Such persons conceive of (their motives for) their actions in terms of, for example, honourable and dishonourable, noble and despicable, brave and cowardly, caring and selfish, merciful and merciless, loving and loveless, self-disciplined and impulsive, independent and dependent, well groomed and badly groomed, neat and chaotic, or famous and unknown – to name just a few of those qualitative distinctions, and reject or decide to perform the corresponding actions based on that judgement. I call (with Charles Taylor[29]) the first kind of evaluation 'weak evaluation' and the second kind 'strong evaluation'. Both kinds of evaluation are acts of embedded reflective self-awareness as long as they are performed in the context of one's evaluative reasoning in a concrete situation concerning (one's decisions for) one's action in that situation.[30] In that case they are expressions of the person's ability to provide an account of herself as a rational agent in the context of action.

How does the desire for unity of one's self express itself in acts of embedded weak or strong evaluation? Take the following example. When asked 'What are you doing?', I come to realize – while providing the purpose of my present action – that my action is irrational. At the same time, I

---

[28] My notion of 'deciding what to do' is neutral with respect to the question whether our future is open or not, that is to say, whether we have a free will in the metaphysical sense, or whether our future is governed by deterministic laws of nature. In the last case, our future can still be epistemically open – as David Velleman argues in Velleman 1989 and in Velleman 2000, pp. 22-24 – and nothing more is required to, in general, conceive of our actions as resulting from our own reasoning.

[29] See, for example, Taylor 1982b and Taylor 1985. As to the 'language of qualitative contrast', see especially Taylor 1985, p. 21. As to Taylor's notion of 'strong evaluation', I refer to the by now well-known citation: "Motivations or desires do not only count in virtue of the attraction of the consummations, but also in virtue of the kind of life and kind of subject that these desires properly belong to." (Taylor 1985, p. 25)

[30] Shaun Gallagher and Anthony J. Marcel, too, mention Taylor's 'strong evaluation' as an example of embedded reflexive evaluation; see Gallagher and Marcel 1999, p. 295.

conceive of the action as resulting from my own reasoning, that is to say, I do not suddenly 'find myself' performing that action. But I now assess the action as hindering or preventing the achievement of an adopted purpose that I really want to achieve. And, retrospectively, I would rather have rejected the spontaneous desire, which moved me to act, for the benefit of that longer-term purpose. It is not necessary that someone asks me about my present action; it may also happen that I conceive of my action as irrational while being engaged in executing that action. It may even happen that I conceive of an action as irrational while deciding to perform that action. In all three cases, the person in fact wanted (or wants) to do what she actually does, whereas on the other hand she does not do what she thinks that she would really want to do.[31]

Conceiving of our action as irrational is an expression of our recognition that the action hinders or prevents the achievement of a purpose which we, during our evaluative reasoning concerning that action, should rather have given more – or rather, most weight to. However, I am inclined to interpret our regarding our action as irrational also as an expression of our desire for unity of our self, more specifically, as an expression of our desire for synchronic unity of our will – a desire which inherently belongs to our general self-conception of being rational agents. In conceiving of ourselves as rational agents, we conceive of ourselves as agents that act based on reasons, or so I have argued. And rational agents, at least when they act rationally, are motivated by reasons (to act in a certain way) which they want to be motivated by. But in cases when we conceive of our actions as irrational, there is or was not *one* motive which we want(ed) to be motivated by but there are at least *two*, namely the one we actually are or were motivated by and the one we think that we really want to be or have been motivated by. We might therefore say that, in such cases, there is a 'divergence in our will'. And I believe that we experience this divergence in our will – if we do or did not solve the conflict by rejecting or wholeheartedly accepting a certain motive for action – as a lack of a unity of our self, namely the soundness of our will. And we will only experience something as a lack of unity if we desire that unity.

When we conceive of our action as irrational, we are engaged in weak evaluation; but, as I explained, there is a second kind of assessment of our actions, namely strong evaluation. And strong evaluation may lead to the conception of acting or planning to act wrongly, which – as the conception of acting or planning to act irrationally – is an expression of our desire for unity, although in another way. Let us first pay attention to situations where we might regard our present or planned action as wrong. Firstly, when asked 'What are you doing?' it may happen that, while explaining my action, I conceive of my action as wrong, because I recognize that the purpose of or

[31] I discussed this account of conceiving of one's own action as irrational in Part I of Chapter 3 (Practical irrationality from a self-related perspective).

the motives for my action do not accord with the self-regarding qualitative standards that I implicitly apply to myself and which represent my values. I may, for example, conceive of (my motives for) my action as selfish instead of caring, or merciless instead of merciful, and by this conceive of my action as wrong. Secondly, as for the conception of acting irrationally, the conception of acting wrongly may not only arise when explaining one's action, but also during one's performance of one's action. For instance, when I originally deliberated about my current action, I weighed the alternatives regarding their contribution to the realization of my longer-term purposes. But now, when performing the action, I come to realize that my action is wrong, more specifically, I regard my action as despicable, because I hurt the feelings of my loved ones, and I am full of remorse. In general terms, I recognize that I violate my values by acting as I do, or in other words, that my action does not accord with my own self-regarding qualitative standards. Thirdly, it is also possible that I conceive of a particular alternative as wrong when deliberating what to do and/or how to do it. But whereas I normally reject such an alternative, I now suppress my assessment that the action I am considering does not accord with my self-regarding qualitative standards. I decide to perform the action for the benefit of a longer-term purpose, the realization of which would be obstructed otherwise. However, I may keep being 'informed' by feelings of regret or remorse that I violate my values by acting as I do.

In my view, conceiving of our action as wrong is not only an expression of our recognition that, by performing that action, we violate particular values that we care about, but also of our desire for unity of our self. More specifically, it is an expression of our desire for unity in being a certain kind of person, a desire that inherently belongs to our conceiving of ourselves as having certain values. This is not to say that we necessarily conceptualize ourselves as having those values and as being or striving to be the kind of person which those values characteristically belong to. A person who applies strong evaluation implicitly – by way of the self-regarding qualitative standards that she applies to herself in those reflective acts – conceives of herself as having those values and as striving to be the kind of person which those values characteristically belong to. When such a person, in a concrete situation, assesses a particular (motive for) action as not according with her values and nevertheless performs that action, her assessment still being the same or 'informed' by certain reflexive emotions, then she experiences a lack of unity in being the kind of person that has particular values. She experiences a divergence between the kind of person which she wants (her motives for) her action to belong to – after all, she applies the relevant self-regarding qualitative standards – and the kind of person which she judges that (her motives for) action actually belong to and which she thinks of in a negative way – after all, she conceives of her action as wrong. I label this kind of disunity as 'non-integrity of one's self'. And although we, in the case of embedded strong evaluation, experience this non-integrity within the

context of one concrete action, we nevertheless experience it as lack of diachronic unity of our self. Here is why. A person who applies strong evaluation in a concrete situation of evaluative reasoning concerning a certain action is normally inclined – because she conceives of the self-regarding qualitative standards that she applies as her values, and because it is inherent to the notion of values that they are enduring – to make the same qualitative distinctions in other concrete situations of action where those values could be at stake. Consequently, if such a person conceives of her present action as wrong, she does, at least if it is not the first time that she applies the self-regarding qualitative standard in question, experience 'non-integrity of her self over time' and thus a lack of diachronic unity of her self.

One might object that the experience of 'non-integrity of our self over time' cannot threaten our awareness of diachronic unity of our self. After all, one's conceiving of oneself as the self that has been and continues to be concerns a content-free identity[32], whereas the identity that is at stake when a person experiences 'non-integrity of her self over time' is an identity with a certain content, namely her values. Therefore, the person who experiences 'non-integrity of her self over time', which is a lack of diachronic unity of her personality[33], does not necessarily experience a lack of diachronic unity of her self. The objection is right, but this is not the point. I do not claim that, by experiencing 'non-integrity of one's self over time', one experiences a lack of one's continued selfhood. What I claim is this: because of (1) our pre-reflective sense of continued selfhood, (2) our embedded reflective awareness of our personality and of the changeability of our personality, and (3) the tension between (1) and (2), we (4) have a desire to relieve this tension by striving for unity of our self as an agent with a certain personality over time, which (5) in case of a person applying strong evaluation is the desire for integrity of one's self, and which (6) expresses itself when a person conceives of her action as wrong and/or feels remorse or regret concerning that action.

In conclusion: Two acts of embedded reflective self-awareness express our desire for unity of our self. Firstly, when being engaged in embedded weak evaluation of (our motives for) our action, we may conceive of our present action, the action which we are currently thinking of performing, or an action that we planned to perform, as irrational. This act of conceiving of one's action as irrational expresses a desire for soundness of the will instead of divergence in the will. Secondly, when engaged in embedded strong evaluation of (our motives for) our action, we may conceive of our present action, the action which we are currently thinking of performing, or an action that we planned to perform, as wrong. This act of conceiving of one's action as wrong, which is often informed and accompanied by emotions like regret or remorse, expresses a desire for integrity of one's self. Our desire for

32 See Section 2 (One's sense of one's continued selfhood).

33 In the context of this chapter, I count one's values as belonging to one's personality.

unity of the self – or rather, for soundness of the will and integrity, respectively – is frustrated when we conceive of one of our actions as irrational or wrong. By desiring unity of our self, we strive to relieve the tension between our pre-reflective sense of continued selfhood and our embedded reflective awareness of the changeability of our personality (in the broadest possible sense). And I believe that our desire to relieve that tension is based on our pre-reflective sense of identity over time, which is entailed by our pre-reflective sense of continued selfhood.

6 Explicit self-reflection and the desire for unity of one's self

When we explain our present or planned action by not only referring to the motive or purpose of that action but when we present an autobiographical narrative to make our action intelligible to others or to justify the action, we are engaged in explicit self-reflection. After all, although our act of reflective self-awareness is still related to a particular action, it nevertheless requires conceptualization of oneself qua self; in the aforementioned case, it requires the person's conceptualization of herself as the protagonist of her autobiographical narrative. Even if we do not provide an account of ourselves as the protagonist of a self-narrative, as soon as our explanation or justification concerns one of our past actions, we take a detached stance and conceptualize ourselves as the agent of that action. After all, that reflective act of self-awareness is not entailed by a person's comprehension of herself as being a rational agent in the context of a concrete action. Sometimes, we engage in self-reflection concerning a recent action without being asked by others. We may, for example, come to evaluate our past action because we experience certain reflexive emotions concerning that action. We also, normally, engage in explicit self-reflection when we evaluate alternatives that concern certain longer-term purposes that we want to achieve. Those longer-term purposes are not just purposes of actions that we plan to perform at a later moment, but goals that require complex realization processes, typically dependent on various circumstances and during a longer period of time. When deciding whether to pursue such a goal, we normally take a detached stance and evaluate, for example, whether this goal fits in with the rest of our long-term plans or what others will think of us if we pursued that goal. It is also notable that persons easily switch to meta-reflection when engaged in this sort of evaluations. In short, what is distinctive of explicit self-reflection is that it does imply a detached stance towards one's self but is nevertheless still related to a concrete action.

In the numerous occurrences of acts of explicit self-awareness, we sometimes conceive of our actions in particular ways or have certain reflexive emotions, which exhibit, as I will argue, a desire for unity of our self. When we, for example, try to explain or justify our planned action by providing an autobiographical narrative, we may come to realize that the action is not

intelligible to us as being part of a coherent self-narrative. Or, when we – informed by emotions like regret or remorse – retrospectively reflect on one of our actions, we may conceive of that action as irrational or wrong.

Let us first consider how our desire for unity expresses itself when a person tries to provide an account of herself as the protagonist of her self-narrative.[34] Take the example of the man who is suddenly interested in gardening.[35] The man is asked by his colleague about his plans for the weekend; he responds that he will do some work in the garden. His colleague says "I did not know that you liked gardening, I didn't even know that you had a garden." The first man explains that he recently got a girlfriend, who has a garden and really loves gardening, and that he wants to spend as much time as possible with her. The second man, thinking of his own experiences in relationships, remarks "But you shouldn't give up all your own hobbies. Didn't you train for the marathon?" The first man responds that the marathon as such was never really important to him, but that he likes to exercise his body and especially in the fresh air, and that gardening, at least partly, fulfils the same function. This man tries (in his second and third answer) to make the change of his hobbies intelligible – to others and himself – as being part of a developing self-narrative. Generally speaking, he tries to make changes of his personality (I count hobbies as belonging to one's personality) intelligible to himself as developments of the self that is the protagonist of his self-narrative. And this means that such a person conceives of her self qua personality as the protagonist of a developing narrative[36], in other words, she conceives of her self qua personality as something changeable that, nevertheless, needs to exhibit a certain unity in that changeability. The unity in question is the coherence of her autobiographical narrative. And in my view, we feel the need for coherence of our self-narrative, and strive for that coherence by interpreting and re-interpreting our actions, because the tension between our conception of ourselves as one and the same self over time and our awareness of the changeability of our personality is relieved when we can conceive of our self as the protagonist of a more or less coherent self-narrative.

The second situation of explicit reflective self-awareness in which our desire for unity expresses itself is when we conceive of a past, present or planned action as irrational. Think of a person who retrospectively reflects on yesterday's action of making a sailing trip instead of continuing to write a certain article. That person explicitly reflects on her decision to make the trip. She may think that it was a good idea to give more weight to the enticing idea

---

[34] I should mention that a self-narrative could be coherent in different ways; that is to say, the narrative has not necessarily to be linear.

[35] The example was introduced in Section 3 (One's self-conception as a rational agent).

[36] I should mention that conceiving of oneself as the protagonist of one's self-narrative does not imply that this self-conception is *highly* reflective or that every individual person would assent to the sentence "I think of myself as the protagonist of my autobiographical narrative". The fact that a person conceives of herself as the protagonist of her self-narrative becomes evident when she, for example, tries to make changes of her personality intelligible as developments of her self.

of a sailing trip than to her adopted purpose to finish her article in the shortest possible time. But she may also conceive of her action as irrational, thinking that she should rather have given more weight to her adopted purpose than to the momentary desire she decided on, because she really wants to finish her article in the shortest possible time and yesterday's sailing trip hinders the actualization of that goal. However, as already argued in the foregoing section, by conceiving of that action as irrational, the person does not only recognize that her action hinders the achievement of an adopted purpose which she should rather have given most weight to during her former process of evaluative reasoning concerning that action, but her conception of having acted irrationally is also an expression of her desire for unity of her self, more specifically, for soundness of her will over time. After all, by retrospectively conceiving of her action as irrational, she judges that there is no unity between the motives she now would rather have been moved by and the ones she actually was moved by to perform the action in question. We might say that such a person conceives of herself as a rational agent over time.

One might object that persons, because of their awareness of the changeability of their self, will especially be aware of the changeability of their momentary desires and preferences, and that it is therefore not intelligible that persons conceive of themselves as rational agents over time. However, this is wrong-headed. We are creatures who can think of and adopt longer-term purposes. And longer-term purposes, by their very nature, are meant to be taken into account when we, in future situations, think about what to do and how to do it; after all, longer-term purposes would have no chance of ever being realized if we did not take them into account in future situations of evaluating our alternative actions. Therefore, a person who still wants to achieve a particular longer-term purpose, the realization of which is hindered by her former action, may judge that she should rather have given most weight to that adopted longer-term purpose in her former process of evaluative reasoning instead of having been moved by the momentary desire which she then chose to fulfil. Such a person conceives of herself as a rational agent over time. And that self-conception is threatened whenever she conceives of her past actions as irrational. Therefore, our desire for diachronic unity of our will does inherently belong to conceiving of ourselves as rational agents over time. I must admit that I believe the time period during which we implicitly desire our will to be united to be relatively short; after all, we do normally not conceive of actions that we performed long ago as irrational *now*. By way of comparison, I would like to recall that our pre-reflective sense of continued selfhood has no time limitation either, but nevertheless only covers a certain period of time.[37] I am therefore inclined to say: As for our pre-reflective sense of continued selfhood, our desire for unity of our self has no time limitation, but this does – and again in parallel with our pre-reflective sense of continued selfhood – not imply that it does not actually cover a limited period of time. To

---

[37] See Section 2 (One's sense of continued selfhood).

sum up: persons, although they are aware of the changeability of their self and are thus also aware of the changeability of their momentary desires, preferences, and goals, nevertheless desire soundness of their will over time. This desire belongs inherently to a person's self-conception of being a rational agent over time, and expresses itself when a person retrospectively conceives of certain of her actions as irrational.

Let us now consider situations where a person, engaged in explicit self-reflection concerning one of her actions, conceives of that action as wrong. Take the person who made a sailing trip the day before instead of continuing to write her article. This person, possibly informed by certain negative reflexive emotions, may reflect on her action and conceive of the action as wrong, judging it to be impulsive instead of self-disciplined. And by making that judgement it becomes evident that she cares about being self-disciplined, and that she conceives of her past action as not belonging to a self-disciplined person. As argued in the foregoing section, a person's conception of her action as wrong is not only an expression of her recognition that she violated particular values which she cares about, but also an expression of her desire for unity of her self, more specifically, for integrity of her self. After all, applying a certain self-regarding qualitative standard in her strong evaluative reasoning implies that a person cares about being the kind of person that accords with the standard in question; and because caring implies continuity of that caring, conceiving of her past action as wrong because of having violated particular values is also an experience of disunity of her self, more specifically, of disunity in her being the kind of person that cares about those particular values. This desire for unity – the desire for integrity of one's self regarding one's values – is inherent to a person's conceiving of her self as having particular values.

In conclusion: Three acts of explicit reflective self-awareness express our desire for unity of our self. Firstly, when engaged in justifying our actions (to others or ourselves) by providing a self-narrative, we, normally[38], feel the need to interpret and 'present' ourselves qua self as the – more or less well-defined – protagonist of our autobiographical narrative, and we, especially, feel the need to make changes of our personality intelligible to ourselves as developments of ourselves qua self. The fact that we feel a need to interpret ourselves qua self in this way expresses our desire for coherence of our self-narrative, which is a desire for coherence in our personality across time. Secondly, when engaged in explicit weak evaluation of one of our actions, we may conceive of that action as irrational. This act of conceiving of one's action as irrational expresses a desire for soundness of one's will over time – a desire that actually, in the concrete situation, covers a limited period of time but has no time limitation in itself. Thirdly, when engaged in explicit strong evaluation of (the motive for) one of our actions, we may conceive of that action as wrong. This act of conceiving of one's action as wrong expresses a desire for integrity of one's self,

[38] I do not exclude the possibility that there may be persons without a narrative self-conception.

which is a desire for unity in one's being the kind of person that cares about particular values.

The fact that we, normally, have a desire for coherence in our personality across time, for soundness of our will, and for unity in our being the kind of person that cares about particular values means that we have an inclination to strive to relieve the tension between our pre-reflective sense of continued selfhood and our reflective awareness of the changeability of our personality (in the broadest possible sense). And our desire to relieve that tension is so inherent to our conceiving of ourselves qua self as the protagonist of our self-narrative, as a rational agent, and as being a certain kind of person respectively – after all, every experience of disunity is a potential threat to the respective self-conception – that, as it seems to me, our desire for unity can only be based in that kind of self-awareness which is a prerequisite for all acts of reflective self-awareness, namely our pre-reflective sense of identity over time.

7 Meta-reflection about oneself and the desire for unity of one's self

Meta-reflection is an act of self-awareness in which a person is explicitly concerned with her self qua self. Think of a person who consciously strives to be a particular kind of person and therefore regularly engages in radical self-evaluation. In the act of meta-reflection, a person is detached from her active engagement with the world; the act of reflexive meta-reflection itself shifts her intentions far away from that active engagement.[39,40] Meta-reflection about oneself requires explicit conceptualization and symbolization of one's self. Admittedly, a person's explicit reflection on and narrative justification of one of her past actions also involves interpretation and conceptualization, and the person's account of her self is therefore a constructed model of her self, instead of, as in the case of embedded self-reflection (which provides an account of one's self in the context of action), part of an instantiated model of her self[41], but it is a model that is still related to herself *as* an agent. Meta-reflection about one's self, by contrast, provides

---

[39] See Gallagher and Marcel 1999, pp. 286-287: "[T]he practice of reflective introspection itself shift and redefine our intentions away from what may have been an active engagement with the world. As in some experimental situations, such questions involve shifting the focus of attention away from purposive activity involving meaningful objects and other persons, to our own movements and modes of consciousness."

[40] Gallagher and Marcel who label such self-reflection as hyper-reflection state: "Such hyper-reflection is a third-order cognitive activity, once removed from reflective consciousness, and twice removed from phenomenal experience or the behavioural level at which we find contextualized action". See Gallagher and Marcel 1999, pp. 286-287.

[41] I was inspired by the following statement in Gallagher and Marcel 1999 to label the model in question an 'instantiated model of one's self' (p. 296): "The referent of such reflection and expression is precisely a self that is instantiated in actions across a variety of personally and pragmatically contextualized situations."

a model of one's self that is based on an explicit, conceptual and objectifying self-thematization.[42,43] That is not to say that, in meta-reflection, we are concerned with our 'total self', so to speak. The reverse is true: when we reflect about ourselves qua self, we are always concerned with a particular aspect of our self. A person may, for example, ask herself whether she cares about things that are worth caring about, why she frequently violates certain values which she really cares about, why she cannot stick to her life plans, or why she frequently does not keep her promises.

Because meta-reflection means explicit conceptual self-thematization, we might at first sight expect that meta-reflection about one's self is exceptional, in the sense that only certain persons engage in such self-reflective acts of thematization of their self. Although I do not want to preclude that there may be persons which lack the cognitive abilities to engage in meta-reflection, persons generally have the capability to conceptualize themselves qua self and reflect upon that symbolic self. But meta-reflection about one's self is exceptional in another sense. Except for persons who go through psychotherapy or psychoanalysis, or, for instance, a minority of people whose life-style requires regular radical self-reflection as in certain religious movements, the circumstances in which persons engage in self-thematization do not appear daily. Persons especially engage in meta-reflection about themselves when they are triggered by certain experiences to reflect upon their own life, when they are forced by the social group they belong to, or when they experience emotions like regret, remorse, contrition, feeling guilty, sadness about themselves, or dissatisfaction with themselves.[44] Here are some examples. Think of a young man who has to explain to his parents why he, after having studied for years, has quitted his

---

[42] See also Zahavi and Parnas 1999, p. 262: "[…] *reflective* self-awareness – which, at least in certain forms, is an explicit, conceptual and objectifying thematization of consciousness" (italics in original).

[43] There is a special issue I want address briefly to avoid misinterpretation: the question whether our self-image and our self-evaluation is 'merely' a matter of cognition. My account of meta-reflection about one's self does not imply that it is completely under our control how we symbolize our self, or in other words, what kind of self-image we have. Publications of (and personal discussions with) psychiatrists have convinced me that one's self-image is constrained by 'one's biology' (whether those constraints are inborn or, for example, arose from trauma, or are an effect of medication). See, for example, Jamison 1997 and Kramer 1997.

[44] Charles Taylor argues that persons have a responsibility for radical evaluation, meaning, as I take it, the radical evaluation of one's values (in a broad sense): "We consider people deep to the extent, *inter alia*, that they are capable of this kind of radical self-reflection. This radical self-reflection is a deep reflection, and a self-reflection in a special sense: it is a reflection about the self, its most fundamental issues, […] and what emerges from it is a self-resolution in a strong sense, for in this reflection the self is in question […] Because this self-resolution is something we do, when we do it, we can be called responsible for ourselves; and because it is within limits always up to us to do it, even when we don't – indeed, the nature of our deepest evaluations constantly raises the question whether we have them right – we can be called responsible in another sense for ourselves whether we undertake this radical evaluation or not." (Taylor 1982b, p. 126; italics in original) I tend to agree with Taylor's view, but I will not address the question in the context of this book.

professional education. He tries to make his situation intelligible to himself by providing an extensive self-narrative explaining why his interests and goals changed so radically for a second time. By the way, the young man might also have engaged in that meta-reflection without being asked by his parents, but, for example, because he felt dissatisfied with himself. Or think of a woman who asks herself why her relations with men again and again end in fiasco. She feels sad about herself and asks herself 'What is it about me?' She, like the young man, may come up with a certain autobiographical narrative, which makes her situation intelligible to herself. Or take the example of a man who, being in midlife, asks himself whether he spends his life in the right way. He may struggle with the question 'What things are really worth caring about? Or think, finally, of a teenage girl who is full of contrition, because she, for the umpteenth time, stole money from her parents to buy drugs. She might ask herself why she does not succeed in changing her life.

As for embedded and explicit self-reflection, the question is: How does one's desire for unity express itself in acts of meta-reflection about one's self?[45] Firstly, persons strive for coherence of their autobiographical narratives. After all, by conceiving of their lives as a narrative and of themselves as the protagonist of that narrative, persons want the protagonist of their self-narrative to be well defined.[46] If a person, for example, conceives of her self as having undergone radical changes, she will strive to interpret those changes as developments of herself, or she will – to be able to conceive of her life as a narrative, in the end – try to explain those changes by referring to certain internal or external causes like illness, age, divorce, moving to a country with another culture, or traumatic experiences. A person may also conceive of herself as alternating between different 'selves' – or rather, personalities. And that person, too, will strive to reconcile those diverging notions of her self by searching for a self-narrative that explains the fragmentation of her self.[47] Let me, by way of

---

[45] This is *not* to claim that all acts of meta-reflection about ourselves qua self are (also) expressions of our desire for unity of the self. The woman, for example, who feels sad, because her relations with men again and again end in fiasco, and who seriously asks herself 'What is it about me?' engages in an act of meta-reflection about herself which does not – or rather, not necessarily – show her desire for unity of her self.

[46] By the way, there may be persons without a narrative self-conception; think of young children who do not have developed the property of having a narrative self-conception yet, or not to the full extent. But although the property of having a narrative self-conception is not fundamental for being a person, it is characteristic of human persons to have that property. This is not to deny that the property in question may be developed to different degrees and that persons therefore may be satisfied with different degrees of coherence of their self-narratives. See also Chapter 2 (The relation of oneself to one's self).

[47] I will, in this chapter, not discuss pathological cases of 'multiple selves' or 'divided minds', known as Multiple Personality Disorder or Dissociative Identity Disorder. It is questionable whether those alters can count as selves. For a discussion of pathologically divided minds see, for example, Radden 1999. See also Chapter 6 (Self-conception and the limits to dissociation).

illustration, cite Kay Redfield Jamison, a psychiatrist who describes her life with manic-depressive illness.

> Which of the me's is me? The wild, impulsive, chaotic, energetic, and crazy one? Or the shy, withdrawn, desperate, suicidal, doomed, and tired one? [...]
> And, as with a suicide attempt, living with the knowledge that one has been violent forces a difficult reconciliation of totally divergent notions of oneself.[48]

Reflecting upon her self as to how she experiences and interprets her self at different points in time, the author of these utterances is confronted with two completely different characterizations of her self and thus with a drastic fragmentation of her self; and she clearly struggles to reconcile those two totally divergent notions of her self. In my view, this struggle exhibits the author's desire for unity of her self, and in particular her desire for coherence of her personality by way of being the well-defined protagonist of her autobiographical narrative.

Besides striving for coherence of themselves qua self, persons, secondly, strive for integrity of themselves qua self, that is to say, persons want to be loyal to their values, they want to be able to justifiably conceive of themselves as the kind of person who cares about particular values.[49] A person who (frequently) violated her values may suffer from having sacrificed her values, and she may, triggered by her deep feeling of remorse, engage in meta-reflection about herself, and may want to change herself, after all; she still wants to be able to conceive of herself as the kind of person who cares about the values in question.[50]

---

[48] See Jamison 1997, pp. 68 and 120.

[49] By the way, there may be persons without the property of applying self-regarding qualitative standards to themselves. But although the property of applying self-regarding qualitative standards to oneself is not fundamental for being a person, it is characteristic of human persons to have that property. This is not to deny that the property in question may be developed to different degrees and that persons therefore may be satisfied with different degrees of integrity. See also Chapter 2 (The relation of oneself to one's self).

[50] Let me, at this point, briefly address an objection which some readers might be tempted to raise, namely that I defend a private or egocentric perspective concerning values and ethical action. Well, as this book is devoted to a person's relation to herself qua self, I obviously am concerned with a person's *own* reflexive feelings and thougths, in other words, with her *first-personal* perspective on her actions and herself. Regarding values, I am therefore focussing on the impact of feelings and thoughts of wrongness of one's actions on a person's thoughts and feelings concerning herself qua self. That focus, however, does, firstly, not imply that I, in principle, dismiss the possibility of a third-personal view on the rightness or wrongness of people's actions. It, secondly, neither implies that a person's values are individual character quirks, so to speak – although it does not exclude the possibility that an individual character quirk functions as a value in a person's evaluative reasoning. And, finally, my focus does not imply that I entertain the view that a person regards her action as wrong *primarily* or even *solely* because of what that wrong action means to her self-conception. Nevertheless, a person cannot regard her action as wrong *without* entertaining *self-regarding* thoughts and/or experiencing *reflexive* emotions. In sum, I take the objection as a misinterpretation of my account.

Let me again cite Kay Redfield Jamison to illustrate one's desire for integrity.

> I have, in my psychotic, seizurelike attacks – my black, agitated manias – destroyed things I cherish, pushed to the utter edge people I love, and survived to think I could never recover from the shame. […]
> After each of my violent psychotic episodes, I had to try and reconcile my notion of myself as a reasonably quiet-spoken and highly disciplined person, one at least generally sensitive to moods and feelings of others, with an engaged, utterly insane, and abusive woman who lost access to all control or reason. […]
> These discrepancies between what one is, what one is brought up to believe is the right way of behaving toward others, and what actually happens during these awful black manias, or mixed states, are absolute and disturbing beyond description […][51]

The author of these utterances is full of contrition; she obviously suffers from what she did to her loved ones, but she also suffers from herself, because she regularly behaves in a way that does not accord with her values. When engaged in meta-reflection about her self, she makes explicit that she still cares about being the kind of person who is disciplined, sensitive to moods and feelings of others, and whose actions actually show that she cares for the people she loves. She has a desire for integrity, even though she might (depending on the frequency and severity of her violent psychotic episodes) never be able to justifiably conceive of herself as actually being the kind of person that is guided by the values she cares about – or rather, which she cares about in her non-psychotic periods.

People who engage in meta-reflection about their self, because they are confronted with fragmentation of their self or with non-integrity of themselves, experience a drastic instability of their self. They experience a lack of unity of their self which deeply threatens their self-conception. And they often deeply suffer from that experience and want to change themselves. Their emotions of sadness and dissatisfaction about themselves – especially when confronted with a fragmented self – and of shame, remorse, or contrition – especially when confronted with non-integrity of themselves – are expressions of their desire for unity of themselves qua self. And their struggle to reconcile their incompatible notions of themselves or to re-establish their values in their life is a striving for that unity.

The inclination to strive for unity of our self is inherent to our conceiving of ourselves as the protagonist of our developing autobiographical narrative and to our conceiving of ourselves as being the kind of person who cares about particular values. Both self-conceptions are characteristic of human persons. And although both self-conceptions are acts of reflective self-awareness which presuppose awareness of instability and changeability of the self, they are, at the same time, compatible with our conceiving of ourselves as 'something'

---

[51] See Jamison 1997, pp. 120-121.

constant, that is, as the self that has been and continues to be, because both self-conceptions entail an inclination to strive for unity of our self. By entailing the desire for unity of our self, they are thus compatible with the most basic conception that we have of ourselves, namely the conception of ourselves as the self that has been and continues to be, which in turn is based on our pre-reflective sense of continued selfhood. No wonder that a person who does not succeed in reconciling the different notions of herself or who even experiences herself as different personalities ends up in a situation of mental crises.

To summarize: Firstly, meta-reflection about one's self provides a model of one's self that is based on an explicit, conceptual and objectifying self-thematization. Secondly, people, apart from exceptions, engage in reflexive meta-reflection when they are triggered by certain experiences to reflect upon their own life, when they are forced by the social group they belong to, and/or are confronted with certain reflexive emotions like regret, remorse, contrition, feeling guilty, sadness about themselves, or dissatisfaction with themselves. Thirdly, people who engage in meta-reflection about their self, because they are confronted with fragmentation of their self or with non-integrity of themselves, experience a lack of unity of their self which deeply threatens their self-conception. And their struggle to reconcile their incompatible notions of themselves or to re-establish their values in their life is a striving for that unity. Finally, one's conception of oneself as the protagonist of one's developing autobiographical narrative and one's conception of oneself as being the kind of person who cares about particular values, although both self-conceptions are acts of reflective self-awareness which presuppose awareness of instability and changeability of the self, entail the desire for unity of the self. They are therefore compatible with the most basic conception that we have of ourselves, namely our conception of ourselves as the self that has been and continues to be, which in turn is based on our pre-reflective sense of continued selfhood.

8 Conclusion

Personhood requires reflective self-awareness, thereby presupposing a pre-reflective sense of continued selfhood. The most basic act of reflective self-awareness is our conceiving of ourselves as ourselves. Reflective self-awareness does not necessarily require conceptualization and symbolization of oneself qua self. Our conceiving of ourselves as a rational agent, for example, becomes evident in acts of self-reflection that are embedded in the context of a concrete action without the person taking a detached stance towards herself qua self.

When engaged in self-reflection – whether embedded, explicitly, or in form of meta-reflection – we often experience a tension between our conceiving of ourselves as one and the same self over time and our reflective awareness of the changeability or instability of our self. I argued that we have the inclination to strive to relieve that tension by desiring unity of our self, and that the desire for unity of our self is entailed by specific self-conceptions that are characteristic

of us: firstly, our conceiving of ourselves as a rational agent; secondly, our conceiving of ourselves as the protagonist of our autobiographical narrative; and, thirdly, our conceiving of ourselves as the kind of person that cares about particular values.

The fact that persons desire unity of their self becomes evident in the following acts of reflective self-awareness: (1) conceiving of one's action as irrational; (2) feeling the need to make changes of one's personality intelligible to oneself; (3) conceiving of one's action as wrong; (4) experiencing fragmentation of one's self and feeling the need for reconciliation; and (5) suffering from non-integrity of oneself. More specifically:

(1) conceiving of one's action as irrational – when engaged in embedded or explicit weak evaluation of one's action – expresses one's desire for soundness of one's will;
(2) feeling the need to make changes of one's personality intelligible to oneself – when engaged in explaining one's actions to others or oneself (presupposing that the person has the property of a narrative self-conception) – expresses one's desire for coherence of one's self-narrative;
(3) conceiving of one's action as wrong, typically informed and accompanied by emotions of regret or remorse – when engaged in embedded or explicit strong evaluation of (one's motive for) one's action (presupposing that the person has the property of applying qualitative standards to herself) – expresses one's desire for integrity of one's self;
(4) experiencing fragmentation of one's self and feeling the need for reconciliation, often accompanied by emotions like sadness about oneself or dissatisfaction with oneself, and trying to re-interpret one's self and, probably, to change oneself – when engaged in meta-reflection about one's self (presupposing that the person has the property of a narrative self-conception) – expresses one's desire for coherence of one's self-narrative;
(5) suffering from non-integrity of oneself (that is, experiencing an impossibility to justifiably conceive of oneself as actually being the kind of person that is guided by the particular values one cares about), normally accompanied by feelings of remorse or contrition, and trying to change oneself – when engaged in meta-reflection about one's self (presupposing that the person has the property of applying qualitative standards to herself) – expresses one's desire for integrity of one's self.

We would not experience divergence in our will, fragmentation of our self, or non-integrity of our self if we did not, at the same time, conceive of our self as one and the same constant 'thing' over time. And we, presumably, would not so typically suffer from those experiences if our conceiving of ourselves as the self that has been and continues to be were not the most basic conception that we have of ourselves, which in turn is based on our pre-reflective sense of identity over time. That is why I believe that our pre-reflective sense of identity over time, finally, is the basis of our desire for unity of our self. Could our desire for unity of our self be based on anything else?

Chapter 6
Self-conception and the limits to dissociation

As I am devoted to exploring our relation to our selves, the phenomenon of dissociation from aspects of oneself, for example, from particular intentions, actions, thoughts or experiences, calls for explanation. What exactly is dissociation from oneself? Can we distinguish different kinds of dissociation from oneself? Are their limits to a person's dissociation from herself; or is a complete dissociation conceivable? Answers to these questions are not of philosophical interest only. The notion of dissociation is widely used in psychology and psychiatry. It therefore seems to me that those sciences may benefit from a philosophical clarification[1] of the possibilities and impossibilities of a person's dissociation from aspects of herself.

I start from the assumption that we can distinguish between three categories of dissociative phenomena: (1) lack of first-person access – a dissociative state of mind that people can retrospectively become aware of, for example, when you suddenly realize that you don't know whether you passed a particular point on your route; (2) dissociative experiences – for example, when a certain impulse to act strikes you unexpectedly and, although you do not understand that impulse to act, you cannot rid yourself of it; and (3) dissociative mental acts – for example, when you distance yourself from a former action, since you regard yourself as misled by others. As my concern is with dissociations *of ourselves* from ourselves, that is to say, with dissociations that we perform in contrast to dissociations that happen to us, I will discuss phenomena of the first and second aforementioned categories for the benefit of eventually getting to grips with the third. Next to dissociative phenomena that are our shared experience, I will also examine pathological cases. It is, however, not my primary concern to illuminate pathological phenomena; they function like thought experiments. After all, extreme cases are often helpful in clarifying and testing our (philosophical) concepts.

I will argue that there are limits to a person's dissociation from herself and that these limits are not of a psychological but of a conceptual nature: certain

---

[1] This is not to claim that this chapter on dissociation is in that sense unique. I refer to, for example, Flanagan 1996, Radden 1999, Radden 2004, Stephens and Graham 2000, and a number of papers published in Gallagher and Marcel 1999.

dissociative mental acts are inconceivable on conceptual grounds. It would, however, be impossible to argue for that claim without a robust philosophical theory of the concept of a person in place. I will therefore perform a conceptual analysis of the notion of dissociation in the light of a philosophical account of being a person. The leading questions are: what kinds of dissociations can count as dissociations of a person from her self and what are the constraints on those dissociations? I take it to be fundamental for being a person to conceive of oneself as oneself, though this fundamental first-person self-conception does not require identification of oneself as oneself or symbolization of oneself qua self. It will turn out to be crucial for the proposed account of dissociation that first-person self-conceptions in contrast to third-person self-conceptions – self-images of being such and such or of being a certain kind of person – are not mediated by identification or symbolization.

I will begin with an exploration of the domain of dissociation and then present a philosophical account of being a person. While unfolding that account, I will successively argue which dissociations are conceptually possible and which are impossible, respectively. Finally, I will – against the background of my account of being a person – provide a concluding overview of possible dissociations and of the limits to people's dissociations.

1 The extreme case of dissociation

1.1 Taking the extreme case literally

The most extreme case of dissociation that we can think of – or rather, that we can formulate – is a person's dissociation from herself qua *self*. That description may seem to apply to patients suffering from Dissociative Identity Disorder. After all, those people seem to be able to switch from their current self to one of their alters. Whether interpreted as a person's dissociative mental act or as something that happens to her, characterizing that phenomenon as dissociation from one's *self* sounds paradoxical. Do people speak metaphorically when they talk about dissociation from one's self, or do they mean what they literally say, namely that there still *is* a self and that the self-self relation has collapsed? I take the case of dissociation from one's self as the most extreme case on the grounds that the person's self-self relation must have collapsed.[2] After all, a person who is dissociated from her self must have lost her capability to conceive of herself as herself, in

[2] There are phenomena of dissociation in the context of people's awareness of their body that strike us as extreme as well, although in another sense. G. Lynn Stephens and George Graham, for example, refer to a paper that describes the case of a patient who experiences his head as alien; see Stephens and Graham 2000, p. 122. I will here not pay attention to that kind of dissociative phenomena.

other words, she must have lost the quality of being a self-comprehending creature. But is this possible?

Let us have a look at both the conceiving and the conceived side of the dissociation relation. Firstly, with respect to the conceiving side: In order for a person to perform a dissociative mental act or to have a dissociative experience, she has to conceive of herself as herself, in that moment. And if a person conceives of herself as herself in the moment of dissociation, then her self is on the side of the conceiver not of the conceived. In other words, her (pre-)reflective awareness of *being-self* is on the subject side of the dissociation relation not on the object side, so to speak. Hence (regarding the conceiving side of the dissociation relation) the person is not dissociated from herself qua self.

Secondly, with respect to the conceived side: In order for a person to dissociate herself or to experience herself as dissociated from her self, she has to conceive of herself as having 'that self'. We do, however, not first-personally conceive of ourselves as having (or being) this or that self. Such a self-conception would be highly reflective (we might even say hyper-reflective); it would require conceptualization of having or being a particular self, which in turn means symbolization of one's self. Conceiving of oneself as oneself, by contrast, means nothing more than being reflectively aware of oneself, or, in my terminology, conceiving of oneself as *being-self*. It requires neither conceptualization of oneself qua self nor symbolization of one's self. Therefore, a person cannot dissociate herself or experience herself as dissociated from her self. Hence, it is simply impossible to make sense of the notion of dissociation from oneself qua self – at least, if we take that expression literally, namely as a person's dissociation from her fundamental first-person self-conception of herself as herself.

This argument against the possibility of a person's dissociation from her self obviously presupposes a certain concept of a person and the self. Although that concept has not been presented yet[3], I decided not to delay my rejection of the extreme case for two reasons: firstly, the idea that such an extreme case might be possible may keep us busy – if only as a background thought; secondly, my denial of this extreme case sheds light on the theory that will emerge.

1.2 Some non-literal interpretations of the extreme case

Since I first linked the notion of dissociation from oneself qua self with patients that suffer from Dissociative Identity Disorder (DID) and then

[3] That is to say, in this chapter. It's true, I unfolded my account of a person and the self in earlier chapters, but I remind the reader that this book is also suitable for *selective* reading. The reader who has read *all* foregoing chapters will, I promise, discover new implications of my account of personhood.

argued that that notion is paradoxical and meaningless, I want to emphasize that I do not want to downplay the seriousness of DID. More specifically, I do not deny that there is something it is like for DID patients to be multiples or to have alters. The pathological phenomenon of people who seem to experience themselves as multiple by far exceeds the normal situation of persons conceiving of themselves as playing various roles. Let me briefly speculate about some alternative interpretations of that phenomenon, thereby introducing some of the notions and distinctions that we will need in a convincing account of 'dissociation from oneself'.

Firstly (a very speculative proposal), there might be human bodies inhabited by multiple selves. Think of streams of consciousness that are not integrated, thereby presupposing that acts of conscious (self-)awareness are relative to one of those separate streams. Sydney Shoemaker's remark regarding a thought experiment discussed by Parfit may illustrate that option:

> Suppose that Parfit's description of his case is right, i.e., that both streams of consciousness belong to one person, and suppose that a particular experience, say a pain, occurs in one stream but not in the other. Was that experience conscious? One might suppose that this translates into the question: "Was the person conscious of it?" But the answer to that question seems to be: "Yes and no. He was conscious of it in the right hemisphere, but not conscious of it in the left hemisphere." This suggests that the property of *being conscious* is really a relational one, and that our example should be described as one in which the experience is conscious relative to one stream and not relative to the other.[4]

If this option were real, that is to say, if people with completely separate streams of consciousness existed, then DID patients might be such people. These people would lack first-person access to (and thus first-person knowledge of) everything that occurs in their non-current streams of consciousness, while still conceiving of themselves as themselves.

In a second interpretation of the phenomenon of multiples, DID patients might 'just' have ego-alien experiences and attribute them to their alters. Having ego-alien experiences means that a person experiences certain of her actions, thoughts, intentions, experiences, or perceptions as ego-alien or emanating from an alien source.[5] The attribution of these experiences to an alter must be based on the delusion of having (particular) alters.

Proponents of a third interpretation suggest that multiples have a dissociative capacity, which has developed in childhood due to emotional, physical and sexual abuse, to avoid experiencing a certain episode.

---

[4] Shoemaker 1996, p. 191.

[5] This kind of dissociative experiences is also called 'disownership' experiences; see, for example, Radden 1999, p. 350.

> To put it roughly, through dissociation the subject is able to avoid experiencing or dealing with an intolerable episode by turning it over to an alternate personality (or *alter*) who undergoes those experiences in his place.[6]

If this interpretation is correct and multiples indeed can bring about such dissociations, the question still is whether those dissociative mental acts can count as acts of the person qua person (that is to say, qua self-comprehending creature) or whether the dissociative acts take place on a sub-personal level. In my view, the dissociations may either be coping strategies of the human organism and just happen to the person, or the person may indirectly be able to bring about the dissociative state of mind, for example by self-hypnosis[7]. Anyway, on this interpretation, the person (qua person) lacks first-person access to 'her' perceptions, sensations, desires, beliefs, intentions and actions, during a certain period.

A fourth interpretation is characterized by scepticism regarding the proliferation of selves. Owen Flanagan writes:

> It is not at all surprising that a person with powers to do so might create an alter to fend off the pain of abuse and the recognition that someone he or she loves is an abuser. But it is obscure why there is so often proliferation beyond one alter. One hypothesis is that proliferation is a complex effect of therapeutic suggestion, involving the therapist's belief in multiplicity and his conveying to the patient the possibility of fitting the description (indeed, a savvy patient may well know that multiplicity is a genuine and increasingly popular way to be damaged or express damage).[8]

Since this chapter is not devoted to the phenomenon of Dissociative Identity Disorder but to the notion of a person's dissociation from herself qua self, I do not examine whether the above list of possible interpretations is exhaustive, whether DID is a container notion for various disorders, or whether a certain interpretation may be rejected by psychologists, psychiatrists, or neuroscientists. What I claim is this: DID patients may lack first-person access to certain perceptions, sensations, intentions or actions, they may have ego-alien experiences, they may have the self-image of being a multiple, they may have developed the ability to indirectly bring about a dissociative state of mind, for example by self-hypnosis, but the phenomenon of DID cannot be described as a person's dissociation from her self – at least, if we do not speak metaphorically but mean what we literally say.

---

[6] Owen Flanagan citing the philosopher Stephen E. Braude, see Flanagan 1996, p. 75 (italics in original).

[7] Owen Flanagan mentions the possibility that DID depends, among other things, on the individual's "powerful abilities to self-hypnotize"; see Flanagan 1996, p. 76.

[8] Flanagan 1996, pp. 75-76.

Dissociation of oneself from oneself qua self (or being dissociated from oneself qua self) is not the most extreme case of dissociation but a conceptually impossible one. This is, however, not to say that the notion of alters is abandoned from our further investigation of the domain of dissociation.

2 A wide notion of dissociation

2.1 Relevant distinctions

The notion of dissociation is ambiguous. Its meaning ranges from, firstly, distancing yourself from someone or something, via, secondly, temporary disintegration of your thoughts and activities, to, finally, pathological situations of thought insertion and Dissociative Identity Disorder – to name just some of the contexts where we apply the term dissociation. In the next three subsections, I will investigate the domain of dissociation in the aforementioned broad sense of the notion.

My primary structuring criterion is the active-passive distinction; more specifically, I distinguish people's dissociative acts from dissociations that happen to them. Unsurprisingly, the first category consists of phenomena of evaluative dissociations. Later on, in the light of a robust philosophical account of self-conception, we will be able to further subdivide that category. The second category, that is, dissociations that happen to us, will be subdivided into phenomena of dissociative experiences and of lack of first-person access; the former are experienced at the time of the dissociative state whereas the latter are not.

Phenomena of lack of first-person access are further subdivided into, firstly, a temporary lack of first-person access to certain perceptions, sensations, actions, and memories; secondly, a more or less permanent lack of first-person access to certain desires, beliefs, and intentions; and, thirdly, in case of multiples, lack of first-person access to one's alters. Phenomena of dissociative experiences are subdivided into: firstly, experiences of non-rational agency, that is, knowing what you are doing but not understanding why you are doing what you are doing; secondly, non-agency experiences, that is, experiencing yourself as the subject of your thoughts, feelings, impulses for actions, or actions, without experiencing yourself as their source; and, thirdly, experiencing oneself as a multiple. We will notice that various non-pathological forms of dissociative experiences and lack of first-person access are not extraordinary, but seem to be a normal human condition.

2.2 Dissociations that happen to us: lack of first-person access

Most people will recognize the following scenario from their own experience: Driving on autopilot, you suddenly recognize that you do not know whether you passed a particular point on the route. You then come to realize that you must have had certain perceptions and must have performed adequate actions in order to manoeuvre your car through the traffic. But you were not reflectively aware of them; you *lacked first-person access* to those perceptions and actions. We might say that you were in a dissociative state of mind, that is to say, the set of your psychological states – for example, your perceptions of the environment and your conscious experiences – was temporarily disintegrated.

People are frequently and regularly in situations where they lack first-person access to certain perceptions, sensations, actions, and memories. Just think of the enormous number of routine actions that we perform daily: we act purposefully, react adequately to perceptions and sensations, and adjust the chain of activities if necessary, but we are not reflectively aware of what we experience and which activities we perform. Hence, lack of first-person access seems to be a normal human condition.

The phenomenon of lack of first-person access has a second variant. We do not only *temporarily* lack first-person access to certain of our psychological states (like in the auto-pilot case), we also seem to lack first-person access to certain desires, beliefs, and intentions, *more or less permanently*. A person may, for instance, sincerely deny that she holds a particular set of beliefs or has certain desires, although these beliefs and desires – in the view of her observers – seem to manifest themselves in her behaviour. Presumably this happens in cases of repressed beliefs or desires.[9] Although the person in question may indeed lack first-person access to those beliefs and desires, one's repressed beliefs and desires are (by their nature), in principle, candidates for first-person self-knowledge. Other mental states, by contrast, may be subconscious in ways described by Freud and by their very nature not possible subjects of first-person self-knowledge. And if, as Freud claimed, such subconscious mental states are a characteristic part (the 'Es') of human consciousness, then a dissociated state of mind regarding some desires, beliefs, and intentions is characteristic of human consciousness.

In a third variant – the pathological case of Dissociative Identity Disorder – lack of first-person access seems to be a *switching condition* regarding various selves of those people. If we take the speculative interpretation that those people actually have different selves for granted, we may characterize this phenomenon of lack of first-person access in the following way: The person is

[9] One might label such phenomena self-deceptions. What is important in the context of this discussion of 'dissociations that happen to us', however, is the fact that the person in question indeed *lacks first-person access* to those beliefs or desires, and this situation might not obtain for all kinds of self-deceptions (see also notes 23 and 31 in this chapter).

reflectively aware of, for example, her intentions, perceptions, and actions (during their actual occurrence and) when the same self is the dominant self, but she is not aware of those intentions, perceptions, and actions in periods in which another self is dominant.

This (third) variant of lack of first-person access is similar to the second in the sense that other selves of a person and those selves' desires, beliefs, intentions, perceptions, sensations, and memories might be called subconscious with respect to the person's current self-consciousness. Nevertheless, the third variant differs from the second: Depending on which self of a person is in charge, desires, beliefs, intentions, perceptions, sensations, and memories seem to switch from sub-consciousness to consciousness, so to speak. The third variant also differs from the first (that is, temporary lack of first-person access to certain perceptions, sensations, actions, and memories). After all, although a person who is a multiple might come to realize – like the person driving on autopilot – that she qua human being must have had certain perceptions and must have performed certain actions, for example, in order to have arrived at her current location, she nevertheless does not believe that *she* must have had those perceptions and must have performed those actions.

The phenomena of lack of first-person access might give rise to the following question: Can people by their own mental act *bring about* a state of mind in which they are not conscious of certain of their perceptions, sensations, actions, memories, desires, beliefs, and intentions? We might grant that possibility. Don't people bring about such dissociative states of mind by their own decision, for example by taking drugs? Well, perhaps indirectly – presuming that drugs can indeed cause dissociative states of mind and that people sometimes take drugs by their own decision to attain a dissociative state of mind. But those people still do not bring about the dissociative state of mind by their own mental act; they merely take a drug – hoping that it will *cause* a dissociative state of mind.[10] Even if some people were able to bring about a dissociative state of mind by their own mental act, for example by applying meditation techniques, those mental acts, as it will emerge in due course, cannot count as dissociative acts from aspects of themselves.

2.3 Dissociations that happen to us: dissociative experiences

Let us now turn to a second category of dissociations, which I label *dissociative experiences*. Think, for example, of yourself going to the kitchen and coming to realize that you do not know why you are doing what you are doing. In that situation, you experience yourself as the subject of your activity

---

[10] This is not to deny that people might be held responsible for the consequences of taking that drug.

of going to the kitchen, but you do not know the reason for your going to the kitchen. Therefore, you do not experience yourself as a rational agent, in that situation.[11] After all, although you know what you are doing, you do not understand what you are doing. This experience is a dissociative experience. You experience yourself qua rational agent, who therefore has the disposition to conceive of her actions as actions based on reasons, as dissociated from your own activity. This kind of dissociative experience may best be described as an *experience of non-rational agency*. Like the recognition of lack of first-person access to certain perceptions, sensations, and activities, experiences of non-rational agency are not extraordinary but our shared experience.

A non-rational agency experience is similar to the phenomenon of temporarily lacking first-person access to one's actions in the sense that, in both cases, the dissociation is not a conscious mental activity of the person but something that happens to her. However, the former differs from the latter because a person who has a non-rational agency experience is conscious of her action but does (temporarily) not remember the purpose of that action, whereas in the 'autopilot case' the person is not conscious of her actions, but she knows their purpose, that is to say, she knows that she acted in pursuit of a particular goal – for example, driving to the station.

Next to experiences of non-rational agency, we can distinguish a second kind of dissociative experiences, which are sometimes referred to as 'disownership' experiences[12], but which I – for reasons that shortly will become clear – label *non-agent experiences*. To start with, there are pathological cases. People may, for instance, experience certain of their thoughts as alien or as emanating from an alien source (this phenomenon is known as 'thought insertion' and is associated with schizophrenia).[13] Similar phenomena are experiences of alien feelings or of 'made' volitional acts. Jennifer Radden, for example, describes a patient's experience of 'dissociated will':

> The patient feels that her actions result from a consciousness and will other than her own. Not merely does she experience an irresistible and unwanted impulse whose origin feels to be external, however. In addition, the impulses in and movements carried out by her body feel to be another's. Rather than being active in bringing them about, she experiences her body as the other's passive instrument.[14]

Although patients who suffer from 'made' actions, 'made' impulses to act,

---

[11] Obviously, there are diverging philosophical accounts of what kinds of behavioural structures count as rational actions. As I take it, a person conceives of what she does (or did) as a rational action if she understands what she does (or did), that is to say, if she knows the reason for her action. I elaborated on my account of rational agency in Chapter 2 (The relation of oneself to one's self). See also this chapter, Section 3.5 (Conceiving of oneself as a rational agent).

[12] See Radden 1999, p. 350.

[13] See Radden 1999, p. 350.

[14] Radden 1999, p. 351.

inserted feelings, or inserted thoughts experience those actions, impulses to act, feelings, or thoughts as ego-alien, they nonetheless experience them as theirs – at least, in a certain restricted sense: these thoughts or impulses to act are still experienced as happening to *them* but they lack the quality of 'mineness' (as, for example, Jennifer Radden[15], and G. Lynn Stephens and George Graham[16] convincingly argue). Thus, in the sense of ownership of the experience, the patients ascribe the thoughts and impulses to themselves, while in the sense of ownership of the mental action, they ascribe them to other sources.[17] I call this kind of dissociative experiences *non-agent experiences*, since the person does not experience herself as the source of her thoughts, feelings, impulses for actions, or actions, while still experiencing herself as their subject.

Not all non-agent experiences are pathological, however. There seem to be normal (that is to say, non-pathological) non-agent experiences too. Harry Frankfurt describes an example:

> [T]here are obsessional thoughts, whose provenances may be obscure and of which we cannot rid ourselves; thoughts that strike us unexpectedly out of the blue; and thoughts that run willy-nilly through our heads.[18]

Such experiences can indeed count as non-agent experiences, since those thoughts seem to happen to us instead of, as Harry Frankfurt points out, being something *we think*.[19] But it seems to me that they are experienced as less alien than the pathological cases in which a person, for instance, experiences thoughts "as of their having been placed in the patient's mind from some outside source"[20].

Our brief discussion of experiencing ego-alien experiences could lead to the view that multiples 'simply' have pathological non-agent experiences.[21] The only difference would be that patients with Dissociative Identity Disorder ascribe the experiences in question to a particular alter of themselves, instead of ascribing them to an alien source. There is, however, an important

---

[15] See Radden 1999, especially p. 351. Radden in turn refers to writings of G. Lynn Stephens and George Graham, who make a subject-agent distinction in a person's self-ascription of a mental episode; see Radden 1999, pp. 354-355.
[16] See Stephens and Graham 2000, for example, pp. 152-153.
[17] If this is right, then Shoemaker's view on the self-intimation of, for example, our beliefs or desires (see Shoemaker 1990, pp. 50-52) needs refinement. The self-intimation of beliefs or desires would turn out to be a partial self-intimation, so to speak, namely, 'merely' concerning one's own experience of the belief or desire, not concerning that belief or desire being one's own mental act.
[18] Frankfurt 1976, p. 59.
[19] See Frankfurt 1976, p. 59.
[20] Radden 1999, p. 350.
[21] I mentioned this view as one of the possible interpretations of DID in Section 1.2 (Some non-literal interpretations of the extreme case).

distinction between those two kinds of dissociative experiences. The patients suffering from, for example, ego-alien impulses to act or ego-alien actions still conceive of themselves as the subjects of those ego-alien experiences. Multiples, by contrast, do not conceive of themselves as the subject of the thoughts or actions of their alters – otherwise they would not be multiples.

I must admit that I believe that a multiple's conception of herself as being a multiple is based on, for example, ego-alien experiences or recognitions of lack of first-person access to certain episodes, accompanied by the belief that she is a multiple, rather than on (dissociative) experiences of being a multiple.[22] In contrast to the other two kinds of dissociative experiences, I just cannot think of any situation whatsoever where something like experiencing oneself as a multiple occurs to people as a normal dissociative experience. Therefore I cannot – by extrapolation from those normal dissociative experiences – imagine how it might be like for multiples to be different selves. This is an important difference with other peculiar cases of dissociated minds. After all, we know from 'normal' experiences, firstly, what it is like to have a dissociative experience of non-rational agency (that is, knowing what one is doing, but not knowing why); therefore we have a glimpse of what it might be like to have, for example, Alzheimer's disease. We know, secondly, what it is like to have a dissociative experience of non-agency (that is, experiencing oneself as the subject of, for example, certain thoughts but not as their source); and therefore we have a glimpse of what it might be like, for instance, to experience thoughts as inserted or to experience an irresistible impulse whose origin feels to be external. But there is, to my knowledge, no non-pathological variant of the dissociative experience of being different selves – if what it is like for DID patients to be different selves can count as an experience at all.

2.4 Dissociative mental acts of a person qua person: evaluative dissociations

Although the question may sound odd at first sight, we might ask whether people by their own mental act can bring about a state of mind in which they have dissociative experiences, instead of the dissociated state of mind just happening to them. In everyday language, we say about people that they

[22] An objector might try to brush aside my speculations on the experience of being a multiple by doubting whether a multiple ever experiences herself as a multiple. The objector might suggest that multiples just experience one alter at a time. I share the objector's doubt. Her suggestion, however, that multiples experience one alter at a time describes – in the light of my account of self-conceptions – a conceptually impossible situation. I have not yet – at least, in this chapter – presented the conceptual apparatus to discuss that suggestion, but here is a brief argumentation: Persons, first-personally, don't experience this or that self; they just experience (and conceive of) themselves as *being-self* – the self that has been and continues to be. I refer to Section 3 (Self-conception), especially the discussion on multiples in Section 3.4 (Dissociation and pre-reflective self-awareness).

dissociate themselves from someone or something if they say or show that they are not connected with the person or thing in question (usually in order to avoid trouble or blame). Here the term 'dissociation' refers to a person's explicit dissociative reflective act.

All dissociative reflective acts that are reflexive by nature are relevant to our analysis; more specifically, all kinds of reflective dissociations from 'aspects' of oneself, for example, from particular desires, intentions, beliefs, actions, decisions, adopted goals and plans, or things one cares about, are important. Think of a person who tries to resist her desire for chocolate because of regarding that desire as resulting from an addiction – she distances herself from her desire to eat chocolate.[23] Or think of a person who does not feel responsible for a former action because of regarding herself as the victim of a deception and her action as based on that deception; this person distances herself from the action in question. I label this kind of dissociation *evaluative dissociation*. The evaluative act may be retrospective or, as it is in the first example, embedded in the context of deliberation about one's current or a future action.

Evaluative dissociation clearly is a dissociative act on the level of the person. It requires reflection on one's actions or desires. And to reflect on one's actions or desires in turn requires having the capacity to conceive of oneself as oneself, and of oneself as the subject of those actions and desires. This capacity is the fundamental feature of personhood. Therefore a mental act that requires that capacity is a mental act on the level of the person. It remains to be seen which 'aspects' of a person might be objects of her evaluative dissociation; after all, we are concerned with the limits to a person's dissociation from herself.

3 Self-conception

Without a philosophical theory of the concept of a person in place, it is impossible to propose a defensible set of answers to the questions that arose during our analysis of the notion of dissociation.[24] How could we ever decide whether certain dissociative phenomena may count as dissociative mental acts of a person qua person, in other words, as dissociative mental acts on a personal level, if we did not presuppose a concept of a person? We need a concept of a person in order to find out which aspects of ourselves we can dissociate from and which aspects are such that we cannot dissociate from them. On the other hand, attempts to provide an intelligible and coherent set of answers to questions concerning the limits to a person's dissociation from

[23] It is arguable that certain acts of evaluative dissociations – for example, distancing oneself from one's desire for chocolate because of regarding that desire as resulting from an addiction – can count as self-deceptions. However, that qualification is not relevant for the phenomenon to count as an act of evaluative dissociation.

[24] I must admit that my view of the concept of a person (that is, persons count as persons owing to their capacity to conceive of themselves as themselves) has already crept into my description of the various phenomena of dissociation.

herself may enhance our understanding of the concept of a person and its implications, and may help to fine-tune the concept.

3.1 First-person and third-person self-conception

I take a person to be a creature that conceives of herself as the subject of her current perceptions, sensations, desires, beliefs, intentions, actions, thoughts, and memories, and that conceives of those perceptions, intentions, actions, thoughts, etcetera as *her* perceptions, intentions, actions, and thoughts. A person ascribes her current perceptions, intentions, actions, thoughts, etcetera to herself based on her *first-person access* to those perceptions, intentions, actions, and thoughts. It is characteristic of first-person access to, for example, our perceptions or thoughts, that we have direct knowledge of those perceptions and thoughts; that is to say, we do not come to know that we have certain perceptions or thoughts by observing ourselves having those perceptions or thoughts, and by identifying ourselves as ourselves. Self-ascriptions that are based on first-person access do not require mediation by observation of oneself (or introspection of one's psychological states) and identification of oneself as oneself.[25] Ernst Tugendhat calls our knowledge that these predicates apply to ourselves "unmediated epistemic self-consciousness" ("unmittelbares epistemisches Selbstbewußtsein"[26]). He characterizes these kinds of self-ascriptions in the following way: we ascribe certain predicates to ourselves "not only by actually

[25] I thus reject a perceptual model of such self-ascriptions. In a perceptual model, a person comes to know that she has a certain thought or desire by (a kind of) perception of that thought or desire. The problem with the perceptual model of self-conceptions is that one's first-person self-ascriptions – for instance, one's conception of one's thought as one's thought – and, in the end, one's fundamental self-conception of oneself as oneself turn out to be susceptible to the error of misidentification. After all, how could we explain – if the conceived 'thing' is an object of the conceiver's perception – that a person indeed relates to herself as herself when she conceives of her thought as her thought? How could we know that the object = the subject? We might misidentify ourselves. By contrast, according to the presented account of first-person self-conceptions, which rejects a perceptual model, we do not identify ourselves at all. For detailed argumentations against the perceptual model – also called "Reflexionsmodell" (Tugendhat's and Frank's term) – or, more generally, the subject-object model of self-consciousness, I refer to Shoemaker 1986, especially pp. 11-17, Tugendhat 1979, for example, pp. 69-62, Frank 2002, for example, pp. 130-133, and Zahavi and Parnas 1999, especially pp. 261-263. For a denial of the possibility of misidentification, see, for example, Baker 2000, pp. 70-71 and 136-137, Shoemaker 1986, pp. 11-12, Tugendhat 1979, especially p. 83, or Zahavi and Parnas 1999, p. 262.

[26] Tugendhat 1979, p. 27: "Genauer müssen wir vom *unmittelbaren* epistemischen Selbstbewußtsein sprechen, in Abgrenzung gegen einen weiteren Begriff von epistemischem Selbstbewußtsein, der alles Wissen umfaßt, das sich in Sätzen der Form "ich weiß, daß ich - - -" artikuliert. Denn es gibt natürlich Sätze dieser Form, in denen das Prädikat nicht für einen Zustand steht, der bewußt ist, für einen Zustand, von dem der betreffende ein unmittelbares Wissen hat. Solche Sätze können sowohl meine Person als köperliche betreffen […] als auch meinen Charakter und meine Verhaltensweisen […]." (Italics in original)

saying 'I' but from the perspective of saying 'I'".[27] Inspired by Tugendhat's explications[28], I call self-conceptions that are based on first-person access *first-person self-conceptions*.

Ascribing certain intentions or thoughts to oneself does not only mean that that person conceives of herself as having those intentions or thoughts but also that the person conceives of herself as herself. First-person self-ascriptions like '*I* am terrified' ('*I*', in the sense of 'I know'), '*I* hear a noise,' '*I* am just thinking about this sentence', or '*I* am just drinking water' are evidences of the person's (reflective) capability to conceive of herself as herself. In other words, conceiving of oneself as oneself is implied by the more specific first-person self-conceptions. If a person did not conceive of herself as herself, she could not ascribe to herself anything whatsoever.

Conceiving of oneself as oneself is what Lynne Rudder Baker calls a highly developed form of a first-person perspective; it is the "defining characteristic of all persons"[29]. The capacity to conceive of oneself as oneself is essential to being a person, because the persistence of a person as a person depends on that capacity. This means, for example, that a person who does not conceive of herself as herself anymore in fact has ceased to exist qua person.[30] I call the conception of ourselves as ourselves *the fundamental first-person self-conception*.

Although conceiving of oneself as oneself clearly is an act of *reflective* self-awareness, this kind of reflective self-awareness (like the specific first-person self-ascriptions of, for example, having a certain thought) does *not* imply that the person, in that reflective act, takes a *detached* stance towards herself, identifies herself as herself, or conceptualizes and symbolizes herself qua self. By contrast, a person's conception of herself as a certain kind of person, for example as someone who is sensitive to the feelings of others, requires that the person takes a detached stance towards herself qua self, thereby conceptualizing and symbolizing herself qua self. That kind of self-conception requires the formation or entertainment of a certain self-image. Think, for example, of a person's conception of herself as pursuing a certain long-term goal, for instance of becoming a physician; or think of a person's conception of

---

[27] Tugendhat 1979, p. 87, my translation. The German original reads: "[Es gibt] nämlich Prädikate, die ich mir nicht nur faktisch "ich"-sagend, sondern aus der Perspektive des "ich"-Sagens zuspreche".

[28] In his book *Selbstbewußtsein und Selbstbestimmung*, Ernst Tugendhat distinguishes between unmediated epistemic self-consciousness ("unmittelbares epistemisches Selbstbewußtsein") and mediated epistemic self-consciousness ("mittelbares epistemisches Selbstbewußtsein"); see Tugendhat 1979, especially pp. 27 and 85-87. I was inspired by his explications to draw a clear distinction between self-ascriptions based on first-person and third-person access, respectively.

[29] See Baker 2000, p. 91.

[30] More specifically, ceasing to exist qua person means ceasing to exist simpliciter. Nevertheless, the body that formerly constituted the person may still be a (human) creature that has the capacity of being pre-reflectively aware of itself. That creature might be able to perform acts of self-awareness without being self-conscious in the fullest sense, that is to say, without having the capacity of being reflectively aware of itself.

herself as having a certain biography, for instance of being the victim of slander and malicious intent in crucial situations of her life.[31] Both self-conceptions require a self-image, however vague that may be. After all, in order to conceive of yourself as becoming a physician or as being betrayed by others in various situations in your life, you have to think of yourself qua self in other circumstances in the future or in the past; and that means that you perform an act of mediated reflective self-awareness – mediated by your conceptualization and symbolization of yourself qua self. Self-images also presuppose mediation by identification. And in certain cases, this mediation by identification is especially obvious; think of my self-conception of being C.S. or of my visual image of myself as a 2-year-old child. In short, self-images are not intrinsically reflexive; they are mediated self-conceptions – mediated by symbolization and identification. I label them *third-person self-conceptions*.

> It is characteristic of all these kinds of mediated self-knowledge that it is in principle available to others in the same way as to myself, probably even easier. Whether I was born in Berlin is better known by my mother than by myself. Whether I am cowardly or not is not something that I can have direct knowledge of; it can only become evident in my actions.[32]

Needless to say that in order for a creature to have third-person self-conceptions, that creature has to conceive of herself as herself.

To sum up: (1) first-person self-ascriptions are not mediated by identification, or conceptualization and symbolization; (2) first-person self-ascriptions are evidences of the person's capacity to conceive of herself as herself; and the capacity to conceive of oneself as oneself is the fundamental first-person self-conception, because one's personhood depends on that capacity; (3) third-person self-conceptions are mediated self-conceptions (self-images) – mediated by identification, or conceptualization and symbolization; (4) third-person self-conceptions presuppose the capacity to conceive of oneself as oneself.

3.2 Dissociation and first- and third-person self-conceptions

The fact that self-images are mediated by identification, or conceptualization and symbolization makes it intelligible that we can (easily) dissociate ourselves from former self-images by referring to external or internal

---

[31] Self-images may often be self-deceived. We may, for instance, exclude what conflicts with what we want to believe about ourselves; and some people may even be massively self-deceived.

[32] Tugendhat 1979, p. 27, my translation. The German original reads: "Charakteristisch für alle diese Weisen des nicht unmittelbaren Wissens von sich ist, daß solches Wissen einem Anderen in prinzipiell derselben Weise und gegebenfalls sogar leichter zugänglich ist. Ob ich in Berlin geboren bin, weiß meine Mutter besser als ich. Ob ich feige bin oder nicht, davon habe ich kein unmittelbares Wissen, das kann sich nur in meinen Handlungen zeigen."

conditions that influenced our former identification, or conceptualization and symbolization. Such dissociative acts are acts of evaluative dissociation. A person may, for example, distance herself from her former, youngish self-image of becoming a veterinary surgeon because of now regarding that self-image as influenced by films and literature. Another person may distance herself from her former self-image of being continuously betrayed by others because of now regarding that self-image as influenced by her depressive moods. And it is conceivable that I would distance myself from my former self-conception of being C.S. if I now regarded that self-conception as based on wrong information.

But could I distance myself then, in the very same moment, from *that* understanding? No, because in that case I would not conceive of myself as myself, in the current moment. My understanding of something, in the present moment, my regarding something as such and such, in that moment, my conceiving at all is inextricably bound up with my conceiving of myself as myself, in that very same moment. J. David Velleman writes:

> If there is a part of your personality with which you necessarily think about things, then it will be your mental standpoint, always presenting a reflexive aspect to your thought. You will be able to think about this part of your personality as "it," but only from a perspective in which it continues to function as the thinking "I" [...] 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 [...] [H]e cannot attain a cognitive perspective from which his understanding isn't "I".[33]

But how about dissociations from one's first-person self-conceptions? If first-person self-ascriptions of our current desires, intentions, thoughts, or actions do not require mediation by identifying oneself as oneself, or by conceptualizing and symbolizing oneself qua self, how then can we ever dissociate ourselves from certain of those desires, intentions, thoughts, or actions? Well, we can obviously *retrospectively* distance ourselves from certain of our desires, intentions, thoughts, or actions by regarding them, for example, as caused by excessive alcohol consumption. Retrospective evaluative dissociation, however, cannot count as (part of) a first-person self-conception. It is a detached reflective act of self-awareness that requires thinking of oneself in other circumstances in the past; it is hence performed in the context of one's current *third-person self-conceptions*.

But people sometimes distance themselves from, for example, certain of their *current* desires or actions, while still ascribing those desires or actions to themselves (otherwise they could not distance themselves from those desires or actions). Take the example of a person who has a strong desire for chocolate, and let us imagine and examine possible cases of dissociations

[33] Velleman 2002, p. 114 (italics in original).

from that desire. Firstly, a person, who wants herself not to be moved by her current desire for chocolate, might try to reject or suppress that desire. She, in a certain sense, distances herself from her desire for chocolate by – in a Frankfurtian sense – not identifying with that desire. But can this rejection of her desire count as an act of dissociation from her current desire? It depends. Assuming that the person experiences herself not only as the subject of her desire for chocolate but also as the source of that desire (as the one who brings that desire about, so to speak), then her rejection of her desire for chocolate cannot count as an act of dissociation. However (and that is the second case), the person might try to resist her desire for chocolate by regarding that desire as, for instance, resulting from a chocolate addiction. This person experiences herself as the subject of her desire for chocolate but not as its source. She has a dissociative experience of non-agency and may therefore distance herself from her desire for chocolate. This person's rejection of her desire for chocolate can count as a dissociative act, and it is indeed part of a first-person self-conception – after all, the person conceives of her desire for chocolate as her desire, at least in the sense of being the subject of that desire. Now consider a third case. A person has a desire for chocolate that she experiences as ego-alien, as a 'made' desire, so to speak; she experiences her body as another's passive instrument. The person might try to resist that alien impulse to eat chocolate. Does this person dissociate herself from her desire and impulse to eat chocolate, and, if so, is that dissociative act part of a first-person self-conception? The answer, although this case is pathological, does not differ from the second case. The person distances herself from her ego-alien desire for chocolate – from an ego-alien aspect of herself, so to speak – as part of her current first-person self-conception. To avoid misunderstanding: not her dissociative experience counts as a dissociative act; the dissociative act is embedded in her self-ascription (which is a reflective act of self-awareness) of having an ego-alien desire for chocolate.

This leads to the following tentative conclusions. Firstly, we can dissociate ourselves from our former third-person or first-person self-conceptions. These dissociations are phenomena of retrospective evaluative dissociation. They require a detached stance towards ourselves qua self and are performed in the context of one's current third-person self-conception. Secondly, in special cases, dissociative acts can be embedded in people's current first-person self-conceptions, namely when they, based on pathological experiences of non-agency regarding, for example, their current desire or action, dissociate themselves from that desire or action. Thirdly, we cannot dissociate ourselves from our own dissociative act, in that very same moment, for we cannot dissociate ourselves from our understanding of something, in that very same moment. After all, our understanding of something, in the present moment, our regarding something as such and such, in that moment, our conceiving at all is inextricably bound up with our conceiving of ourselves as ourselves, in that very same moment.

As we have seen, dissociative experiences of non-agency can be grounds for dissociative acts. But how about experiences of non-rational agency and of being a multiple, and phenomena of lack of first-person access – can they be grounds for people's dissociative acts too? The third question can be answered immediately. Lack of first-person access in its three variants (firstly, temporarily concerning certain perceptions, sensations, and actions, secondly, more or less permanently regarding certain subconscious desires, beliefs, and intentions, and, thirdly, as a switching condition regarding one's alters) can neither be part of one's third-person nor of one's first-person self-conceptions. After all, lacking first-person access to certain perceptions, subconscious desires, or alters means not being reflectively aware of them; and not being reflectively aware of certain 'aspects' of oneself entails the person's impossibility to dissociate herself in a mental act on the personal level (in other words, qua reflectively self-conscious being) from those 'aspects' of herself.

The first and second question – whether experiences of non-rational agency and of being a multiple, respectively, can be grounds for dissociative acts – require elaboration on two aspects of my account of personhood: firstly, our self-conception of being rational agents and, secondly, the notion of pre-reflective self-awareness. Let me start with the latter.

3.3 Pre-reflective self-awareness

The fact that people – for example, when driving their car or performing other routine tasks – may lack first-person access to certain of their perceptions, sensations, and actions does obviously not imply that those perceptions and sensations are not cognitively processed, not taken into account in one's actions, or that the actions are inadequate. It does, moreover, not imply that people, in those situations, lack their self-awareness (of being the self that has been and continues to be).

We can act purposefully and react adequately to perceptions and sensations without being reflectively aware of those perceptions, sensations and actions. However, non-reflectively – or rather, pre-reflectively – we have to be aware of them and of ourselves as their subject, otherwise we could not do things like driving on autopilot. And we are not only pre-reflectively aware of ourselves as the subject of the various perceptions, sensations and actions, as distinct unconnected events, we are pre-reflectively aware of ourselves as one and the same subject of those perceptions, sensations and actions. In short, we are pre-reflectively aware of ourselves as the self that has been and continues to be; we have what I call a pre-reflective sense of

[34] In Chapter 5 (Selfhood: Unity in changeability), I elaborated extensively on the notion of a pre-reflective sense of (continued) selfhood. In the context of this chapter, I limit myself to a brief explanation.

(continued) selfhood.[34] After all (take again the autopilot case), we perform complex series of interconnected actions in pursuit of our overall purpose of driving to a particular place; and that means, more specifically, that what one is prepared to do is related to the overall purpose of driving to a particular place and to what one is just now doing (and to certain of one's current perceptions and sensations), and what one is just now doing is related to what one has just been doing (and to certain of one's current perceptions and sensations). Hence, although we may lack first-person access to the perceptions, sensations and actions involved, we nevertheless experience ourselves (pre-reflectively) as one and the same perceiving, thinking, and acting subject, in the various acts of pre-reflective self-awareness involved – otherwise we would not be able to perform complex structures of interrelated and adequate actions like driving on autopilot.

In sum: we perform various complex routine tasks without being reflectively aware of the action structures that we perform in the pursuit of those tasks and of various perceptions that we have during their performance, and we even adapt our action structures in reaction to those perceptions without being reflectively aware of those adaptations. All this would be impossible if we were not pre-reflectively aware of ourselves as the self that has been and continues to be.

Probably, the moments and periods in which we are *reflectively* aware of ourselves, for example when deliberating about how to perform a certain task, or when reflecting on what we have done so far and whether that was the most effective and efficient way to actualize a particular goal, are more seldom than we normally think. A significant part of our purposeful behaviour does not require our reflective awareness of that behaviour, but it would be impossible without a pre-reflective sense of (continued) selfhood. This is not to deny that lots of complex action structures, which we perform without being reflectively aware of them in that moment, in principle require our reflective capabilities. For example, we could never have thought of the purposes of those action structures if we had not been able to think of ourselves in other circumstances in the future, and we could never have come to perform those actions 'automatically' if we had not first trained ourselves to perform those actions, thereby reflecting on what we are doing.

3.4 Dissociation and pre-reflective self-awareness

The fact that we pre-reflectively experience ourselves as a self – the self that has been and continues to be – makes it intelligible that we are able, for example, to drive our car while lacking first-person access to certain perceptions, which we nevertheless take into account, and to our actions, which we nevertheless adjust depending on those perceptions. Isn't it somewhat surprising that we, while performing complex action structures without being reflectively aware of them, can perform other acts of reflective

self-awareness? While driving on autopilot, we may, for instance, deliberate about what to prepare for dinner, discuss with a colleague how to convince our customer, reflect on past conversations with our parents, or evaluate our career opportunities.

Could this phenomenon help us to understand the pathological phenomenon of lack of first-person access to one's alters? Think of the following proposal: The person driving on autopilot and, at the same time, deliberating about how to convince her customer may in fact have two pre-reflective senses of selfhood in that moment, and just lack first-person access to one of them. To put the proposal in general terms: the pathological phenomenon of lack of first-person access to one's alters might be understood as a phenomenon of being dissociated from non-current senses of selfhood. Is that proposal conceivable, given my account of self-conception and pre-reflective self-awareness?

No, a sense of selfhood is not identifiable as being this or that sense of selfhood; it is content-free, so to speak, and it is inextricably bound up with the experiencing subject. A pre-reflective sense of selfhood is nothing more than experiencing oneself as the self that has been and continues to be. Even if we assumed, for the sake of argument, that people can have different streams of consciousness, each of which might establish a sense of selfhood (in a certain moment), then nonetheless a sense of selfhood is just that: pre-reflectively experiencing oneself as the self that has been and continues to be. There is no sense of selfhood without a creature experiencing itself as *being-self*. Hence, if a creature lacks access to a certain stream of consciousness, then talk about a sense of selfhood belonging to that stream of consciousness does not make any sense. We might lack first-person access to other streams of consciousness than the one which our reflective awareness is currently connected to, more specially, we might lack first-person access to experiences, for example perceptions of external objects or feelings that are processed in other streams of consciousness, but we cannot lack first-person access to other senses of selfhood. Our sense of selfhood and our fundamental first-person self-conception of ourselves as ourselves are inextricably bound up with ourselves as the experiencing and conceiving subject, respectively. They are both content-free, that is to say, nothing more than experiencing, respectively conceiving of ourselves as ourselves, the self that has been and continues to be. Neither experiencing nor conceiving of ourselves as ourselves requires any identification of oneself as oneself. Therefore, if a person, driving on autopilot, suddenly recognizes that she, for example, is not aware of having passed a particular point on the route, then she may try to find out whether she, given her current location, must have passed that point. But she will never doubt that *she* either has passed that point or not.

Could we imagine a situation when someone has reason to think that one of his alters must have performed certain actions or must have had certain experiences? Take, for instance, a person who cannot remember anything of the recent period and who cannot believe that *she* has done and experienced

what others tell her to be the case and what she has every reason to believe, given, for example, her current location and condition. Is it conceivable that this person indeed doubts that *she* had those experiences and performed those actions, and therefore thinks that one of her alters must have experienced those things and performed the actions? Well, a person may indeed doubt whether she indeed has experienced and done what others tell her to be the case, and she may therefore, for example, try to find more evidence for what happened. But wouldn't it be very strange if the person concluded that an alter of herself was in charge? What could be the basis for her belief that she is a multiple?

We can imagine different scenarios of people believing that they have an alter, or even more than one. (I already gave a sketch of those scenarios when I presented interpretations of the phenomenon of Dissociative Identity Disorder as non-literal interpretations of the extreme case.[35]) Firstly, a person may think that she has an alter if she has ego-alien experiences like 'made' thoughts, 'made' feelings, 'made' impulses to act, or 'made' actions. She may not only distance herself, as part of her current first-person self-conception, from those thoughts, feelings, etcetera, but she may also form the third-person self-conception of having an alter who brings about those thoughts, feelings, impulses to act, or actions. That person, however, despite of her self-image of being a multiple, still experiences herself as the subject of the thoughts, feelings, impulses to act or actions, although not as their source. We are, by contrast, imagining a person who did not experience herself as the subject of the perceptions and actions in question; after all, the person from our example cannot remember anything of that period.

Secondly, a person might have the self-image of being a multiple, for example because her therapist told her that her pathological symptoms are evidence of being a multiple.[36] Confronted with actions that she performed in a certain period and things that happened to her then, and given her lack of first-person access to those actions and experiences, she might ascribe those actions and experiences to an alter of herself. These self-ascriptions thus happen in a detached, reflective stance, and are mediated by a

---

[35] See Section 1.2 (Some non-literal interpretations of the extreme case).

[36] Not only certain psychologists, but also some psychiatrists and neuro-scientists hold the view that persons can have more than one self. A Dutch team of researchers, for instance, describe experiments with DID patients, which show that these patients can be in different personality states and that their access to memories of traumatic experiences depend on these personality states. They write: "Our findings reveal the existence of different regional cerebral blood flow patterns for different senses of self." (Reinders et al. 2003, p. 2119) The qualification that such a patient has different *senses of self* – since she, in certain circumstances, lacks access to trauma memory and exhibits corresponding emotional responses – is an interpretation by the research team. Even if the researchers may be rightly inclined to talk about different autobiographical selves, because the patient, in different situations, has first-person access to different memories, the patient herself, at each moment, does not experience and conceive of herself as a *particular* self but simply as herself. She may, however, adopt the view of the psychiatrists and neuro-scientists in question and entertain the corresponding self-image.

constructed model of her self – her self-image of being a multiple. That self-image, because of being a self-image, is a third-person self-conception like 'I am C.S.', 'I am kind', or 'I am an artist'. It may therefore be false or a delusion.

Thirdly, the person may experience herself as a multiple and therefore think that an alter of herself must have been in charge when 'she' performed certain actions. I must admit that it is hard for me to say anything meaningful about that case, because, as I already mentioned during our investigation of the notion of dissociation[37], I cannot – by extrapolation from normal dissociative experiences – imagine how it might be like for multiples to be multiples. It seems to me highly unlikely that the something it is like for multiples to be multiples really can count as the *experience* of being a multiple, but as far as I can see we cannot exclude that possibility on conceptual grounds.

Let us therefore for the sake of argument assume that the experience of being a multiple exists; and let us try to analyse the implications of that experience for a person's sense of selfhood. Firstly, the person in question has a pre-reflective sense of selfhood, in the moment of her experience of being a multiple. After all, a person who experiences herself as a multiple, in any case conceives of herself as herself, in the moment of her reflective awareness of the experience of being a multiple. And conceiving of herself as herself (in other words, the reflective awareness of *being-self*) requires the person to have a pre-reflective sense of selfhood, in the present moment. Secondly, the person has first-person access to her experience of being a multiple; otherwise the experience could not count as a person's current experience. In other words, the person's reflective awareness of her experience of being a multiple is direct – not mediated by identification, or conceptualization and symbolization. Finally, as a consequence of the first two points, a creature who thinks that she has an alter because of experiencing herself as a multiple still (pre-reflectively) experiences herself as one and the same thinking, perceiving, and feeling subject, in various acts of (pre-reflective) self-awareness that may appear in the same moment – including her experience of being a multiple (if such an experience exists at all).

In conclusion (and presuming that people can have different streams of consciousness), a person may lack first-person access to experiences that are processed in other streams of consciousness. Such a person may be said to be in a dissociative state of mind. But the person can never be said to lack first-person access to her other, non-current senses of selfhood, for the expression 'her other senses of selfhood' is just nonsense against the background of my theory. A person cannot have different senses of selfhood. A sense of selfhood is just that: pre-reflectively experiencing oneself as the self that has been and continues to be. Even a person who experiences herself as a multiple, that is to say, who has first-person access to the experience of being a multiple

---

[37] See Section 2.3 (Dissociations that happen to us: dissociative experiences).

(assuming that such an experience exists), still conceives of herself as herself whenever she is reflectively aware of herself. After all, that reflective self-awareness requires that the person experiences herself as one and the same perceiving, thinking, feeling, and acting subject in various acts of pre-reflective self-awareness that may appear at the same moment, including the act of awareness of being a multiple.

3.5 Conceiving of oneself as a rational agent

One phenomenon of dissociation has not been analysed yet against the background of my account of personhood: the experience of non-rational agency, that is, the experience of being conscious of one's activity but (temporarily) not knowing the purpose of that activity. The fact that this experience confuses us becomes understandable if knowing the purpose of one's actions, in a certain moment, is inextricably bound up with one's conception of oneself as a rational agent, in that very same moment, and if conceiving of oneself as a rational agent belongs to our first-person self-conception. Let me explain.

We conceive of our actions as our actions and (leaving aside ego-alien experiences) of ourselves as the source of those actions. Conceiving of certain behavioural structures as actions means understanding those behavioural structures as purposeful behaviour, more specifically, as behaviour that has a certain purpose, which we conceive of as our purpose, as our motivating reason for the action in question.[38] In other words, understanding a behavioural structure, which we perform at a certain moment, as our action means that we conceive of ourselves, in that moment, as a person who acts based on reasons. Conceiving of our actions as actions is thus tied to conceiving of those actions as based on reasons. And that is, as I take it, what conceiving of oneself as a rational agent amounts to. Conceiving of ourselves as rational agents does not imply that we explicitly think 'I am a rational agent'. Our disposition to understand our actions, which is tied to conceiving of our actions as our actions, exhibits our self-conception of being rational agents.[39] And because conceiving of one's action as one's action is a first-person self-conception, conceiving of oneself as a rational agent is inextricably bound up with that first-person self-conception.

It cannot be overemphasized that the self-conception of being a rational agent – in the sense described above – is a first-person self-conception; it does not require conceptualization and symbolization of oneself qua self.

---

[38] I elaborated on the notion of conceiving of one's actions as actions based on reasons in Chapter 2 (The relation of oneself to one's self) and in Chapter 3 (Practical irrationality from a self-related perspective).

[39] I discussed 'One's self-conception as a rational agent' in more detail in Chapter 5 (Selfhood: Unity in changeability).

However, people can also entertain a third-person self-conception of being a rational agent. Someone may, for example, have the self-image of being a person who is not moved by unevaluated spontaneous desires, who rejects spontaneous desires for the benefit of long-term goals, and who has coherent short- en long-term goals. That person can obviously distance herself from that kind of self-conception, in an act of retrospective evaluative dissociation (or as part of her evaluative reasoning concerning her current action) and, for example, decide to let herself slide from moment to moment, so to speak, doing whatever she feels like in the present moment. But this clearly does not imply that the person is no longer disposed to understand her actions as based on reasons. Even if she does whatever she feels like, she will still conceive of that action as her action; and by this she is disposed to understand her action as based on reasons, since that disposition is tied to her first-person self-ascription of her action.

We feel confused if we cannot understand our actions as based on reasons, that is to say, if we do not know why we are doing what we are doing. In such a situation, a person, on the one hand, knows that she is doing what she is doing and is therefore disposed to understand that action structure, on the other hand, she does not know the purpose of what she is doing and can therefore not conceive of the behavioural structure in question as an action based on reasons. This person has an experience of non-rational agency. I label that experience a dissociative experience, for you experience yourself qua agent – or rather, qua rational agent, because a creature that conceives of itself as itself is disposed to understand its actions as actions based on reasons – as dissociated from your own action.

Could we avoid feeling confused when we do not know why we are doing what we are doing? If I am right that for creatures like us (that is, creatures that conceive of themselves as themselves) experiencing one's behaviour as an action structure entails the disposition to understand that action structure, then ceasing to feel the confusion would imply that we could cease conceiving of ourselves as (rational) agents. Could we perhaps bring about a state of mind, by a mental act on the personal level, of being not disposed to conceive of our actions as based on reasons? In that mental act, a person would dissociate herself from her first-person self-ascription of performing a certain action – which is impossible against the background of my account of self-conceptions. After all, first-person self-ascriptions are not mediated by identification, or conceptualization and symbolization. If a person pre-reflectively experiences herself as the subject of a certain action structure and is reflectively aware of herself as the subject of that action structure (which is not up to her, so to speak), her conception of herself as the subject of her action structure entails the disposition to understand that action structure, that is to say, to know its purpose.

It is, however, imaginable that people can indirectly bring about a state of mind, by a mental act on the personal level, of not being disposed to conceive of their actions as based on reasons. A person might, for example,

decide to drug herself with alcohol in order to stop conceiving of herself as a rational agent. She might then, due to the effect of the alcohol, drift off into a mental state where her capability to conceive of herself of herself is (temporarily) disturbed. This person, even if she pre-reflectively experiences herself as the subject of her behaviour, may not *conceive* of herself as the subject of her behaviour since her capability to conceive of herself as herself might be temporarily diminished. And because her disposition to understand her actions as actions based on reasons is tied to her conceiving of her actions as her actions, she also (temporarily) lacks the disposition to understand what she is doing. In short, the person has indirectly brought about a state of mind in which her reflective capacities are damaged to such an extent that she does not conceive of her actions as her actions at all and in which she therefore lacks the first-person self-conception of acting based on reasons.

I should mention that my argument that a person cannot dissociate herself from her first-person self-conception of being a rational agent (a self-conception that is exhibited by her disposition to understand her actions, which in turn is tied to conceiving of her actions as her actions) differs, despite their similarity, from J. David Velleman's argument that "the desire to act in accordance with reasons cannot be disowned by an agent". He writes:

> Note that the desire to act in accordance with reasons cannot be disowned by an agent, although it can be disowned by the person in whom agency is embodied. A person can perhaps suppress his desire to act in accordance with reasons […] The only way for a person truly to suppress his concern for reasons is to stop making rational assessments of his motives, including this one, thus suspending the process of rational thought. And in suspending the processes of rational thought, he will suspend the functions in virtue of which he qualifies as an agent. Thus, the sense in which an agent cannot disown his desire to act in accordance with reasons is that he cannot disown it while remaining an agent.[40]

In order to compare Velleman's argument with mine, we have to replace 'agent' by 'rational agent', since Velleman's agent is what I call a rational agent; after all, Velleman's agent has a "desire to act in accordance with reasons". Although I agree with Velleman's concluding statement that a rational agent cannot disown her desire to act in accordance with reasons while remaining a rational agent, I disagree with his argument in two respects. Velleman states that a person cannot suppress her desire to act in accordance with reasons other than by suppressing the functions in virtue of which she counts as a rational agent. In my view, if a person could suppress the functions in virtue of which she counts as a rational agent, she

---

[40] Velleman 1992, pp. 141-142.

by the same mental act would suppress the functions in virtue of which she counts as a person; after all, our disposition to conceive of our actions as based on reasons is tied to our personhood.[41] However, a person cannot suppress the functions in virtue of which she counts as a rational agent – at least, if we take 'suppress' to be a mental act by which the person dissociates herself from a certain aspect of herself. Suppressing the functions in virtue of which a person counts as a rational agent would in the end imply that a person could dissociate herself from her fundamental first-person self-conception of herself as herself. But that dissociation is conceptually impossible against the background of my account of self-conception.

4 Dissociation of oneself from oneself: possibilities and impossibilities

The most essential points about the account of self-conceptions presented here are: firstly, the acknowledgement of a pre-reflective sense of (continued) selfhood as a prerequisite for self-conceptions; secondly, the distinction between first-person and third-person self-conceptions; thirdly, the clarification that first-person self-ascriptions of one's actions are bound up with the disposition to conceive of those actions as based on reasons; and finally, the recognition of the special position of the first-person conception of oneself as oneself. As we have seen, this account of self-conceptions makes it intelligible that people can dissociate themselves from certain 'aspects of themselves' but not from others. I take the expression 'aspects of oneself' in the broadest possible sense, meaning, for instance, desires, intentions, beliefs, actions, decisions, adopted goals and plans, things a person cares about, perceptions, sensations, processes of reasoning, processes of reflection, self-related reactive attitudes, self-narratives, or a person's alters.

4.1 Dissociation of oneself from aspects of oneself

Dissociations *of oneself* are dissociative mental acts that are performed by a person qua person; they are mental acts on the personal level, so to speak. The qualification of that category of dissociations as dissociations *of oneself* is meant to distinguish them from dissociative (mental) acts that happen on sub-personal levels and from dissociative reactions of the human organism to certain external or internal events. As to the former, we may think of the

[41] Let me briefly recapitulate why: Owing to our personhood – that is, our defining capacity to conceive of ourselves as ourselves – we conceive of ourselves as the subject of our actions and of those actions as our actions. And whenever a person conceives of a certain behavioural structure as her action, that very fact entails the disposition to know the purpose (or motivating reason) for her action, in other words, to conceive of her action as based on reasons.

temporary uncoupling between our conscious rational guidance and the performance of routine tasks, or we might think of mechanisms of repression of sexual feelings or of memories of traumatic experiences. Concerning dissociative reactions of the human organism, we might think of reactions of the human organism to drugs or of coping mechanisms ('freeze state') of the organism in extremely threatening situations. As emphasized earlier, my concern is with people's dissociations that can count as dissociations of oneself qua person; I do not claim to provide an account of dissociations in general.

Let me organize our most important findings concerning the possibilities of dissociations *of oneself* from (aspects of) oneself. I will first recall the main lines of my account of self-conceptions and then use that account as a framework to structure the possible dissociations of oneself from (aspects of) oneself and those that are impossible by their very nature.

(1) Talk about dissociation *of oneself* implies talking about persons, that is to say, creatures that conceive of themselves as themselves. A creature that does not conceive of itself as itself cannot dissociate *itself* from anything.

(2) Conceiving of oneself as oneself is the fundamental first-person self-conception. It is a self-*conception,* as it requires reflective capabilities; and it is a *first-person* self-conception, as it does not require identification of oneself as oneself, or conceptualization and symbolization of oneself qua self. And it is, in two ways, the *fundamental* first-person self-conception: firstly, because the capability to conceive of oneself as oneself is fundamental for personhood, secondly and entailed by the first, because it is a prerequisite for all other (first- and third-person) self-conceptions.

(3) First-person self-conceptions are self-ascriptions that are based on first-person access and which therefore do not require identification of oneself as oneself, or conceptualization and symbolization of oneself qua self. Examples are self-ascriptions of one's current perceptions, desires, intentions, actions, or thoughts. Third-person self-conceptions, by contrast, require mediation by identification of oneself – for example, my self-conception of being C.S. or my visual image of myself as a two-year old child – or by symbolization of oneself qua self – for example, the self-conceptions of having a certain biography, of being a certain kind of person, or of pursuing a certain long-term goal.

(4) One self-conception – the self-conception of being a rational agent – deserves to be mentioned separately. We have the implicit first-person self-conception of being rational agents, that is to say, of acting based on reasons. That self-conception is bound up with our conception of our actions as *our actions*. The first-person self-conception of acting based on reasons has to be distinguished from a self-image of being a rational kind of person, which can have different content for different people, or people may not have that third-person self-conception at all.

(5) All acts of conceiving of oneself are reflective acts of self-awareness. They require a pre-reflective self-awareness, more specifically, a pre-reflective sense of (continued) selfhood: pre-reflectively experiencing oneself as the self that has been and continues to be. A sense of selfhood is not identifiable as this or that sense of selfhood. It is content-free – nothing more than experiencing oneself as *being-self* (over time).

Dissociations *of oneself* are dissociative mental acts that are performed by a person qua self-comprehending creature. This characterization forms a constraint on possible dissociations of oneself. It implies that we can only dissociate ourselves from something if we are reflectively aware of ourselves, in that moment. And there is a second constraint. We can only dissociate ourselves *from* something, more specifically, from aspects of ourselves if we, in that moment, conceive of ourselves as having those aspects, whether that self-conception is a first-person or a third-person self-conception. Needless to say, a dissociative mental act of a person is always an act of reflective self-awareness.

The following kinds of dissociation of oneself from aspects of oneself are conceivable against the background of my account of self-conceptions.

(a) We can dissociate ourselves from former third-person self-conceptions. Because of being third-person self-conceptions they were mediated by identification, or conceptualization and symbolization; and we can dissociate ourselves from those self-images by now regarding those mediations as influenced by internal or external conditions. Such evaluative dissociations can be part of explicit reflections on, for example, our former goals, or things we cared about. They are performed in the context of current third-person self-conceptions; after all, they require thinking of oneself in other circumstances in the past or in the future, and hence require conceptualization and symbolization of oneself.

(b) We can retrospectively distance ourselves from former first-person self-ascriptions, for example desires, intentions, thoughts, or actions, by regarding them as caused by, for instance, excessive alcohol consumption. Such retrospective evaluative dissociations require that we think of ourselves in other circumstances in the past; they are therefore performed in the context of current third-person self-conceptions. It is important to notice that we – although in a restricted sense – still ascribe those desires, intentions, thoughts, or actions to ourselves, that is to say, we still conceive of ourselves as their subjects.

(c) People can retrospectively distance themselves from, for example, certain actions, although they may have never experienced themselves as the subjects of those actions. Retrospectively, however, they may ascribe those actions to themselves as part of their current comprehension of a certain situation and, at the same time, distance themselves from those actions, because they regard them as caused by, for example, excessive

alcohol consumption. It should be emphasized that that distancing can only count as dissociation of a person from aspects *of herself* if having performed these actions is part of the person's current third-person self-conception.

(d) People can – as part of their first-person self-ascriptions – partially dissociate themselves from, for example, current desires or actions if they do not experience themselves as their source. Those dissociative acts of reflective self-awareness can only be partial, because in order to count as a person's dissociation from an aspect *of herself*, the person still has to conceive of those desire or actions as hers – although in the restricted sense of merely being their subject. Dissociations of this kind are based on (pathological) experiences of non-agency.

It should not pass unnoticed that dissociations of oneself from aspects of oneself, because they are acts of reflective self-awareness, are performed in the context of a person's current conception of herself – whether that self-conception is based on third- or first-person access. And a *current* act of self-conception is, if for the sake of argument we use the terminology of a subject-object model[42], inextricably bound up with the conceiving subject; it can never be the object of a person's conception, in that very same moment. For example, when we (in an act of first-person self-conception) conceive of our action as our action, or when we (in an act of third-person self-conception) conceive of ourselves as a certain kind of person, we cannot, in the same moment, ascribe that act of conception to ourselves. The act of self-conception itself is not conceived of, in the present moment. Therefore, people can never dissociate themselves from their current act of self-conception. This brings us to the limits regarding dissociation of oneself from (aspects of) oneself.

4.2 The limits to dissociation of oneself from oneself

Certain dissociations are by their very nature impossible against the background of my account of self-conceptions.

(a) People cannot dissociate themselves from their fundamental first-person self-conception of conceiving of themselves as themselves. After all, dissociation of oneself from anything whatsoever requires conceiving of oneself as oneself.

(b) People cannot dissociate themselves, or rather, cannot dissociate themselves altogether, from aspects of themselves that they currently have first-person access to. After all, dissociation of oneself from, for example, a thought or action that one currently has first-person access to

[42] By formulating (as I do in the sentence that this note refers to) "a *current* act of self-conception can never be the object of a person's conception", I in fact use the terminology of a subject-object model, which could mistakenly be seen as evidence that I in fact presuppose a perceptual model or "Reflexionsmodell" of self-consciousness. See also note 25 in this chapter.

requires (experiencing and) conceiving of oneself as the subject of that thought or action.

(c) People cannot dissociate themselves from current third-person self-conceptions; that is to say, people cannot dissociate themselves from third-person self-conceptions in the same moment in which they entertain those self-conceptions. After all, the act of conceiving of oneself as one does is inextricably bound up with the conceiving subject. However, a person can obviously distance herself from a third-person self-conception immediately after having formed that self-conception; the dissociative act may even be part of the same process of deliberation in which that third-person self-conception was formed originally.

(d) People cannot dissociate themselves from their current acts of conceiving of (aspects of) themselves. After all, one's current act of conceiving is inextricably bound up with oneself as the comprehending creature; it can never be on 'the object side' of one's self-conception, so to speak. A person can, for example, not dissociate herself from regarding certain of her actions as wrong if she, in that moment, regards them in that particular way. Retrospectively, she obviously may distance herself from her former judgement by conceiving of that judgement as, for example, strongly influenced by others.

(e) People cannot dissociate themselves, if they conceive of an action as their action, from implicitly conceiving of themselves as rational agents. In other words, people cannot distance themselves from their disposition to conceive of themselves as acting based on reasons, at least if they conceive of certain actions as their actions in the full sense (that is to say, as being their subject and their source). After all, conceiving of one's action as an action is tied to conceiving of that behavioural structure as purposeful behaviour and of that purpose as one's purpose, as one's motivating reason for the action in question.

The listed possibilities and impossibilities of people's dissociations from aspects of themselves make it very clear that people cannot dissociate themselves from themselves qua self – if we take that expression literally. Such a dissociative state of mind would entail that the creature's fundamental first-person self-conception – her conception of herself as herself – were (temporarily) broken down. In that case, however, we cannot speak of a person anymore, in that moment. That is why the notion of being dissociated from one's self is paradoxical: it implies that there still is a person, a conceiving self, so to speak, and, at the same time, that the self-self relation of conceiving of oneself as oneself, which is fundamental for personhood, has collapsed.

Chapter 7
Conclusion: Personhood, models of self-conception, and regions of mattering

1 About the project, objections to the proposed account, and the rest of the chapter

ABOUT THE PROJECT

The account of self-self relations presented in this book – that is, relations of a person qua self-comprehending creature to herself qua self – regards self-self relations as inextricably bound up with personhood. Persons and only persons can think and feel about themselves qua self. The important point here, however, is not that persons characteristically think and feel about themselves in one way or the other, but that certain ways of thinking and feeling about themselves qua self are characteristic of persons. Those self-self relations are based on 'models of self-conception', which in turn are bound up with properties that are constitutive of personhood (in its fullest sense): as beings with a first-person perspective we implicitly conceive of ourselves as rational agents, as narrative beings we implicitly conceive of ourselves as the protagonists of our self-narratives, as beings that apply self-regarding qualitative standards we implicitly conceive of ourselves as having values. All three properties with their included self-conceptions constitute separate regions of mattering with different structures and dispositions; all three are hence the basis for mattering-relations with ourselves qua self that we qua persons establish, entertain, or experience. A clear or full understanding of personhood can therefore not refrain from an analysis of self-self relations. And an analysis of self-self relations, at the same time, contributes to a fuller understanding of personhood. Indeed, an account of personhood that fails to confront self-self relations – or rather, those self-self relations in which (fully developed) persons cannot but stand to themselves qua self – is simply deficient.

To uncover these self-self relations, I analysed pre-philosophical notions regarding our selves qua self, pathological phenomena, and various 'normal' ways in which persons – in the context of action, deliberation, and self-evaluation – think and feel about themselves. Such a project simply *has*

to take the first-person perspective as the thread running through the analysis.

ABOUT OBJECTIONS TO THE PROPOSED ACCOUNT

My view on personhood touches upon a number of topics, all of which are subject to continuous philosophical debate. More specifically: (1) it touches upon the metaphysical questions about the nature of persons and the self; (2) it confronts the issue of personal identity; (3) it includes a view on (acts of) self-awareness; (4) it includes an account of rational agency and rational action; (5) it proposes a view on values; (6) it takes up the question of self-knowledge; and (7) it confronts the issue of dissociation and multiple selves. Such an account cannot but give rise to objections by philosophers of different conviction. My view on self-awareness, for example, can neither count as a purely analytical nor as a phenomenological account; it can, hence, expect critiques from both camps. My view on rational action may lead to objections, because the rationality of an action is, as I take it, not a matter of compliance with certain standards. My view on values might, for example, be challenged by moral realists. My view on personal identity may be disputed, as my definition of a person's fundamental identity is, in a way, circular. Finally, my view on 'multiples' and my argument for the impossibility of a person's dissociation from her self, is sure to raise objections by those who think that more than one self can inhabit a person's body. Needless to say, this list of possible objections is far from complete.

One possible set of objections should be mentioned separately, however. An account that is developed in a first-personal rather than a third-personal approach is bound to be accused of relativism or solipsism. Moreover, the cogency of the presented account might be challenged by arguing that first-personal experiences depend on an individual person's 'state of mind'. Such critiques actually attack the presented account precisely because of what is essential to it – namely that it takes the first-person perspective seriously. As to the first objection, I contend that taking the first-person perspective seriously unfolds aspects of personhood that are hidden to a third-personal approach while being essential to a full understanding of personhood. As to the second objection, I claim that certain ways of thinking and feeling about oneself qua self are not contingent, but inextricably bound up with the properties that are constitutive of personhood (in its fullest sense). This is not to say that *all* persons establish or entertain *all* those self-self relations, because, in my view, not all properties that are characteristic of personhood, taken in the fullest sense, are fundamental to personhood.

ABOUT THE REST OF THE CHAPTER

This concluding chapter is not intended to summarize the conclusions of the foregoing chapters but to draw connections between them in the light of the quest for 'a person's relation to herself qua self'. The view that is presented here depends on findings of *all* foregoing chapters. First, I will briefly discuss the main aspects and implications of the proposed account of person- and selfhood. I will then summarize my view on self-self relations. Thirdly, I will focus on those self-self relations that are, if I am right, inextricably bound up with properties that are constitutive of personhood (in its fullest sense). To conclude, I will briefly address some new questions, which evolved during this project and quite a number of which would benefit from an interdisplinary approach.

In what follows, I will put emphasis on those aspects of my view on person- and selfhood that I expect to be most controversial. Table 7.1 provides a list of those claims. No doubt, some critics will also disagree with the concept of a person that forms the basis for my account, namely Lynne Rudder Baker's view that the essential capability of a person is her capability of a (highly developed) first-person perspective. Here, however, I will concentrate on claims that are the result of my *own* project (this is not to diminish the inspiration that I owe to other philosophers).

Table 7.1: Controversial claims

Section 2 ***Personhood, selfhood, and the first-person perspective***

- (a) Personhood and selfhood are two sides of the same coin.
- (b) The first-personal view is always prior.
- (c) Selfhood has a pre-reflective predecessor: namely a sense of (continued) selfhood.
- (d) Whether an action is rational solely depends on the person's self-conception of her action as resulting from her process of reasoning.
- (e) Personhood in its fullest sense involves more than rational agency.
- (f) Strong evaluative reasoning implies caring about being the kind of person whom particular values can be justifiably attributed to.
- (g) Self-narratives are not identity-constituting in a fundamental sense.
- (h) The fundamental self-conception of oneself as oneself is content-free.
- (i) Persons cannot but conceive of themselves as themselves.

Section 3 ***Self-self relations***

- (a) Only persons can stand in relations to themselves qua self.
- (b) Self-self relations – with the exception of the fundamental self-self relation of conceiving of oneself as oneself – do not concern one's self as such but aspects of oneself qua self.
- (c) Meta-reflection about oneself qua self and explicit deliberation about or evaluation of one's actions requires a detached stance and

conceptualization and symbolization of oneself qua self.

(d) Reflective self-self relations do not necessarily require a detached stance. Examples are the fundamental self-self relation of conceiving of oneself as oneself and self-regarding thoughts and feelings that are embedded in one's reasoning in the context of an action.

(e) Persons are not necessarily reflectively aware of their various self-self relations; they may also entertain pre-reflective self-self relations.

Section 4 ***Concealed first-person self-conceptions: the basis for mattering-relations to oneself qua self***

(a) Personhood includes three 'models of self-conception' with their own structures, dispositions, and desires: conceiving of oneself as a rational agent, as the protagonist of one's self-narrative, and as having values.

(b) The fundamental conception of oneself as oneself and the 'models of self-conception' are not themselves 'objects' of one's direct reflective self-awareness.

(c) The fundamental self-conception of oneself as oneself and the three 'models of self-conception' have to be clearly distinguished from seemingly corresponding self-images – for example, the self-image of being a rational kind of person, or the thought 'I am myself'.

(d) The fact that our self-comprehension cannot but respond to our conception of ourselves as ourselves and to the three 'models of self-conception' makes intelligible that certain things concerning ourselves qua self matter to us.

(e) We can distinguish four intrinsic self-regarding concerns: a concern with the persistence of ourselves qua self, with the soundness of our 'will', with the coherence of our self-narrative, and with the integrity of ourselves. Hence, our own irrationalities, inconsistencies, and incoherences matter to us not only because they hinder the accomplishment of our aims, but also much more directly than that.

---

2 Personhood, selfhood, and the first-person perspective

The starting point of an adequate analysis of self-self relations – the relation of oneself qua self-comprehending creature to oneself qua self – is the notion of a person. After all, persons are exactly those creatures that are able to conceive of themselves as themselves. How could a non-person ever stand in any relation to itself qua *self*? It simply does not make any sense to speak about a self if a creature is not able to conceive of itself as itself. My account of personhood thus involves a notion of selfhood.

Persons and selves are not different 'things', but the two notions are two sides of the same coin: the third-personal notion of a person corresponds with the first-personal notion of a self. Nevertheless, there is an important

distinction between the notions of person and self: the first-personal view is always prior. Selfhood – that is, one's reflective awareness of *being-self*, which means nothing else than conceiving of oneself as oneself – is prior to personhood in two respects. Ontologically, the existence and persistence of a person depends on the capability to conceive of oneself as oneself. Epistemologically, it is only once we know that a creature has the capability to conceive of itself as itself that we can ascribe personhood to it.

And there is a further important distinction between the notions of selfhood and personhood: selfhood, in contrast to personhood, has a pre-reflective predecessor, namely a pre-reflective sense of (continued) selfhood. It is only once we conceptually presuppose the 'givenness of the self' that it is intelligible that we are able to conceive of ourselves as ourselves – without identifying ourselves as ourselves. But there is evidence that a pre-reflective sense of (continued) selfhood is not a philosophical invention. If we observe complex purposeful behaviour in, for example, infants, certain non-human animals, or people performing tasks on autopilot, we cannot but presuppose a pre-reflective sense of continued selfhood. After all, complex purposeful behaviour requires that the creature experiences itself as one and the same subject in the various activities that it performs to actualize a particular purpose and in the perceptions and sensations that may appear during the corresponding period of time. However, complex purposeful behaviour, while requiring a pre-reflective sense of continued selfhood, is no evidence of personhood. Personhood requires a *reflective* sense of selfhood.

A person's fundamental property is her capability to conceive of herself as herself. Based on that property, she conceives of her actions as *her* actions and of her motives for her actions as *her* motives for *her* actions – in short, she conceives of herself as acting based on reasons. That is why a person is a rational agent. Rational agents are normally able to provide rationales for their actions. And whether an action *is* rational does not depend on the 'quality' of the reasons provided. It solely depends on the person's self-conception of her action as resulting from *her* process of reasoning, whether her understanding of her action is a rational reconstruction or not.

Again, the first-personal view is prior to the third-personal one. It is only once we know that the agent understands herself as (capable of) reasoning and deciding what to do that we can talk of *reasons* for her actions at all. The same holds for a concrete action. Talk about a person's *reasons* to perform a concrete action only makes sense if we think that she herself thinks there is something to be said in favour of doing what she did, does, or plans to do. It is only once we know how a person understands her own action – for example, which considerations she regarded as relevant to her course of action – that we can ascribe to her particular *reasons* for her action – for instance, (the content of) certain desires, beliefs, or goals. And it is only then that we might judge whether the person's reason was a good reason to perform the action in question – 'good' in whatever sense.

Personhood in the fullest sense involves more than rational agency, in

other words, more than understanding oneself as (capable of) reasoning and deciding. Persons characteristically, though not necessarily, have the capability to apply self-regarding qualitative standards in their reasoning concerning their actions. When a person applies this type of evaluative reasoning – that is, strong evaluative reasoning – she does not only make reference to her goals and purposes but also – or rather, primarily and most importantly – to her values. This is not to say that evaluative reasoning – in the weak sense of weighing alternatives – is not normative. Facts about a certain action alternative – for example, that a person, by performing that action, can effortlessly actualize a certain purpose, fulfil more than one of her current desires, or fulfil the strongest one – are not relevant unless we suppose that such facts, at that moment, are important to the person in question (in the sense of, for that person, arguing in favour of a certain action alternative) so as to constitute reasons for her to perform a particular action. However, performing a certain action because it, say, fulfils two of one's current desires does (leaving aside pathological cases) not imply that one *cares* about being a multiple-desire-fulfiller. By contrast, if a person applies self-regarding qualitative standards in her evaluative reasoning concerning her actions, she implicitly *cares* about being the kind of person whom particular values can be justifiably attributed to.

Personhood in the fullest sense also involves the property of having a narrative self-conception. Persons characteristically, though again not necessarily, (try to) connect their intentions, actions, experiences, practical beliefs, plans, goals, character traits, values, and etcetera. They interpret and re-interpret what they do, feel, think, and experience, and position those elements in a self-narrative. Self-narratives may be constructed in processes of meta-reflection; often, however, they are provided as part of action explanations. It is important to notice that self-narratives, because of consisting of interpretations, are *mediated* self-conceptions. They require that the person takes a detached stance and symbolizes herself qua self. Therefore, self-narratives cannot be identity-constituting in a more fundamental sense.

A self-narrative is a 'super self-image', containing numerous mediated self-conceptions, for example, one's belief of being sensitive to others, one's belief of being C.S., one's belief of being a rational kind of person, or one's belief that money is not important to oneself. A person may someday, say, dismiss her earlier belief regarding money, because her ideals have changed. She may revise her belief that she is C.S., because she finds her original birth certificate that states otherwise. Or she may come to realize that she is not the kind of person that is sensitive to others, and that she is not rational either ('rational' in a qualitative sense). But despite the change in her 'super self-image', she never did *not* conceive of herself as herself. It is this fundamental conception of oneself as oneself that constitutes one's identity in the most basic sense. It is simply impossible that we – in our fundamental conception of ourselves as ourselves – could misidentify ourselves or could

be mistaken regarding the 'content of our self'. When conceiving of ourselves as ourselves, we do not conceive of ourselves as such and such or as a certain kind of person, in other words, we neither identify ourselves as ourselves nor symbolize ourselves qua self.

Personal identity, in this fundamental sense, is sometimes criticized as being empty. And indeed, our fundamental self-conception as ourselves is content-free, so to speak. However, its being content-free is precisely one of the two reasons why we cannot be mistaken concerning our identity and why this kind of identity is our most fundamental one. We can distance ourselves from a former autobiographical narrative and hence from a former 'super self-image', but we cannot distance ourselves from ourselves qua self, or from our self-understanding of ourselves as ourselves. Neither can we deliberately stop conceiving of ourselves as ourselves. As long as we are *someone*, that is, a being with a first-person perspective, we cannot but conceive of ourselves as ourselves (at least, in circumstances when we are reflectively aware of anything at all). That fundamental self-self relation can never collapse – not without damaging or destroying the features by virtue of which we count as a person, in other words, not without us ceasing to exist.

3 Self-self relations

The fact that we are *persons* involves that we entertain, establish, and experience kinds of self-relations that are peculiar to persons. Only persons can stand in relations towards themselves qua *self*. With the exception of the fundamental self-self relation of conceiving of oneself as oneself, all other self-self relations concern *aspects* of one's self. That is to say, the self-regarding thoughts or feelings do not relate to one's self 'as such', but to 'things' that we attribute to ourselves qua person, in contrast to 'things' that we attribute to ourselves qua body. Examples of such aspects of oneself are one's beliefs, intentions, desires, moods, actions, purposes, goals, plans, ideals, values, character traits, personality, or biography.

Obviously, persons can think or feel in various ways about aspects of themselves qua self. Such thoughts or feelings need not be highly reflective, in the sense that the person engages in *meta-reflection* about her character, her values, or the kind of person she would like to be. The self-regarding attitudes need not even be part of an *explicit* deliberation about or evaluation of certain of one's actions. Self-regarding thoughts and feelings are often simply *embedded* in the evaluative reasoning that takes place in the context of one's current action. A person may, for instance, reject a certain spontaneous desire, without explicit deliberation – possibly, because acting based on that desire would hinder the realization of one of her adopted goals, or because she may not want to be moved by that kind of desire.

Whether one's self-regarding attitudes are part of an act of meta-reflection, explicit reflection or embedded reflection, the act itself is a *reflective* act of

self-awareness. And the accompanying self-regarding thoughts or feelings – like feeling regret or remorse about one's behaviour, regarding one's action as irrational or wrong, or feeling satisfied or dissatisfied with one's way of life – are therefore *reflective* self-self relations, that is, relations that a person qua self-comprehending creature establishes, entertains, or experiences regarding aspects of herself qua self. This is not to say that persons (in circumstances when they are reflectively aware of anything at all) are always, or even necessarily, reflectively aware of their various self-self relations. Persons may entertain *pre-reflective* self-self relations; they may, for instance, have repressed (sub-conscious) self-regarding feelings or beliefs.

Reflective self-self relations, because of being reflective, may seem to presuppose a detached stance towards oneself. Thorough analysis, however, refutes that presupposition. First of all, a person's fundamental reflective self-self relation of conceiving of herself as herself does *not* require conceptualization and symbolization of herself qua self. Neither do one's self-regarding thoughts and feelings that are *embedded* in one's reasoning in the context of one's current action. Embedded self-reflection, although the person takes a position regarding her motives for action (for example, regarding her spontaneous desires), does not mean that she conceptualizes and symbolizes herself qua self. Obviously, it is only once the person provides an account of herself as a rational agent in the context of a concrete action, that we can ascribe to her an act of embedded reflection and certain accompanying self-regarding feelings or thoughts. The first-personal view is thus again prior to the third-personal. In practice, the distinction between embedded and explicit reflection may sometimes be difficult (or even impossible) to make; nevertheless, there is a clear conceptual distinction.

4 Concealed first-person self-conceptions: the basis for mattering-relations to oneself qua self

Although persons can think and feel in various ways about themselves qua self, not all reflective self-self relations are contingent. Personhood does not only – by definition, so to speak – include the fundamental reflective self-self relation of conceiving of oneself as oneself, but it moreover includes 'models of self-conception': firstly, conceiving of oneself as a rational agent, secondly, conceiving of oneself as the protagonist of one's self-narrative, and, thirdly, conceiving of oneself as having values. These 'models of self-conception' have their own structures, dispositions and desires, which exhibit themselves in various self-regarding thoughts and feelings, and constitute separate *regions of mattering*. The first of those 'models of self-conception' is bound up with the fundamental property of having a first-person perspective, the second with the property of having a narrative self-conception, and the third with the property of applying self-regarding qualitative standards.

The fundamental reflective self-conception of oneself as oneself and the

three mentioned 'models of self-conception' (all of which obviously form *reflective self-self relations*) can best be thought of as *concealed first-person self-conceptions*. They are c*oncealed,* so to speak, because they are never themselves 'object' of our direct reflective self-awareness – although they are at the heart of our self-comprehension. They are *first-person* self-conceptions, because they do (in contrast to self-images, which are therefore third-person self-conceptions) not require a detached stance towards oneself, and conceptualization and symbolization of oneself qua self. Thoughts like "I am a rational agent" or "I have values" are hence not instantiations of the seemingly corresponding 'models of self-conception', but constructed self-images, which require conceptualization and symbolization of oneself qua self. Our concealed first-person self-conceptions do, however, not require conceptualization and symbolization of oneself qua self; they simply exhibit themselves in various ways in which we comprehend ourselves and in various self-regarding reactive attitudes. To avoid misunderstanding, concealed first-person self-conceptions should not be thought of as pre-reflective self-self relations; they should thus not be confused with, for example, repressed self-regarding feelings or beliefs.

The fact that our self-comprehension cannot but respond to our fundamental conception of ourselves as ourselves and to three 'models of self-conception'[1] makes understandable that certain things concerning ourselves qua self (deeply) matter to us. First of all, owing to our self-conception of ourselves as ourselves, it matters to us whether we *ourselves* will have sensations, perceptions, and experiences in the future. In other words, based on our (highly developed) first-person perspective, we are concerned with the persistence of ourselves qua self. This concern is, I believe, well enough understood in the literature, in contrast to the three concerns that belong to our three 'models of self-conception' and which major parts of this book are devoted to: the concern with one's rational agency, one's self-narrative and one's values.

Before turning to these three self-regarding concerns, let me give some thought to our concealed first-person self-conception of ourselves as ourselves, and – as an example of the three aforementioned 'models of self-conception' – to our concealed first-person self-conception of being a rational agent.

4.1 The fundamental first-person self-conception of oneself as oneself

The self-conception of oneself as oneself is, first of all, not contingent but necessary, because we cannot but entertain that reflective self-self relation – at least if we are, at that moment, reflectively aware of anything at all.

---

[1] It is presupposed that we not only have the property of a first-person perspective, but also the non-fundamental (but characteristic) properties of a narrative self-conception and of applying self-regarding qualitative standards.

Moreover, our fundamental self-self relation is concealed. It exhibits itself in all self-ascriptions of desires, beliefs, feelings, perceptions or sensations, in all self-images, like being such and such or a certain kind of person, and in various other acts of reflective self-awareness, like self-evaluation, self-criticism, reflection on (one's motives for) one's actions, evaluative dissociation from certain aspects of oneself, or dissociative experiences. However, our fundamental self-conception of ourselves as ourselves is as such never itself 'object' of our direct reflective self-awareness. All attempts to come to grips with that self-conception result in thoughts in which we take a detached stance and conceptualize, symbolize, and identify ourselves qua self, for example "I am I" or "I am myself". Such thoughts (because of requiring conceptualization, symbolization, and identification of oneself qua self) are third-personal self-conceptions, which obviously – as they are acts of reflective self-awareness – exhibit the concealed self-conception of oneself as oneself. It follows naturally that even thoughts that seem to deny one's fundamental self-self relation – like "I am not identical with myself", "I am not myself", or "I am (or have) more than one self" – exhibit that person's fundamental self-self relation of conceiving of herself as herself.

4.2 The first-person self-conception of being a rational agent – an example of a 'model of self-conception'

The concealed conception of ourselves as rational agents becomes evident from the fact that we, generally, can provide rationales for our actions, that is, sets of reasons on which our actions are based. This is not to say that we always, or even normally, comprehend our actions (before, during, or after their performance) as based on particular reasons. But as soon as we are asked why we are doing what we are doing, why we did what we did, or why we plan to do what we say we plan to do, we are generally able to provide an account of our actions in terms of reasons. It does not matter whether that account is a rational reconstruction or whether the provided reasons actually are the reasons that moved us to act – at least not for being testimony to our concealed self-conception of being rational agents. That 'model of self-conception' becomes evident from our disposition to comprehend our actions as based on reasons; it is as such not a subject of our direct reflective self-awareness.

The concealed self-conception of being a rational agent is inextricably bound up with one's property of having a first-person perspective, which implies that a person generally conceives of herself as the subject of her actions, of her actions as her actions, of her actions as based on motives, and of those motives as her motives. This concealed self-conception of being a rational agent has to be clearly distinguished from the contingent self-image of being or striving to be a rational kind of person. A person can easily retrospectively distance herself from her former self-image of being or

striving to be a rational kind of person (whatever image of being rational she may have in mind), but she cannot dissociate herself from her 'former' concealed self-conception of being a rational agent. It just makes no sense to speak of a concealed self-conception in terms of former and current, exactly because of its being concealed.

To avoid misunderstanding, people can obviously dissociate themselves from particular reasons for their current or former actions – for example, from certain desires that move or moved them to act – but only owing to their disposition to conceive of their actions as based on (their) reasons. Even a person who experiences a certain desire that moves her to act as ego-alien (that is to say, she conceives of herself as the subject of that desire, but not as its source) and who therefore does not conceive of herself as the source of her action, exhibits the disposition to conceive of her actions as based on (her) reasons. Her concealed self-conception of being a rational agent exhibits itself precisely by her experiencing herself as an agent who does *not* herself decide what she is doing.

4.3 Three regions of mattering based on the three 'models of self-conception'

The three 'models of self-conception' – that is, conceiving of oneself as a rational agent, as the protagonist of one's self-narrative, and as having values – constitute separate regions of mattering: one's rational agency, one's self-narrative and one's values. Whether we engage in self-evaluation and self-criticism or not, whether we thematize the fact that these things matter to us or not, even if we deny that they matter to us, as long as we have the properties which these concerns are bound up with, we simply stand in those mattering-relations to ourselves.

First, the fact that we comprehend ourselves as rational agents – in other words, that we cannot but entertain that 'model of self-conception' – includes a concern with our rational agency. This concern exhibits itself in various self-regarding feelings and thoughts. For example, we generally feel confused if we find ourselves doing something without knowing why we are doing what we are doing. We distance ourselves from those actions which we, while still conceiving of ourselves as the subject of those actions, do not experience as originating from ourselves qua self. And we, more or less frequently, conceive of certain of our actions as irrational (sometimes accompanied by feelings of regret or dissatisfaction with ourselves), namely when we, while conceiving of those actions as resulting from our own reasoning, regard them as hindering or obstructing the achievement of our adopted purposes.

The claim is that those situations bother us not only because they, for example, reveal our forgetfulness, addiction or failure to achieve our adopted goals, but also much more directly than that. They do not fit in with our (concealed first-person) self-comprehension of being a rational agent. If I am

right, it is the situation where we regard our actions as irrational that is especially threatening to our self-conception of being a rational agent. In the first two cases, we may experience those actions as not under our control, in the last case, however, we conceive of our irrational action as resulting from our *own reasoning*. We are confronted with a divergence in our 'will', so to speak. On the one hand, we do (or did) what we ourselves decide(d) to do; but on the other hand, we do (or did) not do what we 'really' want(ed) to do – namely (generally speaking) to avoid actions that hinder or obstruct our adopted purposes and to perform actions that are conducive to the realization of those purposes. According to my account, the experience of what I label a divergence in our 'will' frustrates a desire which is included in our concealed first-person self-conception of being a rational agent and which can best be thought of as a desire to justifiably conceive of ourselves as rational agents. The normative criterion concerning ourselves qua self, which obtains in this desire, is consistency.

A second 'model of self-conception' is our comprehension of ourselves as the protagonist of our self-narrative. It includes a concern with (the coherence of) that narrative. For example, when engaged in explaining our actions (to others or to ourselves), we generally feel the need to account for apparent changes in our goals. This inclination to make these changes intelligible exhibits one's concern with one's self-narrative. Even if someone claims that the coherence of her self-narrative does not matter to her, she, by providing a narrative explanation, *is* concerned with a certain level of coherence; after all, she could otherwise not conceive of herself as the protagonist of that explanatory narrative at all. And moreover: Why should she, if she did not have a narrative self-conception, entertain beliefs about the required level of coherence of her self-narrative anyway?

The fact that (the coherence of) our self-narrative matters to us does not only express itself in our inclination to provide an account of ourselves qua self (that is, of our actions, goals, professional career, experiences, things we care about, etcetera) in terms of a self-narrative, but also in certain positive and negative self-regarding thoughts and feelings. A person may, for example, conceive of herself qua self as irrational when her 'evaluative agenda' is characterized by incoherences, that is to say, when her adopted long-term goals and purposes typically conflict with one another. An incoherent evaluative agenda entails a fragmented self-narrative – at least, a self-narrative with a lot of twists and turns. And my claim is that such fragmentations and incoherences bother us not only because we probably will (or have) achieve(d) less than we can (could have done), but also more directly than that. Experiencing fragmentation of one's self and feeling the need for reconciliation is often accompanied by emotions like sadness about oneself or dissatisfaction with oneself. Significant fragmentation and incoherence, after all, does not fit in with our (concealed first-person) self-comprehension of being the (well-defined) protagonist of our autobiographical narrative. That 'model of self-conception', if I am right,

includes the desire to be able ('able' owing to one's way of living a life) to conceive of ourselves as the well-defined protagonist of our self-narrative. The normative criterion concerning ourselves qua self, which obtains in this desire, is coherence.

Finally, the self-comprehension of having values, which belongs to the property of applying self-regarding qualitative standards, forms a third 'model of self-conception'. It includes a concern with one's values and typically exhibits itself in (often particularly strong) negative self-regarding feelings or thoughts – like conceiving of one's action as blameworthy, objectionable, or wrong, and/or experiencing feelings of regret, remorse, or contrition, or even of not being worthy of self-esteem. Explicitly regarding some of our actions as right happens less frequently than regarding certain of our actions as wrong. Also, the feelings accompanying our regarding our actions as right are less vivid, and they do not typically lead to further reflections on our motivations for action and on the qualitative standards we apply to ourselves. It is the deviation from our own self-regarding qualitative standards that provokes the more salient self-regarding reactive attitudes or reflexive emotions. The claim is that our wrongdoings typically bother us (deeply), and that that characteristically leads to further reflection on our motives for action, to self-evaluation and/or self-criticism.

According to my account, persons who apply self-regarding qualitative standards implicitly *care* about being the kind of person who has particular values. They are concerned with their values, and that concern, if I am right, is a desire to justifiably conceive of oneself as a person who cares about particular values. This is obviously not to say that we explicitly entertain thoughts like "I care about being the kind of person who has (such and such) values", or "I want to justifiably conceive of myself as a person who cares about (such and such) values", or – when disqualifying our actions – "I do not want to be such kind of person". However, our salient self-regarding reactive attitudes and reflexive emotions show that we experience the integrity of ourselves qua self to be at stake when we regard our actions as wrong.

4.4 Intrinsic concerns with ourselves qua self

To conclude, I should again emphasize that the concerns with ourselves qua self are inextricably bound up with the properties that constitute personhood (in its fullest sense). The fact that our rational agency, our self-narrative, and our values – or to put it more abstractly, the consistency, coherence, and integrity of ourselves qua self – matter to us, does not imply that we entertain the corresponding self-regarding beliefs. We do not characteristically think "I am concerned with my rational agency", "I desire to be able to conceive of myself qua self as the well-defined protagonist of my autobiographical narrative", "I care about being a person who has particular values", or "I care

about the consistency, coherence, and integrity of myself qua self". These kinds of thoughts are typically results of acts of meta-reflection on oneself qua self; they are detached from the structures, dispositions and desires that lie at the heart of our self-comprehension. That self-comprehension with its included concerns simply exhibits itself in our 'active engagement' with ourselves qua self – most vividly in certain negative feelings concerning our decisions, actions, or our life.

To recapitulate, my claim is that our own irrationalities, our wrongdoings and incoherences matter to us not only because they hinder the accomplishment of our aims, but also much more directly than that. The *strength* of people's concerns with the consistency, coherence, and integrity of themselves qua self can, however, vary from one person to another, since the properties that constitute personhood – or rather, the 'models of self-conception' with their different structures, dispositions, and desires – can, even in mature stages of personhood, be developed to different degrees.

5 New questions

This book is nearing its end. My quest for our relation to ourselves qua self is, however, not completed. I touched upon a variety of questions concerning personhood, selfhood, self-conceptions, and self-relations, and I dug extensively into quite a number of them – yet, still new questions keep arising. Here is a selection.

5.1 Regarding selfhood

THE QUESTION OF SELF-ALIENATION

When we reflect on our actions, our plans and goals, our character traits, or the things we care about, our attention is directed at *aspects* of ourselves qua self. We cannot reflect on our self '*as such*' or 'in its entirety'. As soon as we try to focus on our self 'as such', our self seems to slip away and we again and again find ourselves thinking about particular aspects of our self, for example, about our typical self-regarding moods and feelings. We simply cannot reflect on our self 'as such', for one's self is unidentifiable (and it is content-free). This implies that we, in a certain sense, do not coincide with ourselves qua self, as soon as we perform acts of self-reflection. We might therefore say that we frequently and regularly – after all, we frequently and regularly perform acts of self-reflection – are in a state of self-alienation. We might even say that a person becomes the more alienated from herself qua self, the more she is engaged in reflection about herself qua self, since such meta-reflection about one's self requires a highly detached stance towards oneself.

Doesn't this conclusion conflict with the ancient dictum "know thyself"; in other words, doesn't explicit reflection on oneself qua self enhance one's self-knowledge? Well, one could dispose of the problem, simply by arguing that we now confuse two different notions of the self: the ancient dictum "know thyself" presupposes a rich notion of the self, whereas the thoughts concerning self-alienation presuppose a 'neutral' (that is, unidentifiable and content-free) self. The question of the connections between self-knowledge and self-alienation deserves more attention than that, however. I am convinced it would give us new insights if we analysed the notions of self-knowledge and self-alienation and their connections in the light of my account of personhood and selfhood. And such a project would obviously benefit from cooperation with psychologists and psychiatrists who are confronted with pathological cases of both self-knowledge and self-alienation.

THE QUESTION OF SELF-COINCIDENCE

Based on my account of selfhood, it is impossible that a person, when she performs an act of reflective self-awareness, coincides with herself qua self. One might think that our fundamental first-person self-conception of oneself as oneself is an exception to that impossibility. After all, that reflective act of self-awareness is nothing else than the reflective awareness of *being-self*. In a certain way, our conception of ourselves as ourselves *is* an exception to the impossibility of self-coincidence in reflective acts of self-awareness. That self-conception, however, is never a 'stand-alone' reflective act of self-awareness; it is always concealed in other acts of reflective self-awareness.

Although it is impossible to coincide with oneself qua self in acts of self-ascription or when one is engaged in explicit self-reflection, there are nevertheless degrees of detachment from oneself qua self, depending on the kind of act of self-awareness. Reflecting on one's life or personality, for example, requires a more detached stance towards oneself than do certain *embedded* acts of self-reflection, which take place in the context of a concrete action, for instance, one's rejection of a particular motivation to act in a certain way. So the question might arise which kind of reflective self-self relations (with the exception of the concealed self-conception of oneself as oneself) are as close as possible to self-coincidence.

Some philosophers believe that a person can experience pure self-consciousness; and those philosophers presumably think that a person, in that experience, coincides with herself qua self. One could also speculate that experiencing self-regarding moods or vague reflexive feelings like "es ist einem langweilig" or "es ist einem Angst" (German expressions that lose their specific meaning if translated) comes as close as possible to self-coincidence.

Anyway, I wonder whether my distinction between three kinds of acts of reflective self-awareness – embedded self-reflection, explicit self-reflection

and meta-reflection – and/or my distinction between first-person and third-person self-conceptions could be fine-tuned in such a way that we could propose an answer to the question which kinds of reflective acts of self-awareness approach self-coincidence as close as possible, in other words, which acts imply a minimum of self-alienation.

5.2 Regarding 'the three properties'

THE QUESTION OF CONNECTIONS BETWEEN THE THREE PROPERTIES THAT ARE CONSTITUTIVE OF PERSONHOOD IN ITS FULLEST SENSE

Cognitive and practical abilities like imagining oneself in this or that situation, taking a position towards one's motives (to act in a certain way), postponing the satisfaction of certain of one's desires, or knowing that one is going to die eventually require a first-person perspective. It is, however, conceivable that some of those abilities require a more mature stage of the capability of a first-person perspective than others. Compare, for example, the cognitive ability of knowing that one is going to die eventually and the practical ability of postponing the satisfaction of certain of one's desires. I am tempted to think that the former requires a more mature stage of the capability of a first-person perspective than the latter.

Assuming that we can identify maturity stages in the capability of a first-person perspective, the question may arise whether certain stages of the capability of a first-person perspective are dependent on one or both of the two non-fundamental properties, that is to say, whether these stages require the non-fundamental properties of a narrative self-conception and/or of applying self-regarding qualitative standards to be developed to a certain degree.

Two similar questions might arise concerning connections between the two non-fundamental properties: Do certain developmental degrees of the property of a narrative self-conception require the property of applying self-regarding qualitative standards to be developed to a certain degree too? And the same question can obviously be raised the other way round. To avoid misunderstanding, the question is not whether there are connections between a person's values and her autobiographical narrative, but whether there are connections on the structural level of the self, that is, on the level of the two properties in question.

THE QUESTION OF CONFLICTS BETWEEN ONE'S PURPOSES AND ONE'S VALUES

According to my account, adopting a purpose and having values requires different properties of the domain of personhood. Both require the property of a first-person perspective, but having values requires more than that, namely the property of making qualitative distinctions and applying them as standards to oneself. The situation is even more complex, because to adopt purposes that

in fact are life-goals requires a narrative self-conception, on top of a first-person perspective.

In this book, I spent some time discussing the question of conflicts between our purposes and our values and I argued for the overriding nature of our values.[2] I nevertheless believe that the question about the conflicts between our purposes and our values deserves more research than is devoted to it in the context of this project.

5.3 Regarding cultural and historical dimensions of personhood

According to my account, personhood – or rather personhood in its *fundamental* sense – is definitely *independent* of cultural or historical dimensions; it is solely dependent on the capability of a first-person perspective. A creature counts as a person if it has the capability of a first-person perspective, that is, the capability to conceive of itself as itself. Period. Therefore, certain questions concerning the possible historical and cultural dimensions of personhood can simply not intelligibly be raised, given my account of person- and selfhood.

Someone might, for instance, speculate that prehistorical persons may have lacked a sense of self or that a sense of self is the product or invention of Western culture. That position is not tenable, however. Firstly, we have to assume that such questions concern a *reflective* sense of selfhood, since a *pre-reflective* sense of selfhood is required for all (complex) purposeful behaviour and not restricted to personhood. Secondly, a reflective sense of selfhood simply arises from a person's fundamental capability of a first-person perspective. It exhibits itself, for instance, in our first-person self-conceptions of our actions as *our* actions or our thoughts as *our* thoughts. It is neither reserved to people that engage in hyper-reflection about their 'self' nor part of an elitist self-image or the result of self-absorption. Thirdly, creatures that lack a reflective sense of selfhood cannot but lack the capability of a first-person perspective (after all, the former arises from the latter) and thereby lack the features by virtue of which they would count as persons. Hence, a reflective sense of selfhood can neither be the product of a certain culture or an invention of certain philosophers, nor is it possible that prehistorical people lacked a reflective sense of selfhood.

Nevertheless, there are lots of questions regarding the historical and cultural dimensions of personhood that might, in the light of my account of person- and selfhood, legitimately be raised and which deserve investigation. Such questions may, for example, concern the possible historical dimensions of certain cognitive or practical abilities that are based on the capability of a first-person perspective. Think of our abilities to perform hyper-reflective

---

[2] See Chapter 4 (Wrongness from a self-related perspective), Section 6 (The overriding nature of the self-related conception of wrongness of action).

acts of self-evaluation or to play various roles, or of our highly developed planning and even contingency planning abilities. Questions regarding the historical dimensions of personhood may also concern the two non-fundamental but characteristic properties of persons – the properties of a narrative self-conception and of applying self-regarding qualitative standards. One might ask whether these capabilities developed simultaneously or whether one developed earlier than the other. One might also speculate about the circumstances that might have favoured the development of those properties and, for example, raise the question whether these properties might disappear if those circumstances did not obtain anymore.

Speculations regarding the cultural dimensions of personhood might even seem to be more obvious. To avoid wrong-headed questions and ideas, I should first call to mind that the questions which are relevant here have to concern the *structural* aspects of personhood, not people's ideas of, say, a good life. The latter, undoubtedly, show (sub)cultural variations – think of the difference between a focus on self-fulfilment or on one's duties to others; but this kind of difference does not imply differences in the structural aspects of personhood. Relevant are those questions that relate to the possible cultural dimensions of the properties constitutive of personhood or the cognitive and practical abilities that are based on those properties. Do contemporary 'normal' and mature persons generally have the properties of a narrative self-conception and of applying qualitative self-regarding standards developed to (approximately) the same degree, or is that degree – or even the development of the property in question – culture-dependent?

As to the property of applying qualitative self-regarding standards, it is a widespread common-sense belief in contemporary Western societies that these societies suffer from the fact that a significant number of people do not have values anymore. One might therefore be tempted to conclude that contemporary Western culture suffers from a lack of people's property of applying self-regarding qualitative standards. Analysis in the light of my account, however, may lead to the conclusion that the people in question do not lack values, but that other things function as values in their evaluative reasoning. In that case, what seemed to be a lack of values turns out to be a different ethical orientation; and that orientation is assessed negatively, as a lack of the 'right' values. A second example of the possible cultural dimensions of personhood concerns the property of a narrative self-conception. Certain literature on the Buddhist view on personhood seems to imply that that property is culture-dependent. I wonder whether an analysis of literature on the Buddhist view on personhood, when performed against the background of my account, would lead to the conclusion that Buddhists – or rather, Buddhists in the true sense – do not have a narrative self-conception. To be honest, I doubt it.

To sum up, my account of personhood does *not* exclude the possibility that certain dimensions of personhood (in its fullest sense) may be

historically or culturally dependent. The question remains whether personhood indeed has cultural and historical dimensions and what those dimensions are – a question that definitely merits investigation.

5.4 Regarding the concept of a good life

According to my account of personhood (in its fullest sense), persons entertain three 'models of self-conception': conceiving of oneself as a rational agent, as the protagonist of one's self-narrative, and as having values. Although the three 'models of self-conception' are not 'objects' of our direct reflective self-awareness, we may become aware of them through explicit reflection on our self-regarding reactive attitudes. They thereby can become part of our self-images, more specifically, of third-personal self-conceptions of being (or striving to be) a certain kind of person.

Some people entertain pronounced self-images around the three concerns. Think of people who believe that they are or should be immune to spontaneous desires, because such desires hinder the realization of adopted aims. Or take, by contrast, people who proclaim that they value irrationality. Or think of people who explicitly care about the coherence of their autobiographical narrative, or who, by contrast, proclaim that they deliberately strive for fragmentation.

Such observations might lead us to the ethical question how the three regions of mattering – one's rational agency, one's autobiographical narrative, and one's values – figure in our (contemporary) concept of a good life.

5.5 Regarding the need for unity of ourselves qua self

THE QUESTION CONCERNING A DEEPER FOUNDATION OF OUR SUFFERING FROM DISUNITY

If I am right, persons typically suffer from divergence in their 'will', fragmentation of their self-narratives, and non-integrity of themselves. In other words, they suffer from inconsistency, incoherence, and non-integrity of themselves qua self. In my view, these phenomena exhibit a need for unity of oneself qua self. I argued – and I based that argumentation solely on conceptual analysis – that our need for unity can only be grounded in our pre-reflective sense of identity over time. Nevertheless, I believe that this subject calls for other kinds of investigation too: it needs cross-fertilization with research in the areas of developmental psychology, psychopathology and the neurosciences.

THE QUESTION HOW MUCH (DIS)UNITY A PERSON CAN STAND

I emphasized that people typically suffer from *dis*unity, that is to say, from divergence in their 'will', fragmentation of their self-narratives, and non-integrity of themselves. This is however not to claim that people always and deeply suffer from each and every bit of disunity, so to speak. People generally accept that they, for instance, sometimes have a 'weak will' and that their lives are not straight-lined, and even that certain twists in their lives are difficult to be conceived of as personal developments, and that they sometimes, in their actions, are insensitive to the fact that certain of their values are at stake. My question is how much disunity a person can stand, given her desire for consistency, coherence and integrity.

Interestingly, and perhaps surprisingly – after all, I argued that persons have a need for unity – the opposite question can also be raised, namely, how much *unity* a person can stand. I think that we have reason to believe that people not only typically suffer from disunity but that they can also suffer from too much unity, more specifically, from an excessive concern with the soundness of their will, the coherence of their self-narrative and the integrity of their self. To put it differently, I am tempted to think that these concerns with ourselves qua self can become obsessive.

Both questions, how much unity a person can stand as well as how much disunity a person can stand, obviously bear on the domain of psychiatry and psychology; and to address those questions hence requires an interdisciplinary approach.

A LAST WORD CONCERNING THE VARIOUS NEW QUESTIONS

It will not have passed unnoticed that the various new questions are less concerned with the three properties that are constitutive of personhood (in its fullest sense) and their included self-conceptions, structures and dispositions than with their connections, with developmental and cultural aspects of personhood, and with issues of normativity and normality. And, in my view, most of these new questions require an interdisciplinary approach.

References

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

Baker, Lynne Rudder (2004), 'On being one's own person', in: Sie, Maureen, Slors, Marc and Van den Brink, Bert (eds.), *Reasons of One's Own*. Aldershot, Hampshire, UK and Burlington, USA: Ashgate, pp. 129-149.

Braude, Stephen E. (2004), 'Memory: The nature and significance of dissociation', in: Radden, Jennifer (ed.), *The Philosophy of Psychiatry. A Companion*. Oxford: Oxford University Press, pp. 106-117.

Brook, Andrew (1999), 'Unified consciousness and the self', in: Gallagher, Shaun and Shear, Jonathan (eds.), *Models of the Self*. Thorverton: Imprint Academic, pp. 39-47.

Butterworth, George (1999), 'A developmental-ecological perspective on Strawson's 'The self'', in: Gallagher, Shaun and Shear, Jonathan (eds.), *Models of the Self*. Thorverton: Imprint Academic, pp. 203-211.

Butterworth, George (2000), 'An ecological perspective on the self and its development', in: Zahavi, Dan (ed.), *Exploring the Self. Philosophical and Psychopathological Perspectives on Self-Experience*. Amsterdam, NL and Philadelphia, USA: John Benjamins Publishing Company, pp. 19-38.

Flanagan, Owen J. (1990), 'Identity and strong and weak evaluation', in: Flanagan, Owen J. and Rorty, Amélie Oksenberg (eds.), *Identity, Character and Morality.* Cambridge, Mass.: The MIT Press, pp. 37-66.

Flanagan, Owen J. (1996), *Self Expressions. Mind, Morals, and the Meaning of Life*. New York, Oxford: Oxford University Press.

Frank, Manfred (2002), *Selbstgefühl*. Frankfurt am Main: Suhrkamp Verlag.

Frankfurt, Harry (1971), 'Freedom of the will and the concept of a person', reprinted in: Frankfurt, Harry, *The Importance of What We Care About. Philosophical Essays*. Cambridge: Cambridge University Press, 1988, pp. 11-25.

Frankfurt, Harry (1976), 'Identification and externality', reprinted in: Frankfurt, Harry, *The Importance of What We Care About. Philosophical Essays*. Cambridge: Cambridge University Press, 1988, pp. 58-68.

Frankfurt, Harry (1982), 'The importance of what we care about', reprinted in: *The Importance of What We Care About. Philosophical Essays*. Cambridge: Cambridge University Press, 1988, pp. 80-94.

Gallagher, Shaun and Marcel, Anthony J. (1999), 'The self in contextualized action', in: Gallagher, Shaun and Shear, Jonathan (eds.), *Models of the Self*. Thorverton: Imprint Academic, pp. 273-300.

Jamison, Kay Redfield (1997), *An Unquiet Mind*. London: Picador.

Kogon, Eugen (1974), *Der SS-Staat*. München: Kindler.
Korsgaard, Christine M. (1997), 'The normativity of instrumental reason', in: Cullity, Garrett and Gaut, Berys (eds.), *Ethics and Practical Reason*. Oxford: Clarendon Press, pp. 215-254.
Kramer, Peter D. (1997), *Listening to Prozac*. New York: Penguin Books.
Legerstee, Maria (1999), 'Mental and bodily awareness in infancy', in: Gallagher, Shaun and Shear, Jonathan (eds.), *Models of the Self*. Thorverton: Imprint Academic, pp. 213-230.
Nida-Rümelin, Julian (2001), *Strukturelle Rationalität. Ein philosophischer Essay über praktische Vernunft*. Stuttgart: Reclam.
Olson, Eric T. (1999), 'There is no problem of the self', in: Gallagher, Shaun and Shear, Jonathan (eds.), *Models of the Self*. Thorverton: Imprint Academic, pp. 49-61.
Radden, Jennifer (1999), 'Pathologically divided minds, synchronic unity and models of self', in: Gallagher, Shaun and Shear, Jonathan (eds.), *Models of the Self*. Thorverton: Imprint Academic, pp. 343-358.
Radden, Jennifer (ed.) (2004), *The Philosophy of Psychiatry, A Companion*. Oxford: Oxford University Press.
Reinders, A.A.T.S., Nijenhuis, E.R.S., Paans, A.M.J., Korf, J., Willemsen, A.T.M., and Den Boer, J.A. (2003), 'One brain, two selves', in: *NeuroImage* 20. Elsevier Inc., pp. 2119-2125.
Scanlon, T.M. (1998), *What We Owe to Each Other*. Cambridge, Massachusetts and London, England: The Belknap Press of Harvard University Press.
Schechtman, Marya (1996), *The Constitution of Selves*. Ithaca: Cornell University Press.
Shoemaker, Sydney (1986), 'Introspection and the self', reprinted in: Shoemaker, Sydney, *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press, 1996, pp. 3-24.
Shoemaker, Sydney (1990), 'First person access', reprinted in: Shoemaker, Sydney, *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press, 1996, pp. 50-73.
Shoemaker, Sydney (1996), 'Unity of consciousness and consciousness of unity', in: Shoemaker, Sydney, *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press, pp. 176-197.
Smith, Michael (1994), *The Moral Problem*. Oxford: Blackwell Publishers.
Stephens, G. Lynn and Graham, George (2000), *When Self-Consciousness Breaks. Alien Voices and Inserted Thoughts*. Cambridge, Massachusetts and London, England: The MIT Press. First MIT Press paperback edition, 2003.
Strawson, Galen (1997), 'The self', reprinted in: Gallagher, Shaun and Shear, Jonathan (eds.), *Models of the Self*. Thorverton: Imprint Academic, 1999, pp. 1-24.
Taylor, Charles (1982a), 'The diversity of goods', in: Sen, Amartya and Williams, Bernard (eds.), *Utilitarianism and Beyond*. Cambridge: University Press, pp. 129-145.

Taylor, Charles (1982b), 'Responsibility for self', in: Watson, Gary (ed.), *Free Will*. Oxford: Oxford University Press, pp. 111-126.

Taylor, Charles (1985), 'What is human agency?', in: Taylor, Charles, *Human Agency and Language*. Cambridge, Mass.: Harvard University Press, pp. 15-44.

Taylor, Charles (1989), *Sources of the Self*. Cambridge: Cambridge University Press.

Tugendhat, Ernst (1979), *Selbstbewußtsein und Selbstbestimmung. Sprachanalytische Interpretationen*. Frankfurt am Main: Suhrkamp Verlag.

Velleman, J. David (1989), 'Epistemic freedom', reprinted in: Velleman, J. David, *The Possibility of Practical Reason*. Oxford: Clarendon Press, 2000, pp. 32-55.

Velleman, J. David (1992), 'What happens when someone acts', reprinted in: Velleman, J. David, *The Possibility of Practical Reason*. Oxford: Clarendon Press, 2000, pp. 123-143.

Velleman, J. David (2000), *The Possibility of Practical Reason*. Oxford: Clarendon Press.

Velleman, J. David (2002), 'Identification and identity', in: Buss, Sarah and Overton, Lee (eds.), *Contours of Agency. Essays on Themes from Harry Frankfurt*. Cambridge, Massachusetts: The MIT Press, pp. 91-123.

Wilson, Timothy D. (2002), *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, Massachusetts and London, England: The Belknap Press of Harvard University Press.

Zahavi, Dan and Parnas, Josef (1999), 'Phenomenal consciousness and self-awareness', in: Gallagher, Shaun and Shear, Jonathan (eds.), *Models of the Self*. Thorverton: Imprint Academic, pp. 253-279.

Samenvatting

Wij kunnen gevoelens hebben ten aanzien van onszelf en denken over onszelf, omdat wij het soort wezens zijn die wij zijn – namelijk personen. Met andere woorden, dankzij ons persoon-zijn kunnen wij ons denkend en voelend tot onszelf verhouden. Ik zal dit aan de hand van drie karakteristieke gevallen illustreren.

Ten eerste leidt het in het algemeen tot negatieve gevoelens en gedachten over onszelf als we in ons handelen door motieven worden geleid waardoor we 'eigenlijk' – althans zo denken en voelen we op dat moment – niet willen worden geleid. Zo kunnen we bijvoorbeeld teleurgesteld zijn over onszelf, ontevreden zijn met onszelf of kwaad zijn op onszelf, als we, door te doen waar we op dat moment zin in hadden, willens en wetens de realisering van een voorgenomen doel in gevaar hebben gebracht. Wij kunnen, ten tweede, op een manier hebben gehandeld die we tegenover onszelf niet kunnen verantwoorden. We hebben bijvoorbeeld het vertrouwen van anderen geschaad of geen mededogen of verantwoordelijkheidsgevoel getoond. In een dergelijke situatie kunnen we ons vertwijfeld of wanhopig voelen en ons realiseren dat we ons, op basis van ons handelen, niet als de persoon kunnen zien die we (nog steeds) zouden willen zijn: meelevend met anderen, verantwoordelijk en betrouwbaar. Als derde voorbeeld kunnen we denken aan iemand, die zijn leven in brokstukken voor zich ziet liggen en het op geen enkele manier meer als een samenhangend geheel kan begrijpen. Die persoon lijdt onder deze situatie, immers hij wil zijn leven als samenhangend en zichzelf als de protagonist van zijn levensverhaal kunnen begrijpen. Kortom, hij heeft negatieve gedachten en gevoelens over zichzelf.

Hoewel de genoemde voorbeelden zonder uitzondering *negatieve* zelfgerelateerde gedachten en gevoelens illustreren, kunnen wij uiteraard ook positieve zelfgerelateerde gedachten en gevoelens hebben, zoals tevreden zijn met ons handelen of met de manier waarop we ons leven leiden. Positieve zelfgerelateerde gedachten zijn vaak echter minder 'krachtig' en zij leiden, anders dan negatieve zelfgerelateerde gedachten en gevoelens, niet typisch tot verdere reflectie over ons handelen en over onszelf. Het zijn de afwijkingen van de ons eigen normen die tot soms hevige zelfgerelateerde emoties kunnen leiden en die ons zelfbegrip aan het wankelen kunnen brengen. Met 'de ons eigen normen' bedoel ik hier *niet* de doelen die wij onszelf hebben gesteld, de plannen die we voor onszelf hebben gemaakt of de gedragsregels die we voor onszelf hebben aanvaard. De normen die hier een rol spelen, zijn als het ware dieper verankerd; zij zijn onlosmakelijk verbonden met bepaalde eigenschappen die kenmerkend zijn voor personen.

Personen hebben als fundamentele eigenschap een hoogontwikkeld eerste-persoons-perspectief, dat wil zeggen, zij begrijpen zichzelf als zichzelf. Ik noem die eigenschap fundamenteel, omdat er zonder het vermogen om zichzelf als zichzelf te begrijpen geen sprake is van persoon-zijn; het is de basis voor iedere vorm van zelf*begrip*. Daarenboven (en wel letterlijk en figuurlijk) hebben personen twee kenmerkende, hoewel niet-fundamentele, eigenschappen: ten eerste een narratief zelfbegrip en ten tweede het op zichzelf toepassen van kwalitatieve normen. Alle drie voor persoon-zijn kenmerkende eigenschappen bevatten modellen van zelfbegrip: als wezens met een eerste-persoons-perspectief begrijpen we onszelf impliciet als rationele actoren; als wezens met een narratief zelfbegrip begrijpen we onszelf impliciet als de protagonist van ons eigen levensverhaal; als wezens die op zichzelf kwalitatieve standaarden toepassen, begrijpen we onszelf impliciet als wezens die waarden hebben.

De genoemde drie eigenschappen met hun eigen modellen van zelfbegrip liggen ten grondslag aan het feit dat het er voor onszelf toe doet op basis van die motieven te handelen waardoor we gemotiveerd willen worden, dat het er voor onszelf toe doet ons leven als een min of meer samenhangend geheel te kunnen begrijpen, en dat het er voor onszelf toe doet onszelf als een persoon te kunnen zien die bepaalde waarden heeft. De drie eigenschappen constitueren daarmee domeinen die er voor onszelf met betrekking tot onszelf qua zelf toe doen. Anders gezegd, op basis van de drie voor persoon-zijn kenmerkende eigenschappen zijn wij qua zelf op een karakteristieke manier voor onszelf van belang. Dit is de centrale these die in dit proefschrift wordt beargumenteerd.

HET ONDERZOEK VAN ZELF-ZELF-RELATIES

In het inleidende hoofdstuk worden allereerst zelfrelaties gecategoriseerd. Dit om *die* zelfrelaties, waar het in dit onderzoek hoofdzakelijk om gaat, van andersoortige zelfrelaties te onderscheiden en er niet mee te verwarren. De volgende indeling wordt gemaakt: (1) reflectieve zelf-zelf-relaties; (2) reflectieve zelf-lichaam-relaties; (3) prereflectieve zelf-zelf- of zelf-lichaam-relaties; (4) lichaam-lichaam-relaties.

Reflectieve zelf-zelf-relaties zijn relaties van een persoon qua persoon – dat wil zeggen, qua zichzelf als zichzelf begrijpend wezen – tot (aspecten van) zichzelf qua zelf. Tot deze categorie behoren de aan het begin van deze samenvatting beschreven voorbeelden, en om dit soort zelfrelaties gaat het in dit onderzoek. De tweede categorie, de reflectieve zelf-lichaam-relaties, betreft relaties van een persoon qua persoon tot (aspecten van) zichzelf qua lichaam. Iemand kan bijvoorbeeld ongelukkig zijn met zijn postuur of zichzelf attractief vinden, of blij zijn of juist ontevreden met haar gekrulde haar. Dit soort zelfrelaties ligt buiten het bestek van dit onderzoek. De derde categorie zijn prereflectieve zelf-zelf- of zelf-lichaam-relaties. Te denken valt bijvoorbeeld aan onderdrukte zelfgerelateerde gevoelens, die in bepaalde

situaties in iemands reacties tot uiting lijken te komen. Deze zelfrelaties zijn prereflectief in de zin dat de persoon in kwestie zich (op dat moment) niet van de bewuste reflexieve houdingen en gevoelens bewust is, maar niet in de zin dat zij aanwezig zouden kunnen zijn bij wezens zonder het vermogen tot een eerste-persoons-perspectief. Prereflectieve zelf-zelf- of zelf-lichaam-relaties veronderstellen wel degelijk persoon-zijn. Zij komen in dit onderzoek (in de hoofdstukken 5 en 6) zijdelings aan bod. De vierde en laatste categorie betreft lichaam-lichaam-relaties: relaties van een persoon qua levend menselijk organisme tot zichzelf. Dit soort zelfrelaties vereist niet het vermogen om zich als zichzelf te begrijpen. Voorbeelden zijn ons natuurlijke instinct om te overleven of de compenserende bewegingen van ons lichaam als we dreigen ons evenwicht te verliezen. Deze categorie zelfrelaties speelt in dit onderzoek geen rol en komt daarom slechts zijdelings aan de orde, namelijk in het kader van speculaties over het fenomeen ‘meervoudige persoonlijkheidsstoornis’.

Zoals reeds gezegd, dit proefschrift is primair gewijd aan reflectieve zelf-zelf-relaties. Het is een poging om greep te krijgen op het feit dat wij qua zelf er op karakteristieke manieren voor onszelf toe doen. Er wordt *niet* gevraagd hoe wij als individuen of als soort ertoe zijn gekomen dat wij op deze karakteristieke manieren voor onszelf van belang zijn. Ook de ethische vraag in welke zelf-zelf-relaties wij zouden moeten staan is niet aan de orde. Het gaat erom de structuren van bepaalde zelf-zelf-relaties en de disposities die daarin een rol spelen te begrijpen. De benadering is analytisch en structureel, niet historisch. Bij de behandeling van de diverse deelproblemen wordt in de regel begonnen bij alledaagse prefilosofische noties of algemeen bekende fenomenen die blijk geven van een reflectieve zelf-zelf-relatie, zoals onszelf qua zelf voor veranderlijk houden, onszelf qua zelf als iets constants beschouwen, onszelf in bepaalde situaties voor irrationeel houden of onszelf distantiëren van sommige van onze verlangens of handelingen. Dit soort fenomenen en prefilosofische noties wordt geanalyseerd tegen de achtergrond van een bepaald concept van persoon-zijn. De verkenningen zijn vaak descriptief, hoewel niet empirisch. Het punt waar het om gaat is conceptueel: dankzij de vermogens op basis waarvan personen als personen gelden (in de volste zin van persoon-zijn), staan zij in kenmerkende relaties tot zichzelf qua zelf.

Een adequate theorie van persoon-zijn, zo wordt gesteld, mag zich niet beperken tot het laten zien welke vermogens het persoon-zijn vereist, maar dient de zelfgerelateerde disposities, neigingen en behoeften te verklaren die kenmerkend zijn voor personen. De in dit proefschrift gepresenteerde analyse van zelf-zelf-relaties neemt derhalve het eerste-persoons-perspectief serieus. Door verschillende manieren te analyseren waarop wij onze handelingen, doelen, motieven, overtuigingen, waarden, karaktertrekken of ons leven begrijpen, worden structuren en disposities blootgelegd die inherent zijn aan bepaalde kenmerkende zelf-zelf-relaties. De manieren waarop we denken en voelen ten aanzien van onszelf, zijn, zo wordt

betoogd, niet slechts contingent, dat wil zeggen onze zelfgerelateerde gedachten en gevoelens zijn niet alleen afhankelijk van bijvoorbeeld onze handelingen en ervaringen, ons karakter en onze biografie of de reactie van anderen op ons gedrag, maar zij hebben ook bepaalde apriorische structuren en omvatten bepaalde disposities, neigingen en behoeften. We kunnen bijvoorbeeld niet anders dan *onszelf* als de bron van onze handelingen (willen) begrijpen. Immers, als wij denken en voelen dat niet *wij*, maar krachten *in* ons bepalen wat wij doen, distantiëren wij ons van die handelingen. En het zou ons, voorzichtig gezegd, ongerust maken als wij regelmatig precies die handelingen zouden uitvoeren die wij op dat moment niet willen uitvoeren, als wij, als het ware, tegen onze eigen wil zouden handelen. Beide gevallen – dat personen zich onder bepaalde omstandigheden van hun handelingen distantiëren en dat het beangstigend zou zijn als 'een vreemde wil zich van ons meester zou maken' – maken duidelijk dat personen zich in principe als de bron van hun handelingen voelen en begrijpen, en dat het er voor hen toe doet zich als de bron van hun handelingen te kunnen voelen en begrijpen.

Reeds deze twee voorbeelden laten zien dat dit onderzoek van zelf-zelf-relaties diverse onderwerpen en gebieden van de filosofie raakt. Een prominente rol spelen de onderwerpen persoon-zijn, het zelf, persoonlijke identiteit en praktische rationaliteit, en daarbij worden de domeinen van de handelingsfilosofie, de wijsgerige antropologie, de 'philosophy of mind', de filosofie van de psychiatrie en de ethiek geraakt. De gekozen benadering – het serieus nemen van het eerste-persoon-perspectief en het analyseren van diverse zelf-zelf-relaties tegen de achtergrond van een concept van persoon-zijn – verheldert verbindingen tussen deze domeinen die anders onderbelicht zijn. Het fenomeen van dissociatie, bijvoorbeeld, is typisch een onderwerp uit de filosofie van de psychiatrie. In dit onderzoek wordt 'dissociatie' echter niet beperkt tot pathologische gevallen, maar in een bredere context geplaatst. Daardoor komen vragen naar voren als 'Van welke aspecten van zichzelf kan een persoon zich distantiëren en wat zijn de grenzen van dissociatie?', 'Welke dissociatieve fenomenen kunnen als mentale akten van de persoon qua persoon gelden?', 'Kan een persoon zich van zichzelf qua zelf dissociëren?'.

De analyse van allerlei fenomenen van zelf-zelf-relaties heeft in dit onderzoek steeds hetzelfde uitgangspunt, namelijk een bepaald concept van persoon-zijn: personen zijn wezens met het vermogen van een eerste-persoons-perspectief, dat wil zeggen, het vermogen zichzelf als zichzelf te begrijpen. Hoewel een bepaald persoonsconcept wordt verondersteld, is dit onderzoek toch een zoektocht naar een bevredigende theorie van persoon-zijn. De circulariteit die in de gekozen benadering aanwezig is, is niet problematisch, omdat het gepresenteerde persoonsconcept steeds rijker en verfijnder wordt. Zo wordt bijvoorbeeld in de loop van het onderzoek duidelijk, dat het genoemde concept van persoon-zijn (personen zijn wezens die zichzelf als zichzelf begrijpen) tekort schiet als het er om gaat

begrijpelijk te maken dat wijzelf qua zelf voor onszelf 'gegeven zijn', dat personen zich niet hoeven te identificeren als zij zichzelf als zichzelf begrijpen. Zichzelf als zichzelf begrijpen veronderstelt, zo wordt geargumenteerd, een prereflectief gevoel van zelf-zijn. Ons prereflectieve gevoel van zelf-zijn zou – hier wordt het betoog enigszins speculatief – de basis kunnen zijn voor ons verlangen naar eenheid van onszelf qua zelf, een verlangen dat zich – als behoefte aan consistentie, coherentie en integriteit – in verschillende van de geanalyseerde zelf-zelf-relaties toont.

DE CENTRALE CLAIMS EN AANNAMES VAN DE HOOFDSTUKKEN

De in deze samenvatting opgenomen tabel geeft een overzicht van de claims en aannames van de verschillende hoofdstukken. Hierbij is het inleidende hoofdstuk niet meer vermeld, omdat de inhoud ervan reeds ruimschoots aan bod is gekomen.

Tabel: Overzicht van de hoofdstukken 2 tot en met 7

Hoofdstuk 2: De relatie van onszelf tot ons 'zelf'
(originele titel: *The relation of oneself to one's self)*
Het tweede hoofdstuk presenteert een structurele theorie van 'het zelf' en bespreekt een bijzondere relatie waarin personen dankzij hun persoon-zijn tot zichzelf staan: de relatie van er-voor-zich-zelf-qua-zelf-toe-doen. De theorie onderscheidt drie componenten van deze relatie, die ieder een eigen domein vormen van het er-voor-zich-zelf-qua-zelf-toe-doen en die hun eigen structuren hebben.
Het hoofdstuk steunt op de volgende centrale aanname: het (volwassen) zelf wordt gezamenlijk geconstitueerd door drie eigenschappen: de eigenschappen van een eerste-persoons-perspectief (dit is de fundamentele eigenschap), een narratief zelfbegrip en het maken van kwalitatieve verschillen.

Hoofdstuk 3: Praktische irrationaliteit vanuit een zelfgerelateerd perspectief
(originele titel: *Practical irrationality from a self-related perspective)*
In dit hoofdstuk worden twee fenomenen van het er-voor-zich-zelf-qua-zelf-toe-doen onderzocht: zijn handelingen irrationeel vinden en zichzelf irrationeel vinden. De stelling is dat onze eigen irrationaliteiten niet alleen daarom er voor onszelf toe doen omdat zij belemmeren dat wij onze doelen bereiken, maar ook op een veel directere manier.
Het hoofdstuk steunt op de volgende centrale aanname: of wij onze handeling als rationeel, irrationeel, rationeel-en-irrationeel, onrationeel, als gebaseerd op een vergissing, of als fout, slecht of onjuist begrijpen, hangt ervan af hoe die handeling in verband staat met een mogelijk proces van evaluatief redeneren.

Hoofdstuk 4: Fout handelen en fout zijn vanuit een zelfgerelateerd perspectief

(originele titel: *Wrongness from a self-related perspective*)

In dit hoofdstuk worden twee fenomenen van het er-voor-zich-zelf-qua-zelf-toe-doen onderzocht: zijn handelingen fout, slecht of onjuist vinden en zichzelf fout of slecht vinden. De stelling is: als we onze handeling fout, slecht of onjuist vinden, dan is deze beoordeling overheersend ten aanzien van andere mogelijke beoordelingen van onze handeling.

Het hoofdstuk steunt op de volgende centrale aanname: doelen en waarden spelen een essentieel verschillende rol in ons evaluatieve redeneren, en wel omdat zij betrekking hebben op twee verschillende manieren van reflectie over ons handelen.

Hoofdstuk 5: Zelf-zijn: Eenheid in veranderlijkheid

(originele titel: *Selfhood: Unity in changeability*)

In dit hoofdstuk wordt de (schijnbare) tegenstrijdigheid besproken tussen het feit dat wij onszelf qua zelf als één en hetzelfde begrijpen (door de tijd heen) en het feit dat wij ons bewust zijn van onze veranderlijkheid. De stelling is dat bepaalde kenmerkende manieren om over onszelf te denken en voelen er blijk van geven dat wij een verlangen hebben naar eenheid van ons zelf.

Het hoofdstuk steunt op de volgende centrale aanname: persoon-zijn – omdat het reflectief zelfbewustzijn vereist – veronderstelt een prereflectief gevoel van zelf-zijn.

Hoofdstuk 6: Zelfbegrip en de grenzen van dissociatie

(originele titel: *Self-conception and the limits to dissociation*)

In dit hoofdstuk worden verschillende fenomenen van dissociatie tegen de achtergrond van een theorie van zelfbegrip geanalyseerd. De stelling is dat het om conceptuele redenen onmogelijk is dat een persoon zich van zichzelf qua zelf zou kunnen dissociëren.

Het hoofdstuk steunt op de volgende centrale aanname: men dient tussen een eerste-persoons-zelfbegrip en een derde-persoons-zelfbegrip te onderscheiden. In tegenstelling tot een eerste-persoons-zelfbegrip vereist een derde-persoons-zelfbegrip dat wij onszelf qua zelf conceptualiseren en symboliseren.

Hoofdstuk 7: Conclusie: Persoon-zijn, modellen van zelfbegrip en wat er voor onszelf qua zelf toe doet

(originele titel: *Conclusion: Personhood, models of self-conception, and regions of mattering*)

In dit hoofdstuk worden verbanden gelegd tussen de voorafgaande hoofdstukken. Daarbij worden juist die aspecten van de gepresenteerde theorie van persoon- en zelf-zijn benadrukt waarvan te verwachten is dat zij bestreden zullen worden.

DE MEEST CONTROVERSIËLE CONCLUSIES

Dit onderzoek naar zelf-zelf-relaties leidt tot een aantal controversiële stellingen. Hier is een selectie daarvan:

1. Persoon-zijn en zelf-zijn zijn twee kanten van dezelfde medaille.
2. Zelf-zijn heeft een prereflectieve voorganger, namelijk een gevoel van zelf-zijn.
3. Of een handeling rationeel is hangt er alleen van af of de persoon zijn handeling beschouwt als voortkomend uit zijn eigen redeneren.
4. Autobiografische levensverhalen zijn niet identiteitsconstituerend in een fundamentele zin.
5. Het fundamentele zelfbegrip van zichzelf als zichzelf is inhoudsloos.
6. Personen kunnen niet anders dan zichzelf als zichzelf begrijpen.
7. Zelf-zelf-relaties hebben, met uitzondering van de fundamentele zelfzelf-relatie van zichzelf als zichzelf begrijpen, altijd betrekking op aspecten van zichzelf qua zelf.
8. Reflectieve zelf-zelf-relaties vereisen niet noodzakelijkerwijs een afstandelijke houding ten aanzien van zichzelf.
9. Persoon-zijn houdt drie modellen van zelfbegrip in, ieder met eigen structuren en disposities: het zelfbegrip als een rationele actor, als de protagonist van het eigen levensverhaal en als een wezen dat waarden heeft.
10. Het fundamentele zelfbegrip van zichzelf als zichzelf en de drie modellen van zelfbegrip moeten worden onderscheiden van daarmee schijnbaar corresponderende zelfbeelden, zoals de gedachte 'ik ben mijzelf' of het zelfbeeld een rationele persoon te zijn.

Curriculum vitae

Christiane E. Seidel was born in Lübeck, Germany, on November 17, 1949. After finishing her pre-universitary education at a gymnasium in Lübeck (1968), she attended a secondary teacher training course in Kiel (1968-1971). Her major subject was mathematics. After teaching at a secondary school in Hamburg (1971-1973), she entered the computer industry and got a position at the Frankfurt training centre of one of the leading multinational computer manufacturers. There she gave and developed courses on operating systems, programme languages and software development methods (1973-1978). Her career in the IT-industry was continued at the training centre of a German computer manufacturer in Wiesbaden (1979-1982). After that she moved to the Netherlands and accepted a job at the Dutch subsidiary of a multinational computer company in Utrecht, where she applied herself to project management, quality management, and consultancy, and held various positions in those areas (1983-2000).

In 1996, she began to study philosophy at Utrecht University and took a Master's degree in Practical Philosophy (cum laude, 2002). Her Master's thesis was titled 'The relation of ourselves to ourselves'. While completing her study, she lectured at Utrecht University (2001-2002), where she developed and taught undergraduate courses in Philosophy of Management and Organisation. In 2002 she became a Ph.D. student at Erasmus University Rotterdam, where she was able to continue her research in the philosophy of personhood. This thesis is the result of that research.